

Population Genetics

How Much Genetic Variation exists in Natural Populations?

Phenotypic Variation

If we closely examine the individuals of a population, there is almost always **PHENOTYPIC VARIATION** - that is, *variation between individuals in their structure*.

Some of the phenotypic variation in a population results from **ENVIRONMENTAL VARIATION** - *differences in phenotype can result from differences in environmental conditions*.

Genetic Variation

However, much of the phenotypic variation within a population results from **GENETIC VARIATION** - that is, *differences between individuals are often a result of differences in the alleles possessed by these individuals*.

Evidence for Genetic Variation among Plant Populations (Molles p. 193)

Turesson attempted to address whether the differences among plant **ecotypes** were: 1) responses to different environmental conditions or 2) genetic differences that emerged as a result of exposure to different environments

In one set of experiments he harvested seeds of *Campanula rotundifolia* from 9 different geographic localities in Europe and grew them in plots in a common garden

The results (Figure 8.8., p. 194 Molles) showed that the differences among plant ecotypes persisted even when grown in the same environmental conditions

Thus, supporting Turesson's hypothesis of genetic differences among populations

These experiments were taken one step further by **Clausen** and his associates, working with Grew clones of several ecotypes of the flowering plant *Potentilla glandulosa* in 3 different experimental gardens: 30 m, 1400 m, and 3000m elevations

Results are indicated in Figure 8.10 (p. 195 of Molles)

First, they did observe genetic differences among ecotypes; cloned plants from different elevations expressed different heights when grown in the same garden

Second, they apparently revealed evidence of adaptation by ecotypes to local environmental conditions

Each plant clone responded differently to the 3 environmental garden types

Quantifying Genetic Variation

Q. How do we quantify genetic variation?

Three lines of evidenced indicated that natural populations possess a great deal of genetic variation

1. Inbreeding Experiments

As a consequence of these experiments, recessive genes become expressed; before they were hidden

2. Artificial Selection Experiments

In artificial selection the individuals chosen to breed the next generation are those that exhibit the greatest expression of the desired characteristic

If over the generations, the selected population changes in the direction of the selection, it is clear that the original organisms had genetic variation with respect to the selected trait

3. Molecular Genetics

The most convincing evidence that populations possess a large amount of genetic variation came from discoveries in molecular genetics

For example, if we examine a protein and find that it is variable, we can infer that the gene coding for this protein also is variable

We can then measure how variable it is (e.g. how many variant forms exist) and in what frequencies

The Technique of Gel Electrophoresis

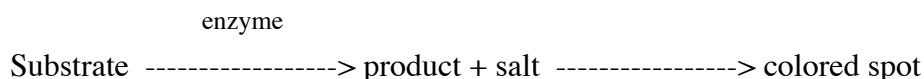
Information about the genetic variation of populations can also be obtained by studying the allozymes (genetic variants of proteins coded for by different alleles at a locus) of populations

Tissue samples for individuals are homogenized in order to release enzymes and other proteins from the cells

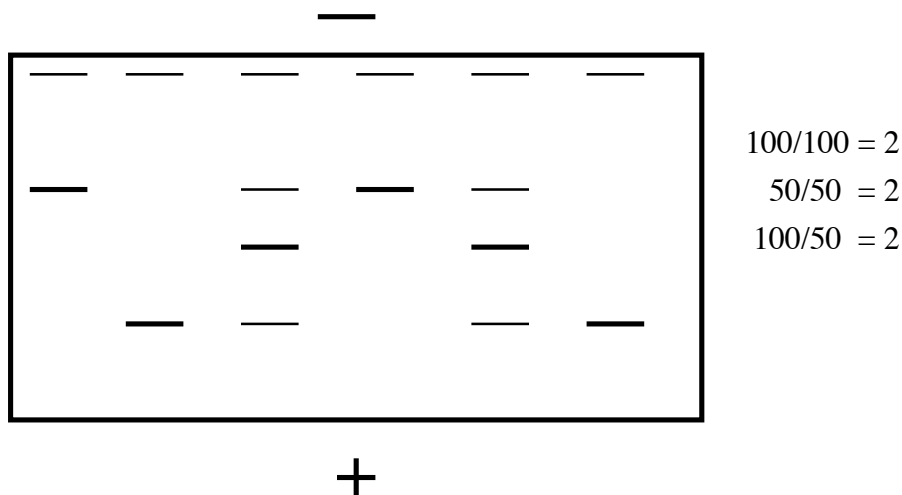
Homogenate supernatant is then placed in a gel and the gel is subjected to an electric current

Each protein in the gel migrates in a *direction* and at a *rate* that depends on the protein's net electric charge and molecular size

The gel is removed from the electric field it is treated with a chemical solution containing a substrate that is specific for the enzyme to be assayed, and a salt that reacts with the product of the reaction catalyzed by the enzyme



The genotype at the gene locus coding for the enzyme can be inferred for each individual in the sample from the number and positions of the spots observed in the gels - the banding patterns observed



There is a relationship between **quaternary structure** of the enzyme and the allozyme phenotype

For example, if the quaternary structure is a dimer, 2 subunits make the functional enzyme enzyme

Subunit assembly is random, therefore, in an individual heterozygous at a dimeric enzyme for slow (S) and fast (F) alleles, three configurations will be seen (SS, SF, and FF)

Because subdivision is random, an SF heterodimer will be 2X more abundant than a homodimer (SS or FF)

Thus, there will be 3 bands with a darker center band

DNA Fingerprinting

DNA fingerprinting allows you to determine genetic variation among closely related individuals based on the specific kinds of bands that are produced on gels

There are duplicated noncoding regions (or repetitive regions) of the DNA referred to as mini- and microsatellite sequences

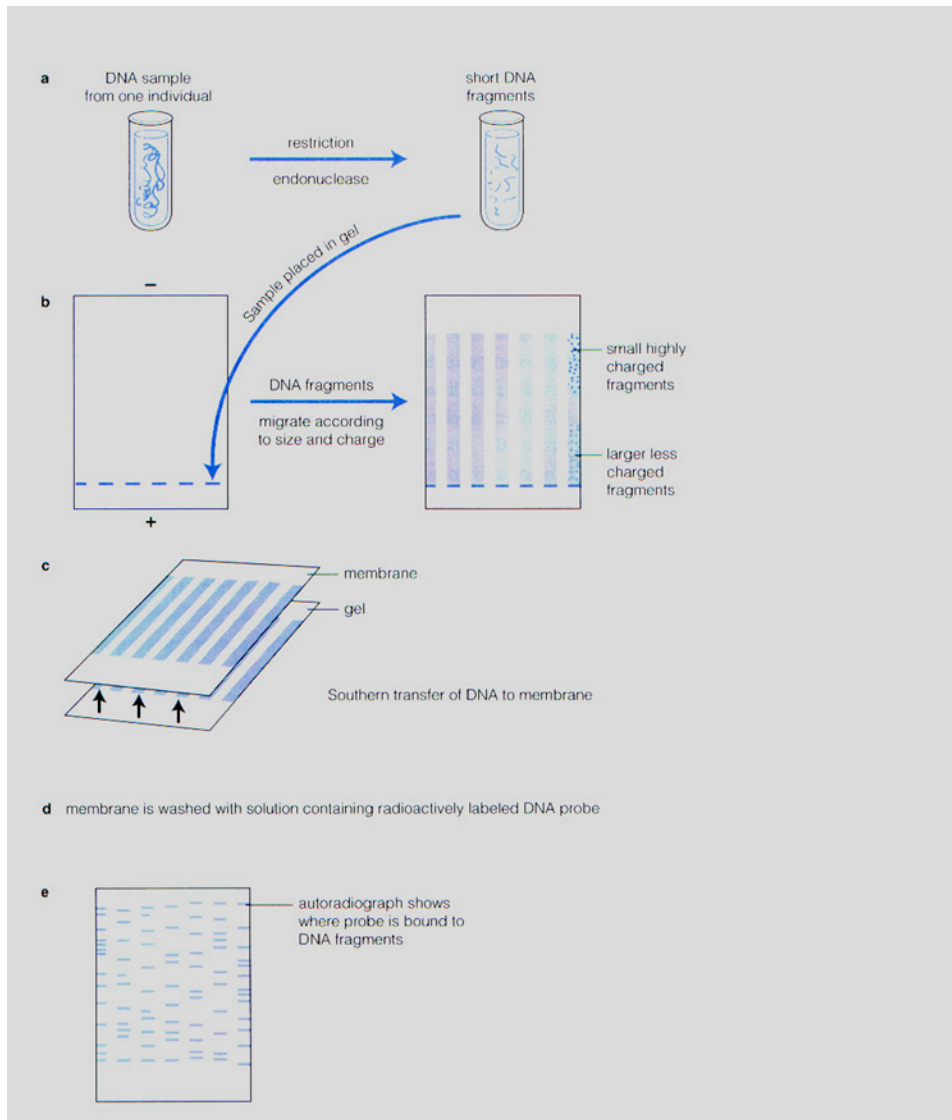
This DNA is similar among closely related organisms – that is, there is a core sequence of nucleotides shared among closely related individuals

However, each individual also has a rather unique sequence - these are highly variable regions and they experience random mutations through time giving

The Process of DNA Fingerprinting (see diagram below)

- a. DNA is isolated from cells and cleaved at specific sites with an endonuclease
- b. The sample containing DNA fragments from each individual is placed in an electrophoretic gel where the fragments are separated by size and charge, producing a streak of fragments of different sizes in each lane of the gel
- c. The DNA fragments are then denatured into single-stranded segments and transferred to a nitrocellulose membrane by a Southern blot transfer
- d. The membrane is then washed with a solution containing single-stranded, radioactively labeled probes for the minisatellite DNA. The probe hybridizes with homologous fragments on the filter
- e. A piece of X-ray film is placed over the membrane and exposed. Each labeled hybrid fragment exposes the film and upon development shows up as a band. The pattern of bands comprises the individuals unique DNA fingerprint

Summary of DNA Fingerprinting



Putting Molecular Data Quantifiable Terms

Polymorphism and Heterozygosity

Polymorphism (P) - the proportion of polymorphic (e.g. variable) loci in a population

We can also determine the average polymorphism by examining the degree of polymorphism in 3 or more populations

Table 2.7
Calculation of the average polymorphism in four populations.

Population	Number of Loci		Polymorphism
	Polymorphic	Total	
1	18	30	$18/30 = 0.60$
2	15	30	$15/30 = 0.50$
3	16	30	$16/30 = 0.53$
4	14	30	$14/30 = 0.47$
			Average: 0.525

The amount of polymorphism is a useful measure of variation for certain purposes, but it suffers from 2 defects: arbitrariness and imprecision

1. Arbitrary because the number of variable loci observed depends on how many individuals are examined

So, in order to avoid the effect of sample size it is necessary to adopt a *criterion of polymorphism*

One criterion often used is that a locus be considered polymorphic only when the most common allele has a frequency of no greater than 0.95 (or rare allele freq no less than 0.05)

2. Imprecise because a slightly polymorphic locus counts as much as a very polymorphic one does

Assume at a certain locus there are 2 alleles with freq of 0.95 and 0.05, while at another locus there are 20 alleles each with a freq of 0.05

More genetic variation exists at the second locus, yet both will weight equally under the 0.95 criterion of polymorphism

Heterozygosity

Heterozygosity - *the average frequency of heterozygous individuals per locus of a population*

Calculations

Obtain the freq of heterozygous individuals at each locus and then average these frequencies over all loci

Table 2.8
Calculation of the average heterozygosity at four loci.

Locus	Number of Individuals		Heterozygosity
	Heterozygotes	Total	
1	25	100	$25/100 = 0.25$
2	42	100	$42/100 = 0.42$
3	9	100	$9/100 = 0.09$
4	0	100	$0/100 = 0$
			Average: 0.19

Heterozygosity is a good measure of variation because it estimates the probability that 2 alleles taken at random from the population for a locus are different

Genetic Variation in Natural Populations

Electrophoretic studies indicate that a considerable amount of genetic variation exists in most natural populations:

	<u>%P</u>	<u>H</u>
Plants	0.26	0.05
Invertebrates	0.47	0.13
Vertebrates	0.25	0.06
Humans	0.30	0.067

Genotypic and Allelic Frequencies

Genotype frequency - the proportion of the population that occurs in each genotype

Allele frequency - the relative proportions of different alleles in a population for a particular gene

Genotype frequency

An electrophoretic examination of the ADH enzyme in a population of field mice reveal 2 alleles for this particular locus: ADH^{100} and ADH^{50}

Examine 100 individuals and find: 50 ADH^{100}/ADH^{100} , 25 ADH^{100}/ADH^{50} , 25 ADH^{50}/ADH^{50}

<u>Genotype</u>	<u># individuals</u>	<u>Genotype frequency</u>
ADH^{100}/ADH^{100}	50	0.50
ADH^{100}/ADH^{50}	25	0.25
ADH^{50}/ADH^{50}	<u>25</u>	<u>0.25</u>
	100	1.00

Note: $\text{freq } ADH^{100}/ADH^{100} + \text{freq } ADH^{100}/ADH^{50} + \text{freq } ADH^{50}/ADH^{50} = 1$

Allele frequency

frequency of allele = freq of homozygous individuals + 1/2 the freq of the heterozygote
for the allele for the allele

Therefore, freq of ADH¹⁰⁰ allele = 0.50 + 1/2 (0.25) = 0.625

$$\text{ADH}^{50} \text{ allele} = 0.25 + 1/2 (0.25) = \frac{0.375}{1.00}$$

$$\text{ADH}^{100} + \text{ADH}^{50} = 1.00$$

Q. How reliable are these estimates?

Two conditions are required for making good estimates of genetic variation:

1. That a random sample of all gene loci is obtained, otherwise, the estimates would be biased
2. That all alleles be detected at every locus

Hardy-Weinberg: Allele Frequencies in the Absence of Evolutionary Change

Q: How do we measure rates of change in the genetic composition of populations?

This is where we begin our discussion of **THEORETICAL POPULATION GENETICS**.

The theory of population genetics is the most fundamental body of theory in evolutionary biology because it provides precise mathematical *predictions*, which can then be *tested*, about how various factors will affect allele frequencies within populations.

How can we predict what allele frequencies will be in the absence of evolutionary change?

This question is important for two reasons:

1. If we want to measure evolution (changes in the frequency of alleles), we need a baseline, or what's called a **NULL HYPOTHESIS** -
2. Before we can construct models that incorporate various types of evolutionary change, we need a model that describes the behavior of alleles in the absence of evolution.

In 1908, the mathematician G. H. HARDY and the geneticist T. WEINBERG, independently published what is now the **HARDY-WEINBERG PRINCIPLE (LAW)**.

The Hardy-Weinberg Principle provides a mathematical description of the behavior of alleles in a population in the absence of evolution

The H-W principle states the following: for a large population of diploid organisms in which mating is random and no evolutionary processes are occurring (e.g. in the absence of evolutionary processes such as mutation, migration, drift, selection), allele and genotype frequencies will reach a stable equilibrium after one generation and do not change thereafter. Or, in the absence of evolution, allele and genotype frequencies for a population will stabilize after one generation and remain constant in subsequent generations.

This stable state is called the *Hardy-Weinberg Equilibrium*.

Overall, there are 2 steps in calculating a H-W equilibrium.

1. The first step is to calculate the allele frequencies in the current generation from the genotype frequencies

<u>genotype</u>	<u>no. indivs.</u>	<u>genotype frequency</u>
ADH ¹⁰⁰ / ADH ¹⁰⁰	60	0.60
ADH ¹⁰⁰ / ADH ⁵⁰	20	0.20
ADH ⁵⁰ / ADH ⁵⁰	<u>20</u>	<u>0.20</u>
	100	1.00

freq of ADH¹⁰⁰ (or p) allele = 0.60 + 1/2 (0.20) = 0.70

freq of ADH¹⁰⁰ (or q) allele = 0.20 + 1/2 (0.20) = 0.30
1.00

Thus we have the genotype and allele frequencies for our population in generation 1

Note:

$$p + q = 1$$

This is always true for 2 alleles at one locus

Once you calculate p, just subtract 1 from it to get q (1 - p = q)

2. Now that we know the genotype and allele frequencies for our population in generation 1, the second step in determining a H-W equilibrium is to *calculate the frequencies of genotypes among the progeny of this population after one round of random mating.*

e.g. if this population were to mate what would be the genotype frequencies of the progeny?

This is important because we want to see if there has been a change from generation to generation

Let's first look at the possible genotypes of the offspring (2nd generation) in our population

Possible mating types and offspring genotypes:

	ADH¹⁰⁰ / ADH¹⁰⁰	ADH¹⁰⁰ / ADH⁵⁰	ADH⁵⁰ / ADH⁵⁰
ADH¹⁰⁰ / ADH¹⁰⁰	ADH ¹⁰⁰ / ADH ¹⁰⁰	ADH ¹⁰⁰ / ADH ¹⁰⁰ ADH ¹⁰⁰ / ADH ⁵⁰	ADH ¹⁰⁰ / ADH ⁵⁰
ADH¹⁰⁰ / ADH⁵⁰	ADH ¹⁰⁰ / ADH ¹⁰⁰ ADH ¹⁰⁰ / ADH ⁵⁰	ADH ¹⁰⁰ / ADH ¹⁰⁰ ADH ¹⁰⁰ / ADH ⁵⁰ ADH ⁵⁰ / ADH ⁵⁰	ADH ¹⁰⁰ / ADH ⁵⁰ ADH ⁵⁰ / ADH ⁵⁰
ADH⁵⁰ / ADH⁵⁰	ADH ¹⁰⁰ / ADH ⁵⁰	ADH ¹⁰⁰ / ADH ⁵⁰ ADH ⁵⁰ / ADH ⁵⁰	ADH ⁵⁰ / ADH ⁵⁰

There are 9 *possible mating types*:

Each of these mating types produces offspring with a characteristic ratio of genotype

To determine the genotype frequencies of the offspring in our population we need 2 pieces of information: *the probability of each mating type and the number offspring of each genotype produced by each mating type.*

We could perform our calculations this way or take advantage of a shortcut.

The short cut we'll use makes use of the fact that diploid organisms produce haploid gametes.

Individuals from our population can produce two types of gametes: they can produce ADH¹⁰⁰ gametes or they can produce ADH⁵⁰ gametes

The frequency of each type of allele in the gametes is the same as the frequency of each type of allele in the parental population.

Recall the frequency of the ADH^{100} allele in our population is 0.7, so the frequency of ADH^{100} gametes is also 0.7.

Similarly, the frequency of the ADH^{50} allele in our population is 0.3, so the frequency of ADH^{50} gametes is also 0.3.

The probability of a ADH^{100} gamete from a male uniting with a ADH^{100} gamete from a female is simply the frequency of ADH^{100} male gametes times the frequency of ADH^{100} female gametes

This is $p \times p$, or p^2 .

Similarly, the probability of a ADH^{50} gamete from a male uniting with a ADH^{50} gamete from a female is simply the frequency of ADH^{50} male gametes times the frequency of ADH^{50} female gametes

This is $q \times q$, or q^2 .

Finally, a ADH^{100} and a ADH^{50} gamete can come together in two ways

1. a ADH^{100} from a male and a ADH^{50} from a female
2. a ADH^{50} from a male and a ADH^{100} from a female.

The probability of a ADH^{100} male gamete uniting with a ADH^{50} female gamete is $p \times q$, and the probability of a ADH^{50} male gamete uniting with a ADH^{100} female gamete is also $q \times p$.

When a ADH^{100} and a ADH^{50} gamete unite this produces ADH^{100}/ADH^{50} offspring

So the frequency of ADH^{100}/ADH^{50} offspring is: $(p \times q) + (q \times p)$, or $2pq$.

So, the genotype frequencies in the offspring (2nd gen.) is given by the binomial expansion:

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

Gamete frequencies and offspring genotype frequencies:

	p (ADH¹⁰⁰)	q (ADH⁵⁰)
p (ADH¹⁰⁰)	p^2 (ADH ¹⁰⁰ / ADH ¹⁰⁰)	pq (ADH ¹⁰⁰ / ADH ⁵⁰)
q (ADH⁵⁰)	qp (ADH ⁵⁰ / ADH ¹⁰⁰)	q^2 (ADH ⁵⁰ / ADH ⁵⁰)

So, the frequency of genotypes in second generation is reflected in frequency of alleles in the male and female gametes of the first generation

Back again to our hypothetical population.

Recall that the frequency of R was 0.7 and that the frequency of r was 0.3.

The genotype frequencies of the offspring are then:

$$p^2 = \text{frequency of ADH}^{100} / \text{ADH}^{100} = (0.7)^2 = 0.49$$

$$2pq = \text{frequency of ADH}^{100} / \text{ADH}^{50} = 2(0.7)(0.3) = 0.42$$

$$q^2 = \text{frequency of ADH}^{50} / \text{ADH}^{50} = (0.3)^2 = 0.09$$

Note that the genotype frequencies have changed from those present in the first generation, but that the allele frequencies have remained stable.

For evolution to occur there must be a change in allele frequencies.

Thus, there has been no evolution.

However, there has been a change in genotype frequency as a result of segregation of alleles and recombination.

These new genotype frequencies represent an EQUILIBRIUM

More specifically, these genotype frequencies represent the H-W equilibrium.

All subsequent generations will have the same genotype frequencies.

Testing for Fit - HW Equilibrium

Use the MN blood group as an example

Observed

<u>Genotype</u>	<u># individuals</u>	<u>frequency</u>
MM	114	0.57
MN	76	0.38
NN	10	0.05

Expected values based on HW Equilibrium

1. Calculate the freq of M and N

$$f M = 0.57 + 1/2 (0.38) = 0.76$$

$$f N = 0.05 + 1/2 (0.38) = 0.24$$

2. Calculate the expected genotype frequencies

$$MM = p^2 = (0.76)^2 = 0.58$$

$$MN = 2pq = 2(0.76)(0.24) = 0.36$$

$$NN = q^2 = (0.24)^2 = 0.06$$

3. Expected genotypes (number of individuals)

$$MM = 0.58 \times 200 = 116$$

$$MN = 0.36 \times 200 = 72$$

$$NN = 0.06 \times 200 = 12$$

4. Perform a Chi-square (χ^2) analysis

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(114 - 116)^2}{116} + \frac{(76 - 72)^2}{72} + \frac{(10 - 12)^2}{12}$$

$$0.034 + 0.222 + 0.333 = \mathbf{0.589}$$

The Hardy-Weinberg Law for 3 alleles

The H-W model can be easily expanded to take account of three or more alleles at a single locus.

We'll quickly look at a 3 allele example

We'll designate these 3 alleles as A^1 , A^2 , and A^3 or p, q, and r respectively

The Table (p. 65 Ayala) demonstrates the equilibrium genotype frequencies for a locus with 3 alleles, with the freq p, q, and r, so that $p + q + r = 1$

For three alleles the genotype frequencies in the second generation can be determined by the multinomial expansion $(p + q + r)^2 = p^2 + 2pq + 2pr + q^2 + 2qr + r^2 = 1$.

Six genotypes are possible with three alleles:

0.2	0.1	0.3	0.1	0.2	0.1
$A^1 A^1$	$A^1 A^2$	$A^1 A^3$	$A^2 A^2$	$A^2 A^3$	$A^3 A^3$
p^2	$2pq$	$2pr$	q^2	$2qr$	r^2

where p is the frequency of A^1 , q is the frequency of A^2 , and r is the frequency of A^3 .

If we assign frequencies to these genotypes in generation 1 you should be able to calculate the genotype frequencies in generation 2, which will be the H-W equilibrium.

Work through this example to make sure you understand how genotype frequencies behave in a 1 gene, 3 allele situation

Allele frequencies:

$$p \text{ or } A^1 = 0.2 + 1/2(0.1) + 1/2(0.3) = 0.4$$

$$q \text{ or } A^2 = 0.1 + 1/2(0.1) + 1/2(0.2) = 0.25$$

$$r \text{ or } A^3 = 0.1 + 1/2(0.3) + 1/2(0.2) = 0.35$$

A Note About Sex-Linked Genes

For sex-linked loci -- those carried on either the X or Y chromosome, the equilibrium frequencies are attained only gradually

In the case of sex-linked genes (e.g. on the X), the equilibrium genotype frequencies for the females (homogametic sex) are the same as in the case for autosomal genes
e.g. if the alleles are A and a, with the freq p and q, there will be p^2 AA, $2pq$ Aa, and q^2 aa females

Calculation of Allele Frequencies for the POPULATION and Approach to Equilibrium for Sex Linked Genes

In the case of sex linked genes, the homogametic sex carries $2/3$ of all genes (on the 2 Xs) in the population

The heterogametic sex carries only $1/3$ (on one X)

Assume that there are 2 alleles (A and a or p and q) in a population and that the freq of A is p_f among the females and p_m among the males

The freq of A in the whole population will be

$$p = 2/3p_f + 1/3p_m$$

Similarly,

$$q = 2/3q_f + 1/3q_m$$

where q, q_f , and q_m are the freqs of the a allele in the whole population, in females and males respectively

When the allele freqs at sex-linked loci are different between males and females, the population does not reach the equilibrium freq in a single generation

Rather, the freq of a given allele among the males in a given generation is the freq of that allele among females in the previous generation

This is because the males inherit their only X chromosome from their mothers

The freq among the females in a given generation is the average of its freq in the females and the males of the previous generation

This is because the females inherit one X chromosome from their fathers and the other from their mothers

Consequently, the distribution of the allele freq of the 2 sexes oscillates, but the difference between the sexes is halved each generation

The population gradually approaches an equilibrium in which both sexes have the same freq

