Animal Genetics & Breeding

PRINCIPLES OF ANIMAL AND POPULATION GENETICS

(UNIT - II)

- Alleles: $p + q = 1$

B b

- Individuals: $p^2 + 2pq + q^2 = 1$

BB Bb bb

Lecture notes on

Hardy Weinberg Law

First Edition

Department of Animal Genetics & Breeding

College of Veterinary & Animal Sciences

Sardar Vallabhbhai Patel University of Agriculture & Technology

Modipuram, Meerut- 250 110

Published Online:

First edition: 2021
Total pages: 42

Published by:

Department of Animal Genetics & Breeding
College of Veterinary & Animal Sciences
Sardar Vallabhbhai Patel University of Agriculture and Technology
Meerut- 250 110, Uttar Pradesh, India

Publication No.

SVP/2020/06/02/230 Dated: July 10, 2021 (for official use)

How to cite this Lecture notes

Tyagi, K 2021, *Hardy Weinberg Law*, lecture notes, Principles of Animal and Population Genetics AGB UNIT II, Sardar Vallabhbhai Patel University of Agriculture & Technology, Meerut, India, Delivered 16, 17 March, 2021. Retrieved online from https://vepub.com

Address Correspondence to:

Dr. Kuldeep Kumar Tyagi
Associate Professor & OIC
Department of Animal Genetics & Breeding
COVAS, SVPUAT, Meerut- 250110 (U.P.) India
drtyagivet@gmail.com
+91 9601283365 (M)

ABOUT

These lecture notes on “Hardy Weinberg Equilibrium” were prepared and delivered to my undergraduate students studying Animal Genetics & Breeding course. This course is offered in the second professional year of Bachelor of Veterinary Science & Animal Husbandry degree at College of Veterinary & Animal Sciences, S.V.P.U.A.T, Meerut, Uttar Pradesh, India. This lecture paves the foundation of population genetics. It illuminates students about the intricacies, concept and application of Hardy Weinberg law in population genetics. Simple derivation on a single locus with two alleles of this law has been further generalized for multiple locus and ploidy levels. Application of Hardy Weinberg law in understanding population structure under different modes of inheritance, selection experiments, understanding influence of evolutionary forces, genetic counselling and detection of genotypic errors have been touched upon. Thereafter special situation of sex linkage and it’s generalization using Jacobsthal series has been introduced. Ternary plots and testing for Hardy Weinberg equilibrium have also been included. All these intricacies make this lecture notes pivotal in understanding population genetics further. Use of explanatory illustrations, examples and figures had deliberately been used to create an interest among the students. I had tried my level best to simplify the concept in easy to understand language. Further constructive suggestions to improve this lecture notes are always welcome from readers on my email and whatsapp.

[KULDEEP KUMAR TYAGI]

DISCLAIMER

These lecture notes on "Hardy Weinberg Law" have been compiled from various resources available in the public domain. Only excerpts from the original works have been used. This is being done for educational purposes in the interest of developing a concise and updated reading material for students with no intent of commercial benefits. References to the source of material used have been included in the footnotes. The author does not claim any ownership of any copyrighted material included by chance in the lecture. If due to inability to trace the original source any copyrighted material got included, it may please be brought to the notice of the author for rectification.

[KULDEEP KUMAR TYAGI]

OTHER LECTURE NOTES BY THE AUTHOR

Series: Animal Genetics & Breeding							
S.No.	Title	Date of Publication	Editions	Pages	Views	Down loads	Access
1	Introduction and importance of statistics and biostatistics	12-10-2020	First	20	40.0k	251	Download
2	Probability	20-08-2020	First	33	12.5k	190	Download
3	Probability	24-11-2020	Second	24	29.5k	277	Download
4	Probability Distribution	07-12-2020	First	33	12.5k	190	Download
5	Introduction of population genetics	23-02-2021	First	16	22.8k	124	Download
6	Genetic structure of population	23-02-2021	First	40	22.1k	207	Download

TABLE OF CONTENTS

Hardy Weinberg Law

1. Introduction

Hardy Weinberg law opened the gateway to understand genetic structure of population over time. It was discovered independently by two scientists and this law is known today by their names. It has remained one of the landmark discoveries in the field of population genetics. Based on probability theory and certain pre-conditions this law is essentially the base for various later discovered models in population genetics. This law provides the basic framework to understand the genetics behind the evolutionary biology of ever changing various species on this earth. It is equally important for animal breeders since the majority of loci which are not under direct selection of breeders follow this law. Therefore understanding of this law becomes an important aspect when we move towards Quantitative genetics from Mendelian genetics. In this lecture we will first try to understand the Panmixia which is the first and foremost requirement for a population to come under Hardy Weinberg equilibrium. Thereafter we will elaborate on Hardy Weinberg law in particular, its history, derivation application and extensions.

2. Panmixia

The word panmixia is a modern latin word that has been derived in the late 19th century from German word *Panmixie*. The etymology of German word *Panmixie* suggests its derivation from Greek word *pan* which means 'all' and *mixis* that means 'mixing'. Thus it sounds like mixing all you have. Panmixia is synonymous to random mating in population genetics. Random mating describes an ideal situation in which all individuals on one sex are equally potential partners of all members of the opposite sex. A population that obeys idealized situations of random mating is said to be panmictic. This assumes that there are no mating restrictions, neither genetic nor behavioural, upon the population and that therefore all recombination is possible. Panmixia have greater place in population genetics because i) it enable us to decipher genetic structure of population over generations ii) it is the complementary force against factor of fixation (F) proposed by Sewell Wright in a population, which he later used to describe F statistics, iii) it help us to understand genetic structure of quite high number of remaining genes as against few genes affected by artificial selection in animal breeding.

3. Hardy Weinberg Law

The law is named after english mathematician Godfrey Harold Hardy and German obstetrician-gynecologist Wilhelm Weinberg. This law was discovered by them in the year 1908. Wilhelm Weinberg delivered an exposition of his ideas in a lecture on January 13, 1908, before the Society for the Natural History of the Fatherland in Württemberg, about three months before the Hardy's notes from April 1908 of that year and five months before Hardy's paper was published in English in June 1908. His lecture was printed in the society's yearbook in September 1908[1]. The Hardy–Weinberg principle, also known as the Hardy–Weinberg equilibrium, model, theorem, or law, states that "gene and genotype frequencies in a large random mating population will remain constant from generation to generation in the absence of mutation, migration and selection".

However it is quite evident now that many other evolutionary forces like genetic drift, mate choice, assortative mating, sexual selection, meiotic drive, genetic hitchhiking, population bottleneck, founder effect and inbreeding etc also affect gene and genotypic frequencies over generations. The effects of these evolutionary forces become more marked if population size is small. In this lecture we will concentrate on Hardy Weinberg Law in an idealized population where its associated preconditions are fulfilled. These preconditions help us to attain simple mathematical derivations to solve the puzzle of gene and genotypic frequencies over generations. Further deviations from these preconditions can be included in the equations once we understand Hardy Weinberg law in its original form.

3.1 Derivation and proof

Hardy–Weinberg law provides us the facility to forecast future gene and genotypic frequencies in a population. To add upon the simplicity it also states that in a large random mating population the gene and genotypic frequencies remain constant in absence of disruptive forces. Therefore if we know the allelic frequencies in a population for a locus and preconditions of Hardy Weinberg law are met then all future generations will have the same allelic and genotypic frequencies. *To understand derivation in its simplest form let us consider a single autosomal locus with two alleles*.

[1] Stern, Curt (1943). "The Hardy–Weinberg law". *Science*. **97** (2510): 137–138. doi:10.1126/science.97.2510.137

The following seven preconditions were observed for derivation of Hardy Weinberg equilibrium equations[2]

Preconditions:

1. **Diploid organisms:** To deduce the equations diploid organisms (2n) having two sets of chromosomes were taken into consideration.
2. **Sexual Reproduction:** The population of diploid organisms considered for deriving equations was considered to be reproduced through sexual mode only.
3. **Non overlapping generation:** Mating systems where only one breeding generation is present at any one time was considered.
4. **Random mating:** Mating between the genotypes under consideration occurs at random. Every individual have equal opportunity to mate with any individual of opposite sex.
5. **Large population:** Parental breeding population under study is large enough such that sampling errors are negligible. Usually for practical purposes a population in hundreds or thousands is generally considered a large population.
6. **Equal allelic frequency in sexes:** The allelic frequency in male and female population must be the same.
7. **No Selection, migration or mutation:** There should be no selection operating against viability and reproduction of different genotypes in the breeding population. The breeding population under study must be closed with respect to gene flow. No individual should move out of the population nor outside individuals should get entered into the breeding population under study. There is no mutation in the breeding population from one allelic state to another.

Step 1: Calculation of allelic and genotypic frequency in parent generation and formation of gametes.

We have learnt in our previous lecture to find allele frequencies in a population. Tabulation for single autosomal locus with two alleles (A, a) will be as follows,

[2] Revised from Hartl DL, Clarke AG (2007). Principles of population genetics. Sunderland, MA: Sinauer

(.. , generation number)

A	**a**	**Genotypes**	**Number $n_{(.. , 0)}$**	**$n_{(.. , 0)} \times$ A**	**$n_{(.. , 0)} \times$ a**
2	0	AA	$n_{(AA , 0)}$	$2 \times n_{(AA , 0)}$	$0 \times n_{(AA , 0)}$
1	1	Aa	$n_{(Aa , 0)}$	$1 \times n_{(Aa , 0)}$	$1 \times n_{(Aa , 0)}$
0	2	aa	$n_{(aa , 0)}$	$0 \times n_{(aa , 0)}$	$2 \times n_{(aa , 0)}$
		Sum	$N_{(.. , 0)}$	$n_{(A , 0)}$	$n_{(a , 0)}$

Genotypic frequencies:

$$f_0(AA) = \frac{n_{(AA , 0)}}{N_{(.. , 0)}} , f_0(Aa) = \frac{n_{(Aa , 0)}}{N_{(.. , 0)}} \text{ and } f_0(aa) = \frac{n_{(aa , 0)}}{N_{(.. , 0)}}$$

$$f_0(AA) + f_0(Aa) + f_0(aa) = 1$$

Allelic frequencies:

Here, $n_{(A , 0)} = 2n_{(AA , 0)} + n_{(Aa , 0)}$, $n_{(a , 0)} = 2n_{(aa , 0)} + n_{(Aa , 0)}$ and $n_{(. , 0)} = n_{(A , 0)} + n_{(a , 0)}$

Now let the allelic frequency of the first allele be denoted by "p_0" and that of the second allele be denoted by "q_0" in the parent generation.

Therefore, $p_0 = \frac{n_{(A , 0)}}{n_{(. , 0)}}$ and $q_0 = \frac{n_{(a , 0)}}{n_{(. , 0)}}$; $p_0 + q_0 = 1$ ..... (1)

Here one should understand that for estimation of gene and genotypic frequencies, parent populations need not to follow preconditions of Hardy Weinberg law. The parent population may or may not be in Hardy Weinberg equilibrium.

Step 2: Union of gametes to produce the first progeny generation.

Now we will impose the conditions of Hardy Weinberg law on the parental breeding population. Now the parental population being large enough will have almost equal numbers of males and females. They will produce gametes bearing allele "A" and "a" in the proportion equivalent to equation 1. The unison of these alleles randomly will result in progeny with genotypic frequencies as depicted in figure given below.

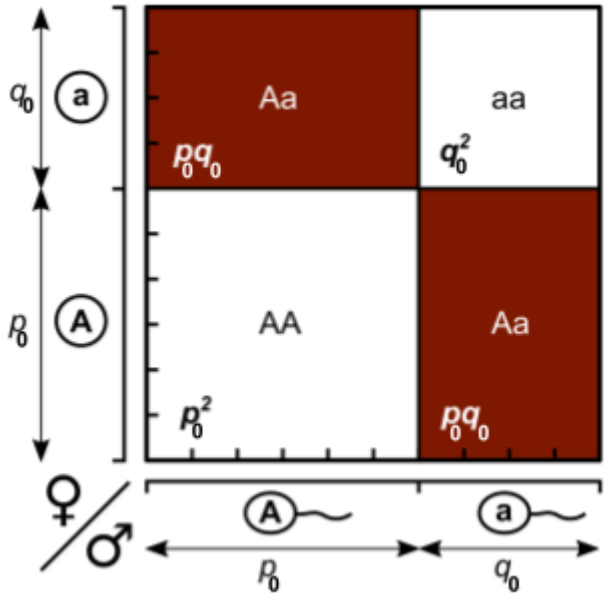

Fig 1: Two dimensional diagramatic presentation for proportion of genotypic frequencies in progeny generation[3].

Now let us recalculate the gene and genotypic frequencies in first generation,

(.. , generation number)

A	**a**	**Genotypes**	**Number $n_{(.. , 1)}$**	**$n_{(.. , 1)} \times$ A**	**$n_{(.. , 1)} \times$ a**
2	0	AA	$n_{(AA , 1)}$	$2 \times n_{(AA , 1)}$	$0 \times n_{(AA , 1)}$
1	1	Aa	$n_{(Aa , 1)}$	$1 \times n_{(Aa , 1)}$	$1 \times n_{(Aa , 1)}$
0	2	aa	$n_{(aa , 1)}$	$0 \times n_{(aa , 1)}$	$2 \times n_{(aa , 1)}$
		Sum	$N_{(.. , 1)}$	$n_{(A , 1)}$	$n_{(a , 1)}$

Genotypic frequencies:

$$f_1(AA) = \frac{n_{(AA , 1)}}{N_{(.. , 1)}} = p_0^2 \Rightarrow n_{(AA , 1)} = p_0^2 \times N_{(.. , 1)} \quad \ldots\ldots\ldots\ldots (2)$$

$$f_1(Aa) = \frac{n_{(Aa , 1)}}{N_{(.. , 1)}} = 2p_0q_0 \Rightarrow n_{(Aa , 1)} = 2p_0q_0 \times N_{(.. , 1)} \quad \ldots\ldots\ldots\ldots (3)$$

[3] Adapted and revised from https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle#/media/File:Hardy%E2%80%93Weinberg_law_-_Punnett_square.svg

$$\text{and } f_1(aa) = \frac{n_{(aa\,,\,1)}}{N_{(..\,,\,1)}} = q_0^2 \Rightarrow n_{(aa\,,\,1)} = q_0^2 \times N_{(..\,,\,1)} \quad \text{..... (4)}$$

$$f_0(AA) + f_0(Aa) + f_0(aa) = 1, \text{ also } p_0^2 + 2p_0q_0 + q_0^2 = 1$$

Now, one should understand that, if parent generation had been produced by a population which was not in Hardy Weinberg equilibrium than,

$$f_0(AA) \neq f_1(AA),\ f_0(Aa) \neq f_1(Aa) \text{ and } f_0(aa) \neq f_1(aa)$$

Allelic frequencies:

Here,

$n_{(.\,,\,1)} = n_{(A\,,\,1)} + n_{(a\,,\,1)} \Rightarrow 2N_{(..\,,\,1)} = n_{(A\,,\,1)} + n_{(a\,,\,1)}$ .....(Since, $n_{(.\,,\,1)} = 2N_{(..\,,\,1)}$ for diploids)

Also,

$n_{(A\,,\,1)} = 2n_{(AA\,,\,1)} + n_{(Aa\,,\,1)}$

$= 2p_0^2 \times N_{(..\,,\,1)} + 2p_0q_0 \times N_{(..\,,\,1)}$ ..... (from eqn 2)

$= 2\,N_{(..\,,\,1)}\,(p_0^2 + p_0q_0)$

$= 2\,N_{(..\,,\,1)}\,p_0\,(p_0+q_0)$

$= 2\,N_{(..\,,\,1)}\,p_0$ ..... (from eqn 1)

Similarly,

$n_{(a\,,\,1)} = 2\,N_{(..\,,\,1)}\,q_0$

Now let the allelic frequency of the first allele be denoted by "p_1" and that of the second allele be denoted by "q_1" in the parent generation.

Therefore,

$$p_1 = \frac{n_{(A\,,\,1)}}{n_{(.\,,\,1)}}$$

$$= \frac{2\,N_{(..\,,\,1)}p_0}{2N_{(..\,,\,1)}}$$

$$= p_0$$

$q_1 = \frac{n_{(a\,,\,1)}}{n_{(.\,,\,1)}}$

$= \frac{2\,N_{(..\,,\,1)}q_0}{2N_{(..\,,\,1)}}$

$= q_0$

and, $p_1+q_1=1$

Hence the allelic frequency remained same in first generation

Step 3: Union of gametes to produce the future progeny generations

The step 2 will be repeated over generations and we will find that

$$f_1(AA) = f_2(AA) = \ldots\ldots\ldots\ldots f_n(AA) = p_0^2$$

$$f_1(Aa) = f_2(Aa) = \ldots\ldots\ldots\ldots f_n(Aa) = 2p_0q_0$$

$$f_1(aa) = f_2(aa) = \ldots\ldots\ldots\ldots f_n(aa) = q_0^2$$

Also, $f_n(AA) + f_n(Aa) + f_n(aa) = 1$, also $p_0^2 + 2p_0q_0 + q_0^2 = 1$ (where, n=0, 1, 2, 3....)

Similarly,

$$p_1 = p_2 = \ldots\ldots\ldots\ldots p_n = p_0$$

$$q_1 = q_2 = \ldots\ldots\ldots\ldots q_n = q_0$$

Also, $p_n + q_n = 1$ (where, n=0, 1, 2, 3....)

3.2 Properties

The population under Hardy–Weinberg equilibrium exhibits the following properties.

1. The allelic and genotypic frequencies remain the same from generation to generation.
2. The genotypic frequencies can solely be determined by the allelic frequencies from the first generation onwards.
3. If preconditions for Hardy Weinberg equilibrium are satisfied in a population, single generation of random mating is sufficient to bring the population

under Hardy Weinberg equilibrium. The population remains under Hardy Weinberg equilibrium until preconditions are not violated.

Example 1: A breeder randomly selects 2000 short horn cattle. The phenotypic colour differentiation was 600 red, 800 roan and 600 white. To nullify the effect of disruptive forces on gene frequency at all the loci he decided to mate them fulfilling all the preconditions for Hardy Weinberg equilibrium. Considering the locus for colouration to be codominant justify that allelic and genotypic frequencies will remain the same from generation to generation?

Solution 1: Let A and a be two codominant alleles for colouration in cattle, such that AA = Red, Aa = Roan and aa = white colour.

Genotypic frequencies in parent generation,

$$f_0(AA) = \frac{n_{(AA,0)}}{N_{(..,0)}} = \frac{600}{2000} = 0.3$$

$$f_0(Aa) = \frac{n_{(Aa,0)}}{N_{(..,0)}} = \frac{800}{2000} = 0.4$$

$$f_0(aa) = \frac{n_{(aa,0)}}{N_{(..,0)}} = \frac{600}{2000} = 0.3$$

Allelic frequency in parent generation

$$p_0 = \frac{n_{(A,0)}}{2N_{(..,0)}} = \frac{2\times600 + 800}{4000} = \frac{2000}{4000} = 0.5$$

$$q_0 = \frac{n_{(a,0)}}{n_{(.,0)}} = \frac{800 + 2\times600}{4000} = \frac{2000}{4000} = 0.5$$

Now this population will be subjected to all the preconditions of Hardy Weinberg equilibrium. Now we don't know how many progeny have been obtained in the first generation. Therefore, Let the number of progeny obtained in the first generation be $N_{(..,1)}$

Genotypic frequencies in first generation

$$f_1(AA) = \frac{n_{(AA,1)}}{N_{(..,1)}} = p_0^2 = 0.5\times0.5 = 0.25, \quad \{\text{Also, } n_{(AA,1)} = 0.25\, N_{(..,1)}\}$$

$$f_1(Aa) = \frac{n_{(Aa,1)}}{N_{(..,1)}} = 2p_0q_0 = 2\times0.5\times0.5 = 0.50, \quad \{\text{Also, } n_{(Aa,1)} = 0.5\, N_{(..,1)}\}$$

$$f_1(aa) = \frac{n_{(aa,1)}}{N_{(..,1)}} = q_0^2 = 0.5\times0.5 = 0.25, \quad \{\text{Also, } n_{(AA,1)} = 0.25\, N_{(..,1)}\}$$

Allelic frequencies:

(.. , generation number)

A	a	Genotypes	Number $n_{(.. , 1)}$	$n_{(.. , 1)} \times$ A	$n_{(.. , 1)} \times$ a
2	0	AA	$0.25\,N_{(.. , 1)}$	$2 \times 0.25\,N_{(.. , 1)}$	$0 \times 0.25\,N_{(..$
1	1	Aa	$0.50\,N_{(.. , 1)}$	$1 \times 0.50\,N_{(.. , 1)}$	$1 \times 0.50\,N_{(..$
0	2	aa	$0.25\,N_{(.. , 1)}$	$0 \times 0.25\,N_{(.. , 1)}$	$2 \times 0.25\,N_{(..$
		Sum	$N_{(.. , 1)}$	$n_{(A , 1)}$	$n_{(a , 1)}$

Now let the allelic frequency of the first allele be denoted by "p_1" and that of the second allele be denoted by "q_1" in the parent generation.

$n_{(A , 1)} = 0.50\,N_{(.. , 1)} + 0.50\,N_{(.. , 1)}$

$n_{(A , 1)} = N_{(.. , 1)}$, Similarly $n_{(a , 1)} = N_{(.. , 1)}$

$n_{(. , 1)} = n_{(A , 1)} + n_{(a , 1)} = 2N_{(.. , 1)}$

Therefore,

$$p_1 = \frac{n_{(A , 1)}}{n_{(. , 1)}} = \frac{N_{(.. , 1)}}{2N_{(.. , 1)}} = 0.5$$

$$q_1 = \frac{n_{(a , 1)}}{n_{(. , 1)}} = \frac{N_{(.. , 1)}}{2N_{(.. , 1)}} = 0.5$$

Now this population will be subjected to all the preconditions of Hardy Weinberg equilibrium. Now we don't know how many progeny have been obtained in the second generation. Therefore, Let the number of progeny obtained in the second generation be $N_{(.. , 2)}$

Genotypic frequencies in second generation

$$f_2(AA) = \frac{n_{(AA , 2)}}{N_{(.. , 2)}} = p_0^2 = 0.5 \times 0.5 = 0.25, \quad \{\text{Also, } n_{(AA , 2)} = 0.25\,N_{(.. , 2)}\}$$

$$f_2(Aa) = \frac{n_{(Aa , 2)}}{N_{(.. , 2)}} = 2p_0q_0 = 2 \times 0.5 \times 0.5 = 0.50, \quad \{\text{Also, } n_{(Aa , 2)} = 0.5\,N_{(.. , 2)}\}$$

$$f_2(aa) = \frac{n_{(aa , 2)}}{N_{(.. , 2)}} = q_0^2 = 0.5 \times 0.5 = 0.25, \quad \{\text{Also, } n_{(AA , 2)} = 0.25\,N_{(.. , 2)}\}$$

Allelic frequencies:

(.. , generation number)

A	a	Genotypes	Number $n_{(.. , 2)}$	$n_{(.. , 2)} \times$ A	$n_{(.. , 2)} \times$ a
2	0	AA	$0.25\, N_{(.. , 2)}$	$2 \times 0.25\, N_{(.. , 2)}$	$0 \times 0.25\, N_{(..$
1	1	Aa	$0.50\, N_{(.. , 2)}$	$1 \times 0.50\, N_{(.. , 2)}$	$1 \times 0.50\, N_{(..$
0	2	aa	$0.25\, N_{(.. , 2)}$	$0 \times 0.25\, N_{(.. , 2)}$	$2 \times 0.25\, N_{(..$
		Sum	$N_{(.. , 2)}$	$n_{(A , 2)}$	$n_{(a , 2)}$

Now let the allelic frequency of the first allele be denoted by "p_2" and that of the second allele be denoted by "q_2" in the parent generation.

$n_{(A , 2)} = 0.50\, N_{(.. , 2)} + 0.50\, N_{(.. , 2)}$

$n_{(A , 2)} = N_{(.. , 2)}$, Similarly $n_{(a , 2)} = N_{(.. , 2)}$

$n_{(. , 2)} = n_{(A , 2)} + n_{(a , 2)} = 2N_{(.. , 2)}$

Therefore,

$$p_2 = \frac{n_{(A , 2)}}{n_{(. , 2)}} = \frac{N_{(.. , 2)}}{2N_{(.. , 2)}} = 0.5$$

$$q_2 = \frac{n_{(a , 2)}}{n_{(. , 2)}} = \frac{N_{(.. , 2)}}{2N_{(.. , 2)}} = 0.5$$

This will continue over generations until preconditions for Hardy Weinberg equilibrium are fulfilled.

Hence,

$$f_1(AA) = f_2(AA) = \ldots\ldots\ldots\ldots f_n(AA) = 0.25$$
$$f_1(Aa) = f_2(Aa) = \ldots\ldots\ldots\ldots f_n(Aa) = 0.50$$
$$f_1(aa) = f_2(aa) = \ldots\ldots\ldots\ldots f_n(aa) = 0.25$$

Similarly,

$$p_1 = p_2 = \ldots\ldots\ldots\ldots p_n = 0.5$$
$$q_1 = q_2 = \ldots\ldots\ldots\ldots q_n = 0.5$$

Note: Since the parent generation might not be in Hardy Weinberg equilibrium, therefore following relation holds

$$f_1(AA) = f_2(AA) = \ldots\ldots\ldots\ldots f_n(AA) \neq f_0(AA)$$
$$f_1(Aa) = f_2(Aa) = \ldots\ldots\ldots\ldots f_n(Aa) \neq f_0(Aa)$$
$$f_1(aa) = f_2(aa) = \ldots\ldots\ldots\ldots f_n(aa) \neq f_0(aa)$$

However,

$p_1 = p_2 = \ldots\ldots\ldots\ldots\ldots\ldots\ldots p_n = p_0$

$q_1 = q_2 = \ldots\ldots\ldots\ldots\ldots\ldots\ldots q_n = q_0$

4. There is a unique set of genotypic frequencies for each value of allelic frequencies as depicted in figure 2. In this figure, the horizontal axis shows the two allele frequencies p and q and the vertical axis shows the expected genotype frequencies. Each line shows one of the three possible genotypes.

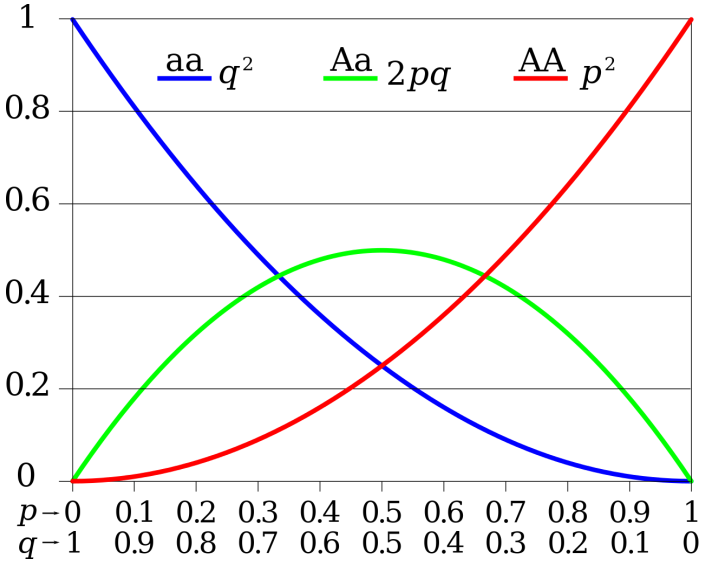

Fig 2: Genotypic frequency distribution in Hardy Weinberg equilibrium population [4]

5. The frequency of heterozygotes remains higher when allelic frequencies oscillate between 0.33 to 0.67 being maximum at 0.5.

3.3 Generalization

The simple Hardy Weinberg derivation can be generalized for more than two alleles and polyploidy using multinomial distribution. The genotype frequencies in the Hardy–Weinberg equilibrium are given by individual terms in the multinomial expansion of $(p_1 + p_2 + \ldots\ldots\ldots\ldots\ldots + p_n)^c$ for "n" alleles and "c" ploids.

Therefore different genotypic frequencies under Hardy Weinberg equilibrium will be

$$(p_1 + p_2 + \ldots\ldots\ldots\ldots\ldots + p_n)^c =$$

[4] https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle#/media/File:Hardy-Weinberg.svg

$$\sum_{k_1, k_2, \ldots\ldots, k_n \in N:\, k_1+k_2+\ldots\ldots+k_n = c} \binom{c}{k_1, k_2, \ldots\ldots, k_n} p_1^{k_1} \ldots\ldots p_n^{k_n}$$

Where, $\binom{c}{k_1, k_2, \ldots\ldots, k_n} = \frac{c!}{k_1!\, k_2!\, \ldots\ldots\, k_n!}$

Allelic frequencies can be calculated using the formula,

$$p_{(1,\ldots,n)} = \frac{1}{c} \left(\sum_{k_1, k_2, \ldots\ldots, k_n \in 1,\ldots\ldots, c:\, k_1+k_2+\ldots\ldots+k_n = c} k_{(1,\ldots,n)} \binom{c}{k_1, k_2, \ldots\ldots, k_n} p_1^{k_1} \ldots\ldots p_n^{k_n} \right)$$

To understand the above formulas, let us consider a diploid population in Hardy Weinberg equilibrium with three alleles, so n=3 and c=2.

Here, c! = 2! =2, Therefore different genotypic frequencies under Hardy Weinberg equilibrium will be

k_1	k_2	k_3	$k_1!$	$k_2!$	$k_3!$	$k_1!\, k_2!\, k_3!$	$\frac{c!}{k_1!\, k_2!\, \ldots\ldots\, k_n}$	$p_1^{k_1} p_1^{k_2} p_3^{k_3}$	Genotypes (G)	$k_1 \times$ G	$k_2 \times$ G	$k_3 \times$ G
2	0	0	2	1	1	2	1	$p_1^2 p_2^0 p_3^0$	p_1^2	$2p_1^2$	0	0
0	2	0	1	2	1	2	1	$p_1^0 p_2^2 p_3^0$	p_2^2	0	$2p_2^2$	0
0	0	2	1	1	2	2	1	$p_1^0 p_2^0 p_3^2$	p_3^2	0	0	$2p_3^2$
1	1	0	1	1	1	1	2	$p_1^1 p_2^1 p_3^0$	$2p_1p_2$	p_1p_2	p_1p_2	0
1	0	1	1	1	1	1	2	$p_1^1 p_2^0 p_3^1$	$2p_1p_3$	p_1p_3	0	p_1p_3
0	1	1	1	1	1	1	2	$p_1^0 p_2^1 p_3^1$	$2p_2p_3$	0	p_2p_3	p_2p_3

Therefore,

$$(p_1 + p_2 + p_3)^2 = p_1^2 + p_2^2 + p_3^2 + 2p_1p_2 + 2p_1p_3 + 2p_2p_3$$

Allelic frequencies

$$p_1 = \frac{1}{2}\left(2p_1^2 + p_1p_2 + p_1p_3\right)$$

$$p_2 = \frac{1}{2}\left(2p_2^2 + p_1p_2 + p_2p_3\right)$$

$$p_3 = \frac{1}{2}\left(2p_3^2 + p_1p_3 + p_2p_3\right)$$

Similarly, this generalization can be extended to any number of alleles with different ploidy levels.

Illustration 1: Derive an expression for calculation of gene and genotypic frequencies for a locus determining ABO blood groups in a Hardy Weinberg population?

Let the proportion of individuals in the Hardy Weinberg population with blood groups A, B, O and AB be w, x, y and z.

Let, p_1, p_2 *and* p_3 be the allelic frequencies of three alleles and as per the previously established generalization, p_1^2, p_2^2, p_3^2, $2p_1p_2$, $2p_1p_3$ and $2p_2p_3$ will be the genotypic frequencies. Therefore as per our previous knowledge of gene action in ABO blood group system

$p_1^2 + 2p_1p_3 = w$ ....(3)

$p_2^2 + 2p_2p_3 = x$ ....(4)

$p_3^2 = y$ ....(5)

$2p_1p_2 = z$ ....(6)

From equation, 5, $p_3 = \sqrt{y}$

From equation, 6, $p_2 = \frac{z}{2p_1}$

Putting these values in eqn 4 we get, $\frac{z^2}{4p_1^2} + \frac{2 \times z \times \sqrt{y}}{2p_1} = x$

On solving we get, $p_1^2 - \frac{p_1 z \sqrt{y}}{x} = \frac{z^2}{4x}$ ....(7)

Also putting the value of p_3 in equation 3 we get

$p_1^2 + 2p_1\sqrt{y} = w$ ....(8)

On subtracting equation 7 from equation 8 we get,

$2p_1\sqrt{y} + \frac{p_1 z \sqrt{y}}{x} = w - \frac{z^2}{4x}$

$\frac{p_1\sqrt{y}(2x+z)}{x} = \frac{4wx - z^2}{4x}$

$p_1 = \frac{wx - (\frac{z}{2})^2}{\sqrt{y}(2x+z)}$

Putting the value of p_1 in equation 6 we get,

$$p_2 = \frac{z\sqrt{\bar{y}}(2x+z)}{2(wx-(\frac{z}{2})^2)}$$

Therefore all the gene and genotypic frequencies can very well be calculated if we know the proportion of individuals belonging to each blood group phenotypically.

Example 2: Calculate the gene and genotypic frequencies for a locus determining ABO blood groups in a Hardy Weinberg population with proportion of individuals having blood groups A, B, O and AB being 450, 210, 40 and 300 respectively.

Solution 2: Here total population = 450+210+40+300 = 1000

The allelic frequency of three alleles as per the illustration will be

$$p_1 = \frac{wx-(\frac{z}{2})^2}{\sqrt{\bar{y}}(2x+z)}\ ,\ p_3 = \sqrt{\bar{y}} \text{ and } p_2 = 1 - (p_1 + p_3)$$

Where, w= 400/1000 = 0.4,
x= 210/1000 =0.21,
y= 40/1000= 0.04
and z=300/1000 = 0.3
Now putting the values of w, x, y and z

$$p_1 = \frac{0.0945-0.0225}{0.2(2\times0.21+0.3)} = 0.5,$$

$$p_3 = \sqrt{\bar{y}} = \sqrt{0.04} = 0.2$$

$$p_2 = 1 - (0.5+ 0.2) = 0.3$$

The genotypic frequencies will be

$$p_1^2 = 0.25,\ p_2^2 = 0.09,\ p_3^2 = 0.04,\ 2p_1p_2 = 0.3,\ 2p_1p_3 = 0.2 \text{ and } 2p_2p_3 = 0.12$$

3.4 Application

The basic application of Hardy Weinberg law resides in the estimation of genetic structure of present, past and future generations. one can find application of Hardy Weinberg law in many more situations which are beyond the coverage under present lecture notes. The Hardy Weinberg law can be applied in the following major ways in the subject of animal genetics and breeding.

3.4.1 Population genetic structure in different modes of inheritance

The Hardy Weinberg law can very well be applied for estimation of allelic and genotypic frequencies under different modes of inheritance using simple algebraic expressions on part population data. This can very well be understood by the following examples given below.

A. Complete dominance under single locus with two alleles

Let the frequency of the dominant allele be "p" and that of the recessive allele be "q". The frequency of recessive allele can be estimated from the genotypic frequency of recessive individuals (say aa) in a Hardy Weinberg population.

$$q = \sqrt{f(aa)}$$

The frequency of the dominant allele is then estimated as

$$p = 1-q$$

Once the allelic frequencies are known, rest of the genotypic frequencies can easily be calculated as under

$$f(AA) = p^2 \text{ and } f(Aa) = 2pq$$

Example 3: The American Hairless Terrier is a breed that includes hairless and coated varieties. The breed's characteristic hairlessness is inherited as a recessive trait. The American Hairless Terrier is usually hairless but some individuals in the breed may be coated. A random mating population of such contain 90.25% hairless dogs. Enumerate the genetic structure of American Hairless Terrier dog population?

Fig 3: American Hairless Terrier[5]

Solution 3: Allelic frequencies

Here, $f(aa) = \frac{90.25}{100} = 0.9025$

Therefore, $q = \sqrt{f(aa)} \Rightarrow \sqrt{0.9025} = 0.95$, Hence, $p = 1-q \Rightarrow 1-0.95 = 0.05$

Genotypic frequencies

$$f(AA) = p^2 \Rightarrow 0.05 \times 0.05 = 0.0025$$

$$f(Aa) = 2pq \Rightarrow 2 \times 0.05 \times 0.95 = 0.095$$

Dominant individuals are 0.25% and heterozygotes are 9.5% in the population.

[5] Photo courtesy: https://www.pinterest.com/pin/4574037107926807/

B. Codominant alleles with lethal recessives

Illustration 2: Burmese head defect (BHD), or congenital frontonasal dysplasia, is inherited as an autosomal codominant condition characterized by improper development and subsequent malformation of the skull and facial features. Recessive kittens with BHD may be stillborn or born live, but kittens born live cannot survive for long and require euthanasia. Whereas heterozygous genotype may have shortened facial structure (brachycephaly), but will not have Burmese head defect. Similar conditions are also found in other cat breeds like Australian Mist, Bombay, Burmilla, Tonkinese etc.

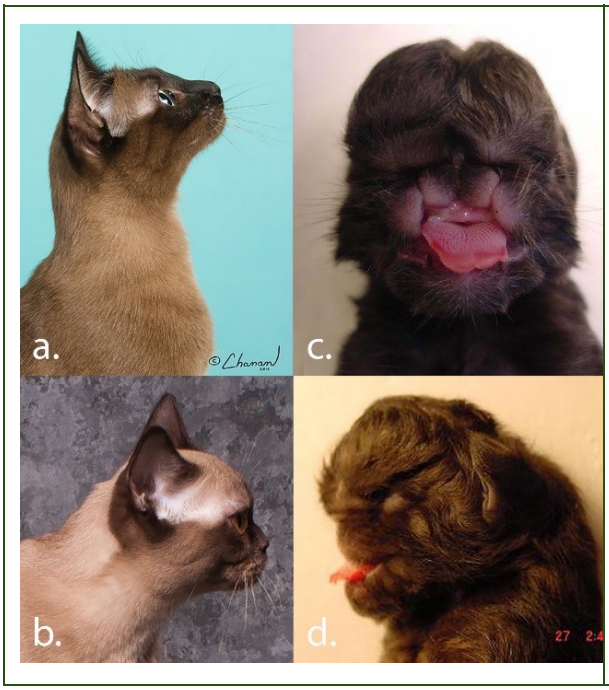

Fig 4: a. Head conformation of a traditional Burmese cat. **b.** the contemporary style with extreme brachycephalic conformation. **c.** and **d.** frontal and lateral view, respectively of kitten with Burmese head defect, showing abnormal development of the maxillary processes and ocular and nasal degeneration. Photographs are courtesy of Nancy Reeves, Isabelle Marchand and Richard Katris – Chanan Photography and are reproduced from Lyons et al (2016)[6] under the Creative Commons licence (CC BY-NC-ND)

It is quite clear from the mode of inheritance that the individuals with recessive genotypes had stillborn or died early in their life in this population.

Therefore, genotype pool will be constituted by

$p^2 + 2pq$

Where, Genotypic frequency of homozygous dominant = p^2

and that of heterozygotes = 2pq, being a codominant trait

Frequency of brachycephalic cats = $\frac{Genotypic\ frequency\ of\ heterozygotes}{p^2+2pq} = \frac{2pq}{p^2+2pq}$

[6] Lyons LA, Erdman CA, Grahn RA, Hamilton MJ, Carter MJ, Helps CR, Alhaddad H and Gandolfi B (2016) Aristaless-Like Homeobox protein 1 (ALX1) variant associated with craniofacial structure and frontonasal dysplasia in Burmese cats. *Developmental Biology 409:* 451–8

Example 4: A sample population of cats were screened phenotypically for brachycephalic condition. In this sample 33.33% cats were found brachycephalic. How much progeny of this cat population will expectedly die because of BHD provided the population is in Hardy Weinberg Population ?

Solution 4: According to given conditions

$$\frac{2pq}{p(p+2q)} = 0.33 \Rightarrow \frac{2pq}{p(1+q)} = 0.33 \Rightarrow \frac{2q}{1+q} = 0.33$$

$$\Rightarrow q = \frac{0.33}{1.34} \Rightarrow q = 0.2, \text{ Therefore, } p = 1\text{-}0.2 = 0.8$$

Therefore frequency of future progeny that will suffering from BHD = q^2

$= 0.2 \times 0.2$

$= 0.04$

Hence around 4% progeny in this population will either be stillborn or need to be euthanized early in their life.

C. Multiple alleles with pleiotropism

Illustration 3: Dominant white and white spotting in domestic cats[7]

Fig 5: White and white spotted cats[8]

The dominant white (W) and white spotting (w^S) mutations that cause white patterning in cats is the result of three alleles W, w^S and w^+ allele. The order of dominance for these alleles is W > w^S > w^+. The dominant allele W also has pleiotropic effect and cats bearing this allele also suffer from some degree of hearing impairment. Considering the allelic frequencies of three alleles as p, q and r. The various phenotypes observed as a combination of these alleles and their respective

[7] Illustration adapted from David, et al. (2014). Endogenous Retrovirus Insertion in the KIT Oncogene Determines White and White spotting in Domestic Cats. *G3: Genes|Genomes|Genetics, 4*(10), 1881-1891. doi: 10.1534/g3.114.013425

[8] Photo courtesy: https://vgl.ucdavis.edu/sites/g/files/dgvnsk8836/files/inline-images/Cat-Dominant-White-and-White-Spotting.jpg

genotypic frequencies will be as follows.

S.No.	Genotype	Genotypic frequencies	Colour	Hearing impairment
1.	WW	p^2	White	Yes
2.	Ww^S	2pq		
3.	Ww^+	2pr		
4.	w^Sw^S	q^2	White spotted	No
5.	w^Sw^+	2qr		
6.	w^+w^+	r^2	Normal	

If the proportion of cats with hearing impairment were "x" and those with normal hearing and colour were "z". We can easily calculate the allelic frequencies in this population.

The allelic frequency for w^+

$r^2 = z$

$r = \sqrt{z}$

Also let the proportion of white spotted individuals be "y", therefore x + y + z = 1

$$q^2 + 2qr = y$$

Substituting the value of "r" in this equation,

$$q^2 + 2q\sqrt{z} - y = 0$$

This is a form of quadratic equation, $ax^2 + bx + c = 0$ where $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$

Since $q \geq 0$ always, therefore $q = \frac{-2\sqrt{z} + \sqrt{4z+4y}}{2}$

$q = \frac{-2\sqrt{z} + \sqrt{4z+4y}}{2} \Rightarrow -\sqrt{z} + \sqrt{z+y}$

Also, p+q+r =1,

$p = 1-(q+r)$

$= 1-(-\sqrt{z} + \sqrt{z+y} + \sqrt{z})$

$= 1-\sqrt{z+y}$

Therefore, $p = 1-\sqrt{z+y}$, $q = -\sqrt{z} + \sqrt{z+y}$ and $r = \sqrt{z}$

Example 4: In a cat population if white cats were 19% and those with normal colour were 49% ? What would be the allelic frequency for three alleles as described in illustration 3?

Solution 4: Let the allelic frequencies for three alleles be p, q and r

According to given conditions

Proportion of white cats (x) = 0.19

Proportion of normal cats (z) = 0.49
Therefore proportion of white spotted cats (y) would have been
= 1 - 0.19- 0.49
= 0.32
Therefore as per illustration 3,
$p = 1 - \sqrt{z + y}$
$= 1 - \sqrt{0.49 + 0.32}$
$= 1 - 0.9$
$= 0.1$
$r = \sqrt{z}$
$= 0.7$
$q = 1 - (p+r)$
$= 0.2$
Hence the allelic frequency of three alleles in this Hardy Weinberg population would have been 0.1, 0.2 and 0.7.

Similarly, under several other combinations of part population data and modes of inheritance, prediction of genetic structure of the whole population can be undertaken using Hardy Weinberg law.

3.4.2 Selection experiments

Hardy Weinberg law is basically applied in the control population in a breeding experiment. In control population or group animals are allowed to mate randomly and thereby as per Hardy Weinberg law the gene and genotypic frequencies at various loci remains the same from one generation to another. Whereas in other treatment groups the mating is planned as per breeding program. All other fixed causal factors in the breeding experiment are tried to be kept constant as much as possible. Therefore any significant differences observed in treatment populations compared to that of the control population can majorly be attributed to the outcome of the breeding program.

3.4.3 Influence of evolutionary forces on population

Another direct application of Hardy Weinberg law is to determine the influence of evolutionary influences on the population under study. If gene and genotypic frequencies in a population do not differ significantly from one generation to another therefore it can be inferred that evolutionary influences do not have

significant effect on the genetic stability of the population. The quantification of these influences will be undertaken in forthcoming lectures.

3.4.4 Genetic counselling

If the allelic frequency for a trait in the population is known then this information can very well be used for genetic counselling. The likelihood of the progeny born with genetic disease can be ascertained prior to the marriage of a person in a given population. We will try to understand the utility of Hardy Weinberg law for genetic counselling using an example.

Example 5: Phenylketonuria (PKU) is an autosomal recessive metabolic disorder that results in mental retardation if untreated during the newborn period. In the United States, one out of 10,000 babies is born with PKU[9]. What is the risk for a female carrier to have a fetus affected with PKU?

Solution 5: Here in the given example
The frequency of recessive homozygous individuals will be

$q^2 = \frac{1}{10000} \Rightarrow q = \frac{1}{100}$

$p = 1-q \Rightarrow 1- \frac{1}{100} = \frac{99}{100}$

Therefore frequency of carriers in the population will be 2pq

$= 2 \times \frac{1}{100} \times \frac{99}{100}$

$= \frac{99}{5000}$

Now probability that a female carrier will marry a carrier male = $\frac{99}{5000}$

Now probability that fetus will be getting recessive allele from female P(A) = $\frac{1}{2}$

Similarly, probability that fetus will be getting recessive allele from male P(B) =

$\frac{1}{2} \times \frac{99}{5000}$

Therefore, the probability that the fetus will suffer from PKU will be P(AB) = P(A). P(B)

[9] Smith I, Cook B, Beasley M. Review of neonatal screening programme for phenylketonuria. BMJ 1991;13:333-5.

$$= \frac{1}{2} \times \frac{1}{2} \times \frac{99}{5000}$$

$$= \frac{99}{20000}$$

$$= 0.00495$$

Therefore there are 0.459% chances that this female will give birth to a child suffering from PKU if she chooses to marry at random without screening of her male partner.

3.4.5 Detecting genotyping errors in large-scale genotyping studies

Genotyping data sets may contain errors that, in some instances, lead to false conclusions. Deviation from Hardy–Weinberg equilibrium (HWE) in random samples may be indicative of problematic assays. Genotyping data sets generally focuses on Single-nucleotide polymorphism (SNP). In genetics, a single-nucleotide polymorphism is a substitution of a single nucleotide at a specific position in the genome that is present in a sufficiently large fraction of the population (e.g. 1% or more).

A SNP in which both forms lead to the same polypeptide sequence is termed synonymous (sometimes called a silent mutation) — if a different polypeptide sequence is produced they are nonsynonymous. A nonsynonymous change may either be missense or nonsense, where a missense change results in a different amino acid, while a nonsense change results in a premature stop codon. SNPs that are not in protein-coding regions may still have consequences for gene splicing, transcription factor binding, or the sequence of non-coding ribonucleic acid (RNA).

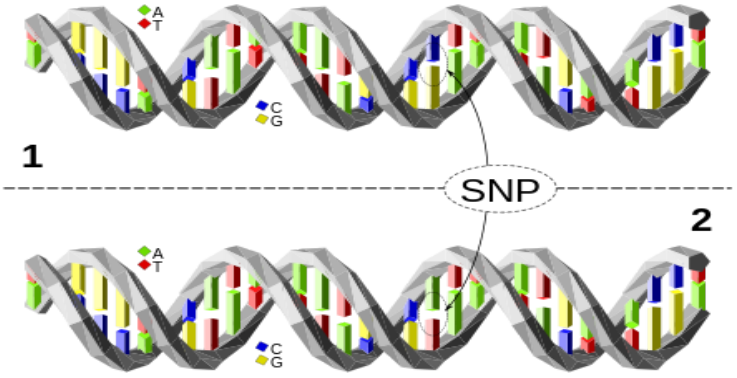

Fig 6: Single-nucleotide polymorphism[10]

[10] Image credits: David Eccles (gringer) retrieved online https://commons.wikimedia.org/wiki/File:Dna-SNP.svg for fair use under the Creative Commons Attribution 4.0 International license

Illustration 4: Detection of genotyping errors by HWE testing[11]

In a study conducted in 2004, a subset of 313 SNPs, whose minor allele frequencies were >5%, was selected for the estimation of deviation from HWE. A total of 36 SNPs (11.5%), with minor allele frequencies ranging between 0.06 and 0.49, were found to deviate from HWE (P<0.05).

Possible explanations for SNPs that showed deviation from HWE were explored. An SNP assay was classified as 'nonspecific' if a primer and/or probe set showed 100% homology with multiple regions in the genome. Identifiable reasons for deviation from HWE in 36 SNPs reported had been tabulated by authors as given below.

Reasons for deviation	N	%
Genotyping error	21	58
Nonspecific	5	14
Unknown	10	28
Total	36	

Thus application of Hardy Weinberg equilibrium in large sequencing data was proved useful in detecting genotyping errors.

3.5 Sex Linkage

We have already understood how an autosomal locus behaves in a population under Hardy-Weinberg equilibrium. We will now extend our knowledge to allosomal locus. So we will try to understand how autosomal locus differs from allosomal locus with respect to principles of Hardy Weinberg equilibrium. The initial allelic frequencies in two sexes can be calculated using the methodology as described in section 3.1 of this lecture notes. Consider that the initial population is not under Hardy Weinberg equilibrium. Therefore, $p_{f0} \neq p_{m0}$ and $q_{f0} \neq q_{m0}$. Although, $p_{f0} + q_{f0} = 1$ and $p_{m0} + q_{m0} = 1$. Mating structure of an allosomal locus having two alleles can be tabulated as under. Blue boxes represent the genotypic frequency of the female population after one random mating. Similarly green boxes represent the genotypic frequency of the male population after one random mating.

(frequency)

[11] Hosking, L., Lumsden, S., Lewis, K. et al. Detection of genotyping errors by Hardy–Weinberg equilibrium testing. Eur J Hum Genet 12, 395–399 (2004). https://doi.org/10.1038/sj.ejhg.5201164

			Male gametes		
	Gamete types		X ($\frac{1}{2}$)		Y($\frac{1}{2}$)
		Allele frequency	$\frac{1}{2}\ p_{m0}$	$\frac{1}{2}\ q_{m0}$	$\frac{1}{2}$
Female gametes	X(1)	p_{f0}	$\frac{1}{2}\ p_{m0}\ p_{f0}$	$\frac{1}{2}\ q_{m0}\ p_{f0}$	$\frac{1}{2}\ p_{f0}$
		q_{f0}	$\frac{1}{2}\ p_{m0}\ q_{f0}$	$\frac{1}{2}\ q_{m0}\ q_{f0}$	$\frac{1}{2}\ q_{f0}$

3.5.1 Genotypic frequencies

The genotypic frequencies in the first generation can easily be calculated for both the sexes separately using the above table .

Female population

f (Homozygous dominant) = $\frac{1}{2}\ p_{m0}\ p_{f0}$

f (Heterozygous) = $\frac{1}{2}\ (q_{m0}\ p_{f0} + p_{m0}\ q_{f0})$

f (Recessive homozygous) = $\frac{1}{2}\ q_{m0}\ q_{f0}$

Cumulative frequency = $\frac{1}{2}\ p_{m0}\ p_{f0} + \frac{1}{2}\ (q_{m0}\ p_{f0} + p_{m0}\ q_{f0}) + \frac{1}{2}\ q_{m0}\ q_{f0}$

$= \frac{1}{2}\ [\ p_{f0}\ (p_{m0} + q_{m0}) + q_{f0}\ (p_{m0} + q_{m0})\ ]$

$= \frac{1}{2}\ [\ (p_{f0} + q_{f0})\ (p_{m0} + q_{m0})\ ]$

$= \frac{1}{2}$ ...(Since, $p_{f0} + q_{f0} = 1$ and $p_{m0} + q_{m0} = 1$)

Male population

f (Dominant population) = $\frac{1}{2}\ p_{f0}$

f (Recessive population) = $\frac{1}{2}\ q_{f0}$

Cumulative frequency = $\frac{1}{2}\ (p_{f0} + q_{f0})$

$= \frac{1}{2}$ ...(Since, $p_{f0} + q_{f0} = 1$)

Therefore cumulative frequencies for all the genotypes remained $\frac{1}{2}$ in both the sexes.

3.5.1 Allelic frequencies

Female population

Now, frequency of dominant allele in first generation of female population after random mating will be

$$= \frac{\frac{1}{2}[p_{m0}p_{f0}+\frac{1}{2}\,q_{m0}p_{f0}+\frac{1}{2}\,p_{m0}q_{f0}]}{\frac{1}{2}[p_{m0}p_{f0}+q_{m0}p_{f0}+p_{m0}q_{f0}+q_{m0}q_{f0}]} \Rightarrow \frac{p_{m0}p_{f0}+\frac{1}{2}\,q_{m0}p_{f0}+\frac{1}{2}\,p_{m0}q_{f0}}{(p_{m0}+q_{m0})(p_{f0}+q_{f0})}$$

$= p_{m0}p_{f0} + \frac{1}{2}\,q_{m0}p_{f0} + \frac{1}{2}\,p_{m0}q_{f0}$ ...(Since, $p_{f0} + q_{f0} = 1$ and $p_{m0} + q_{m0} = 1$)

$= p_{f0}\,(p_{m0} + \frac{1}{2}\,q_{m0}) + \frac{1}{2}\,p_{m0}q_{f0}$

$= p_{f0}\,(p_{m0} + \frac{1}{2}\,q_{m0} + \frac{1}{2}\,q_{m0} - \frac{1}{2}\,q_{m0}) + \frac{1}{2}\,p_{m0}(1 - p_{f0})$ ...(Since, $p_{f0} + q_{f0} = 1$)

$= p_{f0}\,(p_{m0} + q_{m0} - \frac{1}{2}\,q_{m0}) + \frac{1}{2}\,p_{m0} - \frac{1}{2}\,p_{m0}p_{f0}$

$= p_{f0}\,(1 - \frac{1}{2}\,q_{m0}) + \frac{1}{2}\,p_{m0} - \frac{1}{2}\,p_{m0}p_{f0}$ ...(Since, $p_{m0} + q_{m0} = 1$)

$= p_{f0} - \frac{1}{2}p_{f0}\,q_{m0} - \frac{1}{2}\,p_{m0}p_{f0} + \frac{1}{2}\,p_{m0}$

$= p_{f0} - \frac{1}{2}p_{f0}(\,q_{m0} + p_{m0}) + \frac{1}{2}\,p_{m0}$

$= p_{f0} - \frac{1}{2}p_{f0} + \frac{1}{2}\,p_{m0}$ ...(Since, $p_{m0} + q_{m0} = 1$)

$= \frac{1}{2}\,(\,p_{f0} + p_{m0})$

Similarly, it can be proved that,

$q_{f1} = \frac{1}{2}\,(\,q_{f0} + q_{m0})$

Therefore it can be observed that the frequency of dominant and recessive allele in successive generations comes out to be average of respective frequencies of both sexes in their previous generation.

Male population

Now, frequency of dominant allele p_{m1} in first generation of male population will be

$$p_{m1} = \frac{\frac{1}{2}p_{f0}}{\frac{1}{2}(p_{f0} + q_{f0})} = p_{f0} \quad \text{...(Since, } p_{f0} + q_{f0} = 1)$$

Similarly, $q_{m1} = \frac{\frac{1}{2}q_{f0}}{\frac{1}{2}(p_{f0} + q_{f0})} = q_{f0}$ ...(Since, $p_{f0} + q_{f0} = 1$)

Thus, it can be observed that the frequency of dominant and recessive alleles follows a criss cross pattern in subsequent generations.

Also,

$$q_{f1} - q_{m1} = \frac{1}{2}(q_{f0} + q_{m0}) - q_{f0}$$

$$= \frac{1}{2}q_{m0} - \frac{1}{2}q_{f0}$$

$$= \frac{1}{2}(q_{m0} - q_{f0})$$

Therefore it can be established from the above derivation that the difference between allelic frequency of successive generation in male and female population becomes half the difference between the allelic frequency of the female and male (reverse order) population in the previous generation.

3.5.2 Generalization for allelic frequency

The allelic frequencies for recessive alleles in successive generations can be generalized for two sexes in the following manner.

Female population,

Let q_{f0} be the frequency of recessive allele in parental generation. This q_{f0} can be

rewritten as

$q_{f0} = \frac{1}{1} q_{f0} + \frac{0}{1} q_{m0}$

As proved earlier

$q_{f1} = \frac{1}{2} (q_{f0} + q_{m0})$

Similarly in second generation,

$q_{f2} = \frac{1}{2} (q_{f1} + q_{m1})$

$= \frac{1}{2} [\frac{1}{2} (q_{f0} + q_{m0}) + q_{f0}]$ ...(Since, $q_{f1} = \frac{1}{2} (q_{f0} + q_{m0})$ and $q_{m1} = q_{f0}$

$= \frac{1}{2} [\frac{3}{2} q_{f0} + \frac{1}{2} q_{m0}]$

$= \frac{3}{4} q_{f0} + \frac{1}{4} q_{m0}$

Similarly in third generation,

$q_{f3} = \frac{1}{2} (q_{f2} + q_{m2})$

$= \frac{1}{2} [\frac{3}{4} q_{f0} + \frac{1}{4} q_{m0} + q_{f1}]$ ...Since, $q_{f2} = \frac{3}{4} q_{f0} + \frac{1}{4} q_{m0}$ and $q_{m2} = q_{f1}$

$= \frac{1}{2} [\frac{3}{4} q_{f0} + \frac{1}{4} q_{m0} + \frac{1}{2} (q_{f0} + q_{m0})]$ ...Since, $q_{f1} = \frac{1}{2} (q_{f0} + q_{m0})$

$= \frac{1}{2} [\frac{5}{4} q_{f0} + \frac{3}{4} q_{m0}]$

$= \frac{5}{8} q_{f0} + \frac{3}{8} q_{m0}$

Similarly in fourth generation,

$q_{f4} = \frac{1}{2} (q_{f3} + q_{m3})$

$= \frac{1}{2} [\frac{5}{8} q_{f0} + \frac{3}{8} q_{m0} + q_{f2}]$ ...Since, $q_{f3} = \frac{5}{8} q_{f0} + \frac{3}{8} q_{m0}$ and $q_{m3} = q_{f2}$

$= \frac{1}{2}[\ \frac{5}{8}\ q_{f0} + \frac{3}{8} q_{m0}\ +\ \frac{3}{4}\ q_{f0} + \frac{1}{4} q_{m0}\ ]$ ...Since, $q_{f2} = \frac{3}{4}\ q_{f0} + \frac{1}{4} q_{m0}$

$= \frac{1}{2}[\ \frac{11}{8}\ q_{f0} + \frac{5}{8} q_{m0}]$

$= \frac{11}{16}\ q_{f0} + \frac{5}{16} q_{m0}$

Similarly in fifth generation,

$q_{f5} = \frac{1}{2}\ (\ q_{f4} + q_{m4})$

$= \frac{1}{2}[\ \frac{11}{16}\ q_{f0} + \frac{5}{16} q_{m0}\ +\ q_{f3}\ ]$ ...Since, $q_{f4} = \frac{11}{16}\ q_{f0} + \frac{5}{16} q_{m0}$ and $q_{m4} = q_{f3}$

$= \frac{1}{2}[\ \frac{11}{16}\ q_{f0} + \frac{5}{16} q_{m0}\ +\ \frac{5}{8}\ q_{f0} + \frac{3}{8} q_{m0}\ ]$ ...Since, $q_{f3} = \frac{5}{8}\ q_{f0} + \frac{3}{8} q_{m0}$

$= \frac{1}{2}[\ \frac{21}{16}\ q_{f0} + \frac{11}{16} q_{m0}]$

$= \frac{21}{32}\ q_{f0} + \frac{11}{32} q_{m0}$

These results can be summarized in the following tabular form.

Generation	**Multiple of q_{f0}**		**Multiple of q_{m0}**	
	Numerator	**Denominator**	**Numerator**	**Denominator**
0	1	1	0	1
1	1	2	1	2
2	3	4	1	4
3	5	8	3	8
4	11	16	5	16
5	21	32	11	32
.	.	.	.	.
.	.	.	.	.
t	$\frac{2^{t+1}-(-1)^{t+1}}{3}$	2^t	$\frac{2^t-(-1)^t}{3}$	2^t

The multiples of q_{f0} and q_{m0} in each successive generation follows a typical series in the following sequence 0, 1, 1, 3, 5, 11, 21 had been observed. These are called

Jacobsthal numbers. In mathematics, the Jacobsthal numbers are an integer sequence named after the German mathematician Ernst Jacobsthal. In simple terms, the sequence starts with 0 and 1, then each following number is found by adding the number before it to twice the number before that. The Jacobsthal number at a specific point in the sequence may be calculated directly using the closed-form equation.[12]

$$J_n = \frac{2^n - (-1)^n}{3}$$

Hence frequency of recessive sex linked allele after t generations in Hardy Weinberg equilibrium can be deduced to

$$q_{ft} = \frac{2^{t+1} - (-1)^{t+1}}{3 \times 2^t} q_{f0} + \frac{2^t - (-1)^t}{3 \times 2^t} q_{m0}$$

Male population

Let q_{m0} be the frequency of recessive allele in parental generation. As proved earlier

$q_{m1} = q_{f0}$

Similarly in second generation,

$q_{m2} = q_{f1}$

$= \frac{1}{2} (q_{f0} + q_{m0})$ ...Since, $q_{f1} = \frac{1}{2} (q_{f0} + q_{m0})$

Similarly in third generation,

$q_{m3} = q_{f2}$

$= \frac{3}{4} q_{f0} + \frac{1}{4} q_{m0}$

Therefore after t generations, where t > 0

$$q_{mt} = \frac{2^t - (-1)^t}{3 \times 2^{t-1}} q_{f0} + \frac{2^{t-1} - (-1)^{t-1}}{3 \times 2^{t-1}} q_{m0}$$

[12] Sloane, N. J. A. (ed.). "Sequence A014551 (Jacobsthal–Lucas numbers)". *The On-Line Encyclopedia of Integer Sequences*. OEIS Foundation.

Example 6: The allelic frequency of recessive allele in a large female and male population was found to be 0.8 and 0.3 respectively? The population was allowed to mate at random. What will be the resultant allelic frequency of recessive allele in female and male population after three generations of random mating?

Solution 6: We know that the generalized formula for finding allelic frequency of recessive allele in female population after 't' generations of random mating is given by

$$q_{ft} = \frac{2^{t+1}-(-1)^{t+1}}{3\times 2^{t}}\ q_{f0} + \frac{2^{t}-(-1)^{t}}{3\times 2^{t}}\ q_{m0}$$

$$= \frac{2^{3+1}-(-1)^{3+1}}{3\times 2^{3}}\ 0.8 + \frac{2^{3}-(-1)^{3}}{3\times 2^{3}}\ 0.3$$

$$= \frac{16-1}{24} \times 0.8 + \frac{8+1}{24} \times 0.3$$

$$= \frac{12+2.7}{24}$$

$$= \frac{14.7}{24}$$

$$q_{f3} = 0.6125$$

Similarly for males,

$$q_{mt} = \frac{2^{t}-(-1)^{t}}{3\times 2^{t-1}}\ q_{f0} + \frac{2^{t-1}-(-1)^{t-1}}{3\times 2^{t-1}}\ q_{m0}$$

$$q_{m3} = \frac{2^{3}-(-1)^{3}}{3\times 2^{3-1}} \times 0.8 + \frac{2^{3-1}-(-1)^{3-1}}{3\times 2^{3-1}} \times 0.3$$

$$= \frac{9}{12} \times 0.8 + \frac{3}{12} \times 0.3$$

$$= \frac{7.2+0.9}{12}$$

$$= \frac{8.1}{12}$$

$$q_{m3} = 0.675$$

3.6 Ternary plot for Hardy Weinberg Equation

Ternary plot used in population genetics to depict Hardy Weinberg equation is called a de Finetti diagram. It is named after the Italian statistician Bruno de Finetti (1906–1985). This graph is used to draw the genotype frequencies of a diploid population with two alleles. Geometrically, de Finetti diagram is based on an equilateral triangle, and Viviani's theorem. This theorem states that the sum of the perpendicular distances from any interior point to the sides of said triangle is a

constant equal to the length of the triangle's altitude.

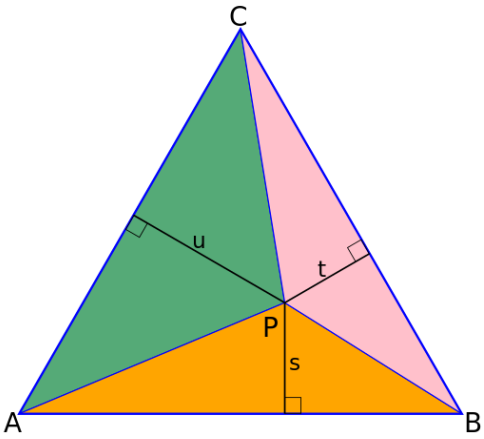

Fig 7: Geometry describing Viviani's theorem[13]

In fig 7, ABC is an equilateral triangle and P is any random point within the triangle at distances of from the three sides AB, BC and CA respectively.
As per Viviani's theorem,

Height (Altitude) of the equilateral triangle = s + t + u

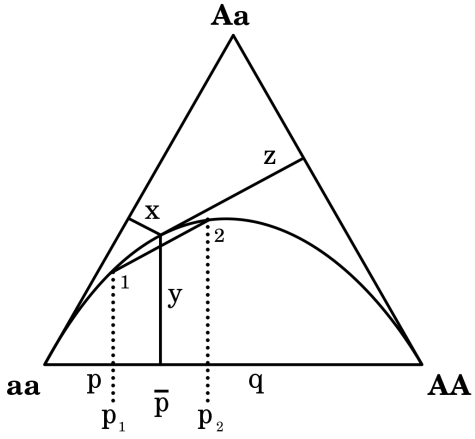

Fig 8: de Finetti diagram[14]

A ternary plot or de Finetti diagram is a barycentric plot on three variables which sum to a constant. It graphically depicts the ratios of the three variables as positions in an equilateral triangle. The ternary plot therefore enables the user to represent three variables in a two-dimensional graph. In population genetics ternary plots have traditionally been used to show the changes in genotypic frequencies of systems with

[13] Photo Courtesy: Krishnavedala, Geometry describing Viviani's theorem. Shared under CC0 1.0 Universal (CC0 1.0) Public Domain Dedication https://commons.wikimedia.org/wiki/File:Viviani_Theorem.svg

[14] Photo Courtesy: Beao, de Finetti diagram. Shared under the Creative Commons Attribution-Share Alike 2.5 Generic license.https://commons.wikimedia.org/wiki/File:De_Finetti_diagram.svg

one locus and two alleles (resulting in three genotypes)[15]. The genotypic frequency is 100% in each corner of the triangle and 0% at the opposite line. The percentage of a specific genotype decreases linearly with increasing distance from this corner. The curved line is the expected Hardy–Weinberg frequency as a function of p. The easiest way to determine the composition is to set the altitude of the triangle to 100% and determine the shortest distances from the point of interest to each of the three sides. The distances (the ratios of the distances to the total height of 100%) give the probability of each of the genotypes as per Viviani's theorem.

Therefore, from fig 8,

$$x = f(AA),\ y = f(Aa) \text{ and } z = f(aa)$$

3.7 Testing for Hardy Weinberg Equilibrium

There are several statistical procedures available that test the null hypothesis of Hardy-Weinberg equilibrium. We will discuss here the $\chi 2$ test, the $\chi 2$ test with continuity correction and Fisher's exact test.

3.7.1 $\chi 2$ test for Hardy Weinberg equilibrium

The χ^2 test for a population in Hardy Weinberg equilibrium can further be generalized for locus with two alleles and those with multiple alleles.

3.7.1.1. Locus with two alleles

Now, recall χ^2 test for 2 x 2 contingency table. The contingency table of 2 x 2 fold is given as follows

	Group 3	Group 4	Sum
Group 1	a	b	a+b
Group 2	c	d	c+d
Sum	a+c	b+d	n=a+b+c+d

Where total number of observations, n = a + b + c + d
$\chi 2$ statistic may be calculated with 1 degrees of freedom as

[15] Cannings C., Edwards A.W.F. Natural selection and the de Finetti diagram. Ann Hum Genet, 31 (1968), pp. 421-428

$$\chi 2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Let us consider a simple situation for a locus with two alleles and three genotypes to understand χ^2 testing for Hardy Weinberg equilibrium. Let D, H and R be the number of homozygous dominant, heterozygous and homozygous recessive individuals in the population.

The generalized contingency table of 2 x 2 fold can easily be compared to genotypic frequencies in a population.

		Male population allelic contribution	
		Dominant	**Recessive**
Female population allelic contribution	**Dominant**	**D**	**$\frac{1}{2}$ H**
	Recessive	**$\frac{1}{2}$ H**	**R**

Therefore from the two contingency tables it can be deduced that,

a = D , b = c = $\frac{1}{2}$ H and d = R

Substituting these values in formula for $\chi 2$ statistic we get,

$$\chi 2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$= \frac{n(DR-\frac{1}{4}H^2)^2}{(D+\frac{1}{2}H)(R+\frac{1}{2}H)(D+\frac{1}{2}H)(R+\frac{1}{2}H)}$$

$$= \frac{n(DR-\frac{1}{4}H^2)^2}{(D+\frac{1}{2}H)^2(R+\frac{1}{2}H)^2}$$

$$= \frac{n(4DR-H^2)^2}{(2D+H)^2(2R+H)^2}$$

Therefore, one can easily calculate $\chi 2$ value for a locus with two alleles and three genotypes. Where the number of individuals in the population are constituted by homozygous dominant (D), heterozygous (H) and homozygous recessives (R) individuals. If the calculated value comes lower than tabulated value at 1 degree of freedom and 5% level of significance $\chi^2_{(1,\,0.05)}$ = 3.84. The null hypothesis of no difference between observed and expected frequencies is accepted. Therefore the differences between observed and expected genotypic frequencies is merely by chance

only and thus the population is said to be in Hardy Weinberg equilibrium.

Illustration 5: Selkirk Rex : The curly coated cat[16]

Fig 9: A white Selkirk Rex[17]

The coat pattern in Selkirk Rex cats follows autosomal incomplete dominant inheritance decided by number of copies of SLK allele. A single copy of the variant that causes the Selkirk Rex coat produces wavy hair, and two copies produce a tighter curl. Additionally cats with two copies have slender body types with long ears. Cats with no copy of SLK alleles have normal coats.

Example 7: Suppose in a survey on 100 Selkirk Rex cat population, there were 10 normal (N/N), 60 cats were having wavy hairs (N/SLK), while the remaining 30 were having tightly curled coats (SLK / SLK). Test whether this population is in Hardy Weinberg equilibrium?

Solution 7: Here, D = 10, H = 60 and R = 30

We know that, $\chi 2$ statistic for a locus with two allele and three genotypes can be given as

$$\chi 2 = \frac{n(4DR - H^2)^2}{(2D + H)^2(2R + H)^2}$$

[16] Serina Filler et al., Selkirk Rex: Morphological and Genetic Characterization of a New Cat Breed, *Journal of Heredity*, Volume 103, Issue 5, September-October 2012, Pages 727–733, https://doi.org/10.1093/jhered/ess039

[17] Photo Courtesy: Ramdrake, A white Selkirk Rex. Shared under the GNU Free Documentation License, Version 1.2 or any later version. https://commons.wikimedia.org/wiki/File:Selkirk_Rex.jpg

$$= \frac{100[4\times10\times30 - (60)^2]^2}{(2\times10 + 60)^2(2\times30 + 60)^2}$$

$$= \frac{100[-2400]^2}{(80)^2(120)^2}$$

$$= \frac{100[-2400]^2}{(80)^2(120)^2}$$

$$= \frac{100[-2400]^2}{(4)^2(2400)^2}$$

$$= \frac{100}{16}$$

$$= 6.25$$

Since the calculated value comes higher than tabulated value at 1 degree of freedom and 5% level of significance ($\chi^2_{(1, 0.05)} = 3.84$). The null hypothesis of no difference between observed and expected frequencies is rejected. Therefore the differences between observed and expected genotypic frequencies is not merely by chance and thus the population is not in Hardy Weinberg equilibrium.

Yates correction

The locus that exhibits more than one form of a gene is said to be polymorphic, provided that frequency of rarer allele must be ≥ 0.01. In such situations if the rarer allele is recessive and its frequency is nearer to 0.01, the frequency of homozygous recessive individuals will be nearer to 0.01% and that of heterozygous individuals will be nearer to 2%. It is quite imperative in exploratory data such frequencies may translate to count fewer than 5 for a particular genotype. In this condition Yates correction is applied to the $\chi 2$ statistic for a locus with two alleles and three genotypes.

$$\chi 2 = \frac{n(|4DR - H^2| - \frac{n}{2})^2}{(2D + H)^2(2R + H)^2} \quad \text{...(Yates correction)}$$

3.7.1.2. Locus with multiple alleles

We have studied in our previous lecture "The genetic structure of population" that if

there are k alleles, then number of possible genotypes in the population will be

$$= \frac{k(k+1)}{2}$$

We will consider a simple case of K=3 for ease of understanding. There will be six genotypes in this case. Let k_1, k_2 and k_3 represent dummy variables for three alleles A, B and C respectively.

k_1	k_2	k_3	G	O	$k_1 \times O$	$k_2 \times O$	$k_3 \times O$	GF	E
2	0	0	AA	n_{AA}	$2 \times n_{AA}$	$0 \times n_{AA}$	$0 \times n_{AA}$	p_1^2	$N.. \times p_1^2$
1	1	0	AB	n_{AB}	$1 \times n_{AB}$	$1 \times n_{AB}$	$0 \times n_{AB}$	p_2^2	$N.. \times p_2^2$
1	0	1	AC	n_{AC}	$1 \times n_{AC}$	$0 \times n_{AC}$	$1 \times n_{AC}$	p_3^2	$N.. \times p_3^2$
0	2	0	BB	n_{BB}	$0 \times n_{BB}$	$2 \times n_{BB}$	$0 \times n_{BB}$	$2p_1p_2$	$N.. \times 2p_1p_2$
0	1	1	BC	n_{BC}	$0 \times n_{BC}$	$1 \times n_{BC}$	$1 \times n_{BC}$	$2p_1p_3$	$N.. \times 2p_1p_3$
0	0	2	CC	n_{CC}	$0 \times n_{CC}$	$0 \times n_{CC}$	$2 \times n_{CC}$	$2p_2p_3$	$N.. \times 2p_2p_3$
Sum				N..	n_A	n_B	n_C	1	
Allelic frequency					$p_1 = \frac{n_A}{2N_{..}}$	$p_2 = \frac{n_B}{2N_{..}}$	$p_3 = \frac{n_C}{2N_{..}}$		

G = Genotypes, O = Given observed frequency, GF = Genotypic frequency (estimation already detailed in section 3.1 of this lecture).

Therefore, $\chi 2 = \sum \frac{(O-E)^2}{E}$

If the calculated value comes lower than tabulated value at 1 degree of freedom and 5% level of significance $\chi^2_{(1,\ 0.05)} = 3.84$. The null hypothesis of no difference between observed and expected frequencies is accepted. Therefore the differences between observed and expected genotypic frequencies is merely by chance only and thus the population is said to be in Hardy Weinberg equilibrium.

4. Summary

The Hardy Weinberg law which was established in the year 1908 has proved to be the backbone of population genetics over time. We have seen that it's equation is simple

but calls for many prerequisites. This has in turn opens a window to many later discoveries on situations when these prerequisites are not fulfilled. We learnt that starting from simple derivation on a single locus with two alleles this law can be generalized for multiple locus and ploidy levels. We also got an idea that Hardy Weinberg law can be applied in understanding population structure under different modes of inheritance, selection experiments, understanding influence of evolutionary forces, genetic counselling and detection of genotypic errors. Thereafter special situation of sex linkage and it's generalization using Jacobsthal series has been understood. We also touched upon and grasped the idea of ternary plots and testing for Hardy Weinberg equilibrium. All these intricacies make this lecture notes pivotal in understanding population genetics further.

****

ABOUT THE AUTHOR

Dr. Kuldeep Kumar Tyagi had completed his B.V.Sc & A.H. in the year 2006 from Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab India. He got admission in a master program in the subject of Animal Genetics and Breeding at Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India after securing 6th rank in All India ICAR-JRF examination. He had completed his Masters in the year 2008 and carried out research on competent fibroblast cells used in somatic cell nuclear transfer. He qualified CSIR Net in his first attempt during the final semester of the masters program itself. He got selected as Assistant Professor in the year 2009 at College of Veterinary Science & A.H. at Navsari Agricultural University, Navsari, Gujarat, India. He enriched his practical knowledge and expertise in the subject of Animal Breeding while disbursing his duties as Scheme Incharge at Livestock Research Station of the same university for 9 years. During the same tenure he also accumulated practical expertise on various aspects of field level breeding programs while heading "All India Coordinated Research Project on Goat Improvement - Surti Field Unit" as Principal Investigator. He completed his Ph.D. in the year 2016 from the same university as an inservice candidate. He had worked on gene expression studies on mammary epithelial cells of buffaloes during his Ph.D. degree program. He had been selected as Associate Professor in the department of Animal Genetics & Breeding, College of Veterinary and Animal Science, Sardar Vallbhbhai, Patel University of Agriculture & Technology, Meerut, Uttar Pradesh, India in the year 2018. Since then he has been heading the same department as Officer-Incharge. He had handled 5 externally funded and 27 institutionally funded research projects. He had coguided two masters students. He has in his credit 62 research papers, 14 research recommendations, 6 lecture notes and 4 success stories. He is a member of 4 professional societies and attended 21 conferences/ symposiums/ workshops. He has remained on a panel of experts for framing question papers for National level, State level examination bodies and various Universities. He is hosting a google site for online teaching https://sites.google.com/view/learnagb and can be reached at drtyagivet@gmail.com for initiating a conversation.

Published by:

Department of Animal Genetics & Breeding
College of Veterinary & Animal Sciences
Sardar Vallabhbhai Patel University of Agriculture and Technology
Meerut- 250 110, Uttar Pradesh, India

To cite this lecture notes:

Tyagi, K 2021, Genetic Structure of Population, lecture notes, Principles of Animal and Population Genetics AGB UNIT II, Sardar Vallabhbhai Patel University of Agriculture & Technology, Meerut, India, Delivered 25, 27 November, 04, 08 December 2020. Retrieved online from https://vepub.com