



**Genetic diversity analysis with  
molecular marker data:  
Learning module**

---

**Basic concepts of  
population genetics**

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 1



## Contents

---

- ▶ Definitions
- ▶ The Hardy-Weinberg principle
- ▶ Examples of calculating allele frequencies
- ▶ Reproduction and mating systems
- ▶ Forces shaping genetic diversity
- ▶ Appendix 1: Critical values of the chi-square distribution

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 2

## Definitions: Population genetics

---

- ▶ The measurement of variability by describing changes in allele frequency for a particular trait over time
- ▶ The analysis of causes leading to those changes

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 3

'Population genetics' is defined in many ways. In general, we can say that population genetics is the study of the application of Mendel's laws and other genetic principles to entire populations of organisms instead of just to individuals. Population genetics is also the study of changes in gene frequencies and, as such, is closely related to evolutionary genetics because evolution depends heavily on changes in gene frequencies. A short introduction to the main factors possibly causing changes in genetic diversity may be found in slides 33 to 42 of this section.

Although inspecting all genetic variables present in a population is virtually impossible, we can examine a population through variation in individual phenotypes (description of certain morphological or physiological traits) and genotypes (molecular markers).

## Phenotype is ...

---

- ▶ The description of all traits of an individual as they concern its morphology, physiology, ecological relationships and behaviour
- ▶ At any given time, the phenotype is the result of the genes of the individual interacting with the environment

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 4

Phenotypic differences can be either qualitative (present or absent) or can be quantitative. Qualitative traits can be classified. Quantitative traits are measured.

Individuals share the same phenotype if they look alike. Some genotypes may have the same phenotype. The distinction between genotype and phenotype is important in those traits that are influenced by the environment: two individuals with the same genotype may result in different phenotypes because of the environment.

## Phenotypic variation

---



---

Copyright: IPGRI and Cornell University, 2003

Population genetics 5

Natural populations are phenotypically diverse. The amount of phenotypic diversity is extraordinary and obvious, even at the most spontaneous observation.

A population of closely related individuals will show low variability. This is especially critical if environmental conditions change and that population does not have the variation to cope with the change. The population could rapidly move towards extinction.

Population genetics deals with phenotypic diversity, especially where it is due to differences in the genotypic composition of individuals.

## Gene and allele

---

- ▶ A gene is the basic physical and functional unit of heredity, passing information from one generation to the next
- ▶ An allele is any of the alternative forms of a gene that can exist at a single locus

Copyright: IPGRI and Cornell University, 2003

Population genetics 6

Not all of a DNA sequence is made up of genes. Genes are those sections of DNA that have a known function. They include a transcribed section and a regulatory element that allows its transcription. The existence of genes is deduced by observing the segregation of variants in the progeny of crosses, whether naturally or artificially produced. This observation was the basis on which Gregor Mendel defined the laws of inheritance at the end of the 19<sup>th</sup> century.

Genes may have two or more alleles. Indeed, a gene may have so many alleles as to constitute an allelic series for that gene. Alleles belonging to a series may show different patterns of dominance to each other. For example, an allele may show a dominant effect, which means that it expresses its phenotype, even if accompanied by a recessive allele. A recessive allele is one where its phenotype is not expressed in an heterozygous individual. If an allele is codominant, its phenotypic effect will be intermediate in the heterozygote relative to the effect of an homozygous dominant and that of an homozygous recessive.

## Genotype is ...

---

- ▶ The description of the complete set of genes that an individual inherits from its parents
- ▶ The genotype of an individual remains unchanged throughout its life, regardless of the environment surrounding and affecting it

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 7

The term 'gene' generally refers to the physical entity that is transmitted from parents to offspring during the reproductive process that influences hereditary traits. An individual's genotype, then, is the sum of all genes inherited from its parents. Genes determine the composition of proteins and also influence external traits and behaviour.

## Genetic variation

---

- ▶ Genetic variation deals with the concept of genotype
- ▶ Genetic variation affecting traits exists in most natural populations, but these traits are influenced by the alleles of many genes in addition to the effects of the environment
- ▶ It is difficult to trace phenotypic differences to the effects of particular genes

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 8

Hidden genetic variation is even more extensive than that observed through the phenotype, so much so that it is virtually impossible for two individuals in a population to have the same genotype at all loci. This genetic variation may be detected through molecular technologies, which reveal polymorphisms, which are then useful as genetic markers. However, even molecular tools are limited and, except for comparisons of entire DNA sequences, most methods are limited to a certain number of genes or loci. Even so, sufficient variation is usually found in samples of genes to make assessing the genetic variation possible for most populations.

As mentioned earlier, because only a small portion of the genome is usually explored in genetic variation studies, questions arise about the reliability of results for extrapolation to natural populations. This becomes an important reason to carefully plan the experiments and pay particular attention to the sampling of both individuals and the loci to be screened.



## What is polymorphism?

---

- ▶ 'Presence of many forms'
- ▶ In genetic terms, it refers to the coexistence of two or more alternative phenotypes in a population or among populations. In general, these diverse phenotypes are caused by alternative alleles of one gene
- ▶ At the molecular level, polymorphism refers to the coexistence of alternative banding patterns or DNA variants when revealed by a given detection method

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 9

A gene or a phenotypic trait is said to be polymorphic if more than one form of the gene or trait exists in a population. Genetic variation, which may cause evolutionary change, is ever-present.

More information about polymorphism in general, and molecular polymorphism in particular, is given throughout the training module *Using Molecular Marker Technology in Studies on Plant Genetic Diversity* ([www.ipgri.cgiar.org/publications](http://www.ipgri.cgiar.org/publications)).

## A population is ...

---

▶ **Ecologically:**

A group of individuals of the same species living within a restricted geographical area that allows any two individuals to interbreed

▶ **Genetically:**

A group of individuals who share a common gene pool and have the potential to interbreed

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 10

Populations are extremely complex entities. In population genetics, the focus is on the local interbreeding unit of a larger population because changes in allele frequencies occur within such limited units and may result in the evolution of adaptive traits. These local interbreeding units are usually called local populations, subpopulations or simply populations. Normally, in a population, members of a species are unevenly distributed. Subdivision of populations are often due to accidents in the environment where they are present. In principle, population size is not infinitely large and does not remain constant.

## Population structure

---

Three levels of population structure are identified:

- Individual organisms
- Subpopulations
- Total population

---

Copyright: IPGRI and Cornell University, 2003

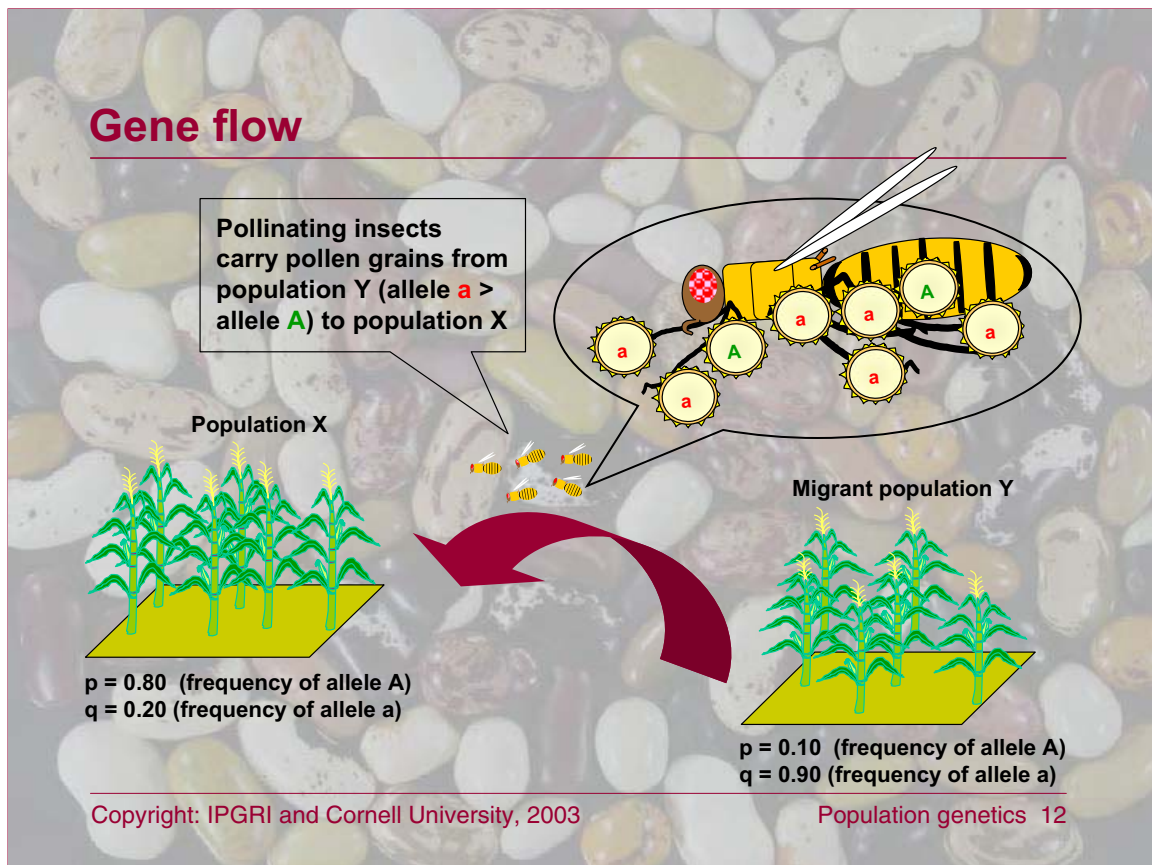
Population genetics 11

A population may be considered as a single unit. However, in many species and circumstances, populations are subdivided in smaller units. Such subdivision may be the result of ecological (habitats are not continuous) or behavioural factors (conscious or unconscious relocation). If a population is subdivided, the genetic links among its parts may differ, depending on the real degree of gene flow taking place.

A population is considered structured if (1) genetic drift is occurring in some of its subpopulations, (2) migration does not happen uniformly throughout the population, or (3) mating is not random throughout the population. A population's structure affects the extent of genetic variation and its patterns of distribution.

See the following slides and the *Glossary* for more details on these new concepts (e.g. gene flow, migration).

## Gene flow



Gene flow is the passage and establishment of genes typical of one population in the gene pool of another by natural or artificial hybridization and backcrossing.

In the drawing above, population Y has a higher frequency of allele a ( $q = 0.90$ ). Pollinators flying from that population will transport more copies of allele a when they move towards another population X. The resulting effect of gene flow is observed in subsequent generations of population X as an increase in the frequency of the migrant allele a.

## Allele frequency

---

- ▶ Allele frequency is the concept used to quantify genetic variation
- ▶ It is defined as a measure of the commonness of a given allele in a population, that is, the proportion of all alleles of that gene in the population that are specifically this type

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 13

An allele is an alternative form of a gene. If a gene corresponds to a specific sequence of nucleotides along a DNA molecule, the alleles represent the different sequences of nucleotides that are possible for that particular locus.

Often the term 'gene' is used synonymously with 'allele' and, hence, 'gene frequency' is sometimes used synonymously with 'allele frequency'. Allelic differences at a single locus in a population indicate genetic variation. This genetic variation needs to be quantified for different genes and for different individuals or populations.

## Calculating the allele frequency

$$P(A) = [2(AA) + (Aa)]/2n$$

- ▶ Twice the number of homozygous genotypes with that allele (because homozygotes carry two copies each of the same allele),
- ▶ plus the number of heterozygous genotypes with that allele (because heterozygotes carry only one copy of a particular allele),
- ▶ divided by two times the total number of individuals in the sample (because each individual carries two alleles per locus)

Copyright: IPGRI and Cornell University, 2003

Population genetics 14

Note that any result obtained with this formula will only be an estimate of the total allele frequency in the population, because only a sample of individuals is usually studied. However, if the sampling of individuals is well done, that is, the size of the sample is sufficiently large, then it can be assumed that our calculation is close to the true allele frequency. As a rule of thumb, allele frequency estimates should be performed, where possible, on samples of 100 individuals or more.

## Genotype frequency

---

- ▶ This is the frequency of a given genotype in a population
- ▶ The frequencies of various types of breeding systems determine the mathematical relationship between the allele and genotype frequencies

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 15

The natural breeding system of individuals can be examined through studies of the frequencies with which alternative genotypes occur in a population. When a population undergoes mating that is random relative to the alleles of interest, certain patterns of genotype frequencies may be expected.

Genotype frequencies are also used to estimate the amount of self-pollination occurring in populations of individuals that have this or mixed type of reproduction. The mating or breeding system, therefore, has a significant effect on the frequency of occurrence of alternative genotypes in a given population.

## The Hardy-Weinberg principle

- ▶ A population with random mating results in an equilibrium distribution of genotypes after only one generation, so that the genetic variation is maintained
- ▶ When the assumptions are met, the frequency of a genotype is equal to the product of the allele frequencies

$$\begin{array}{ccc} \mathbf{AA} & \mathbf{Aa} & \mathbf{aa} \\ \mathbf{p^2} & \mathbf{2pq} & \mathbf{q^2} \end{array}$$

Copyright: IPGRI and Cornell University, 2003

Population genetics 16

The H-W equilibrium states that sexual reproduction does not reduce genetic variation generation after generation; on the contrary, the amount of variation remains constant if there are no disturbing forces acting against it. It establishes the relationship for calculating genotype frequencies under random mating and, in doing so, provides the foundation for many studies in population genetics.

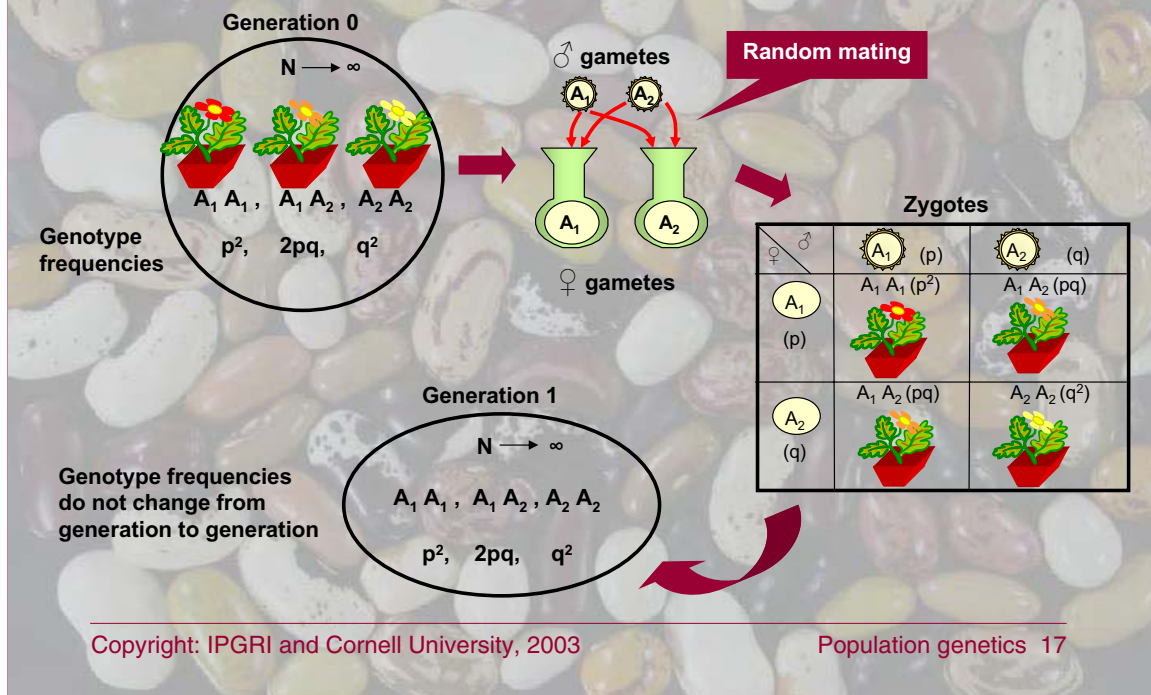
This principle describes the expectations for allele frequencies in an idealized situation where,

- The organism is diploid
- Reproduction is sexual
- Generations are not overlapping
- Mating is random
- Population size is very large
- Migration is negligible
- Mutations can be ignored
- Natural selection does not affect the alleles under consideration

Note that most crop plants violate at least one of these assumptions!



## Demonstrating the H-W principle



The starting point is generation 0. We have a gene with two alleles,  $A_1$  and  $A_2$ . The frequency of allele  $A_1$  is  $p$  and the frequency of allele  $A_2$  is  $q$ . The genotype frequencies in generation 0 are for  $A_1 A_1 = p^2$ , for  $A_1 A_2 = 2pq$  and for  $A_2 A_2 = q^2$ . If random mating occurs, the probability of any allele from the female plant meeting any allele from the male plant will be the same. The table to the right of the slide depicts the four possible genotypes for next generation. The frequency of occurrence of each genotype is given by the product of the frequency of each allele in the genotype (e.g. for  $A_1 A_1$  is  $p \times p = p^2$ ). If the results in the table are summarized, as in the blue inset at the bottom of the slide, we see that the genotype frequencies in generation 1 remain the same as in the previous generation.

## An example of three populations in Hardy-Weinberg equilibrium

Frequencies										
	Genotypes of $G_0$			$G_0$		Genotypes of $G_1$			$G_1$	
Popul.	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	p	q	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	p	q
Pop. 1	0.6	0.2	0.2	0.7	0.3	0.49	0.42	0.09	0.7	0.3
Pop. 2	0.49	0.42	0.09	0.7	0.3	0.49	0.42	0.09	0.7	0.3
Pop. 3	0.4	0.6	0.0	0.7	0.3	0.49	0.42	0.09	0.7	0.3

Copyright: IPGRI and Cornell University, 2003

Population genetics 18

Population genotype frequencies are given in rows. Generations ( $G_0$  and  $G_1$ ) are given in columns. Again, we have one gene with two alleles,  $A_1$  and  $A_2$ . The frequency of allele  $A_1$  is p and the frequency of allele  $A_2$  is q. Genotype frequencies are different for each population in generation 0 (e.g. the frequency of  $A_1 A_1$  in population 1 is 0.6; in population 2, 0.49; and in population 3, 0.4, and so on for the other genotypes). We note, though, that the allele frequencies in the three populations are similar in  $G_0$  ( $p = 0.7$  and  $q = 0.3$ ). In the next generation,  $G_1$ , if all requirements of the H-W principle are met, the genotype frequencies in the three populations balance (now the frequency of  $A_1 A_1$  is 0.49 in the three populations, and the same happens with the frequencies of  $A_1 A_2$  and  $A_2 A_2$ ). The allele frequencies are kept.

## The chi-square test

- ▶ This hypothesis test is useful for determining whether the allelic frequencies are in H-W equilibrium
- ▶ The procedure is as follows:
  - Define  $H_0$  (and  $H_a$ )
  - Define significance level  $\alpha$
  - Perform the statistical test
  - Apply the decision criteria

Copyright: IPGRI and Cornell University, 2003

Population genetics 19

$H_0$  = the hypothesis that says the allele frequencies for trait Q in a given population are in H-W equilibrium.

$H_a$  = an alternative hypothesis that says the allele frequencies for trait Q are not in H-W equilibrium.

We choose a significance level to give us a certain percentage of confidence in our results. The statistical test follows the formula:

Where,

$$ST = \left[ \frac{\sum [O_i - E_i]^2}{(E_i)} \right] = \chi^2_{k - \text{mdf}}$$

ST = statistical test  
O = observed frequencies  
E = expected frequencies

k = number of genotypic classes  
m = number of alleles  
df = degrees of freedom

If our sample allows for only 1 degree of freedom, then the difference in frequencies is reduced by 0.5, a correction factor, such as:

The decision criteria is applied as follows:  $\frac{[|O_i - E_i| - 0.5]^2}{E_i}$

If  $\chi^2_{\text{cal}} \leq \chi^2_{\text{tab}}$  then  $H_0$  is accepted; and, if  $\chi^2_{\text{cal}} > \chi^2_{\text{tab}}$  then  $H_0$  is rejected

Where,

cal = the result of calculating ST with the data obtained in our sample  
tab = the value identified in the table (a  $\chi^2$  table may be found in Appendix 1, click [here](#)).

## Applying the $\chi^2$ test: An example

Genotype	AA	Aa	aa	
No. observed	169	520	311	
No. expected	250	500	250	
$\chi^2$ calculation	$(169-250-0.5)^2/250$	$(520-500-0.5)^2/500$	$(311-250-0.5)^2/250$	
$\chi^2$ value	+25.921	+0.760	+14.641	<b>41.322</b>



Decision criteria:

$$\chi^2_{\text{cal}} (41.322) > \chi^2_{\text{tab}} (3.8)$$

$H_0$  is rejected

Copyright: IPGRI and Cornell University, 2003

Population genetics 20

In this example, let's say that allele frequencies were 0.429 for  $A_1$  and 0.571 for  $A_2$ . Each genotypic class was represented as in the table on the slide above.

The hypotheses being tested are:

$H_0$  = this population is in H-W equilibrium for its allele frequencies

$H_a$  = this population is not in H-W equilibrium for its allele frequencies

Because the number of genotypic classes is 3 and we therefore have only 1 degree of freedom, we apply the correction factor in our calculation of the  $\chi^2$  elements.

The  $\chi^2$  calculated is 41.322. With this figure and with an error margin of 0.05%, the statistical evidence rejects  $H_0$ , which means that this population is not in H-W equilibrium for the trait under study.

## Calculating allele frequencies: Examples

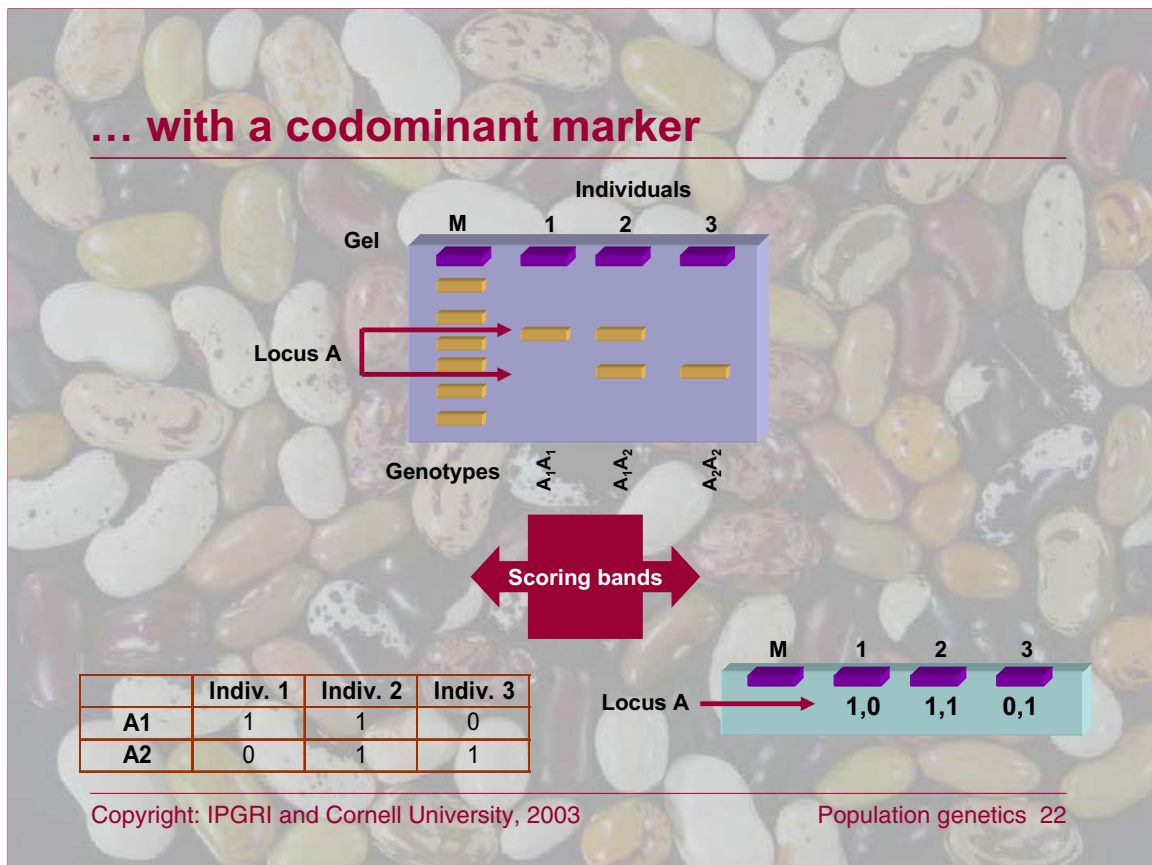
- ▶ Calculating allele frequencies with a codominant marker
- ▶ Calculating allele frequencies with a dominant marker
- ▶ Calculating the allele frequencies with a codominant gene having multiple alleles

Copyright: IPGRI and Cornell University, 2003

Population genetics 21

In the next few slides we show examples of calculations of allele frequencies for results obtained with different marker types. The examples are given with figures that emulate real gels and the bands obtained through the application of molecular markers. For details on molecular marker technologies and the interpretation of bands, see the training module *Using Molecular Marker Technology in Studies on Plant Genetic Diversity* ([http://www.ipgri.cgiar.org/publications/pubfile.asp?ID\\_PUB=912](http://www.ipgri.cgiar.org/publications/pubfile.asp?ID_PUB=912)).

## ... with a codominant marker



(continued on next slide)

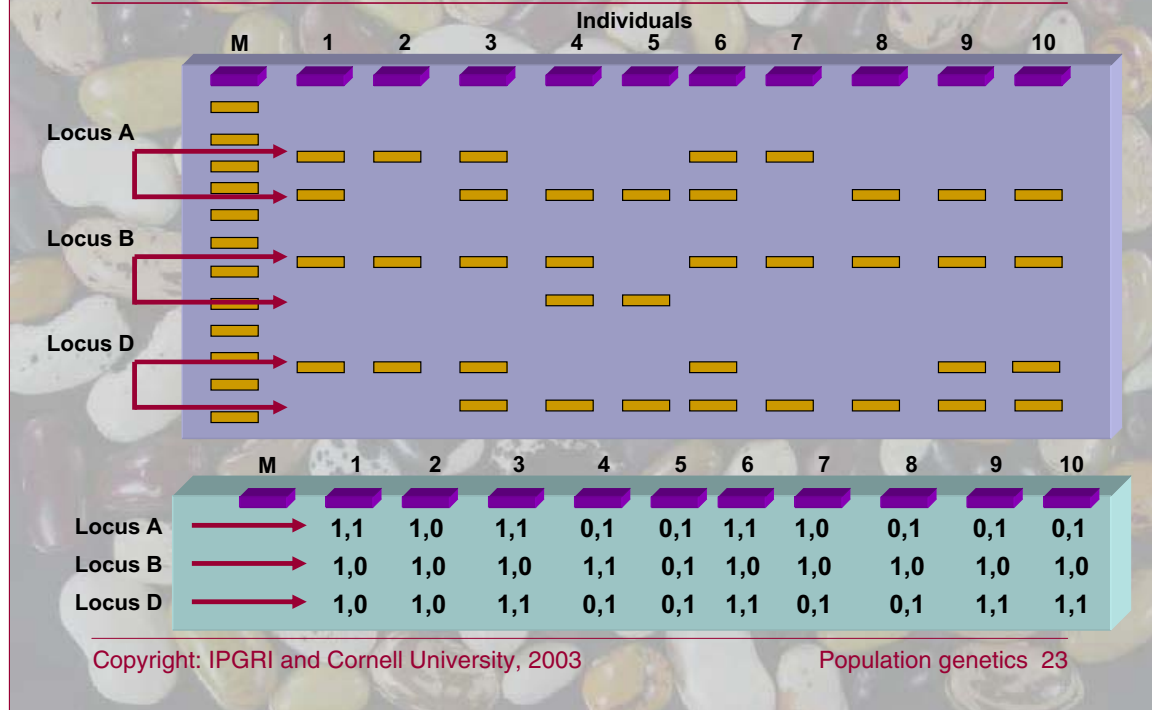
With a codominant marker, the genotypes of the three genotypic classes can be observed for the two homozygotes and the heterozygote. In the drawing above, top centre, we see a gel image with the banding pattern of a codominant marker for a single locus of a diploid organism. We need to score the bands in the gel and convert them to numbers. To do so, each of the band sizes (the band in the same row) is scored and transformed to a 1 if it is present or to a 0 if it is absent. We can do it by band, as in the table at the bottom left, or by genotype, as at the bottom right corner. In the table below we can see the calculations of the expected and observed genotype frequencies, as well as the allele frequencies. (M = size marker.)

Genotypes	A <sub>1</sub> A <sub>1</sub>	A <sub>1</sub> A <sub>2</sub>	A <sub>2</sub> A <sub>2</sub>	Total
Genotype frequency (expected)	$p^2$	$2pq$	$q^2$	1
Number of individuals	$n_{11} = 40$	$n_{12} = 20$	$n_{22} = 140$	$n = 200$
Genotype frequency (observed)	$P_{11} = n_{11}/n = 0.20$	$P_{12} = n_{12}/n = 0.10$	$P_{22} = n_{22}/n = 0.70$	1

$$p = (2n_{11}/2n) + (n_{12}/2n) = P_{11} + \frac{1}{2} P_{12} = 0.20 + \frac{1}{2} (0.10) = 0.25$$

$$q = (2n_{22}/2n) + (n_{12}/2n) = P_{22} + \frac{1}{2} P_{12} = 0.70 + \frac{1}{2} (0.10) = 0.75$$

## ... with a codominant marker (continued)



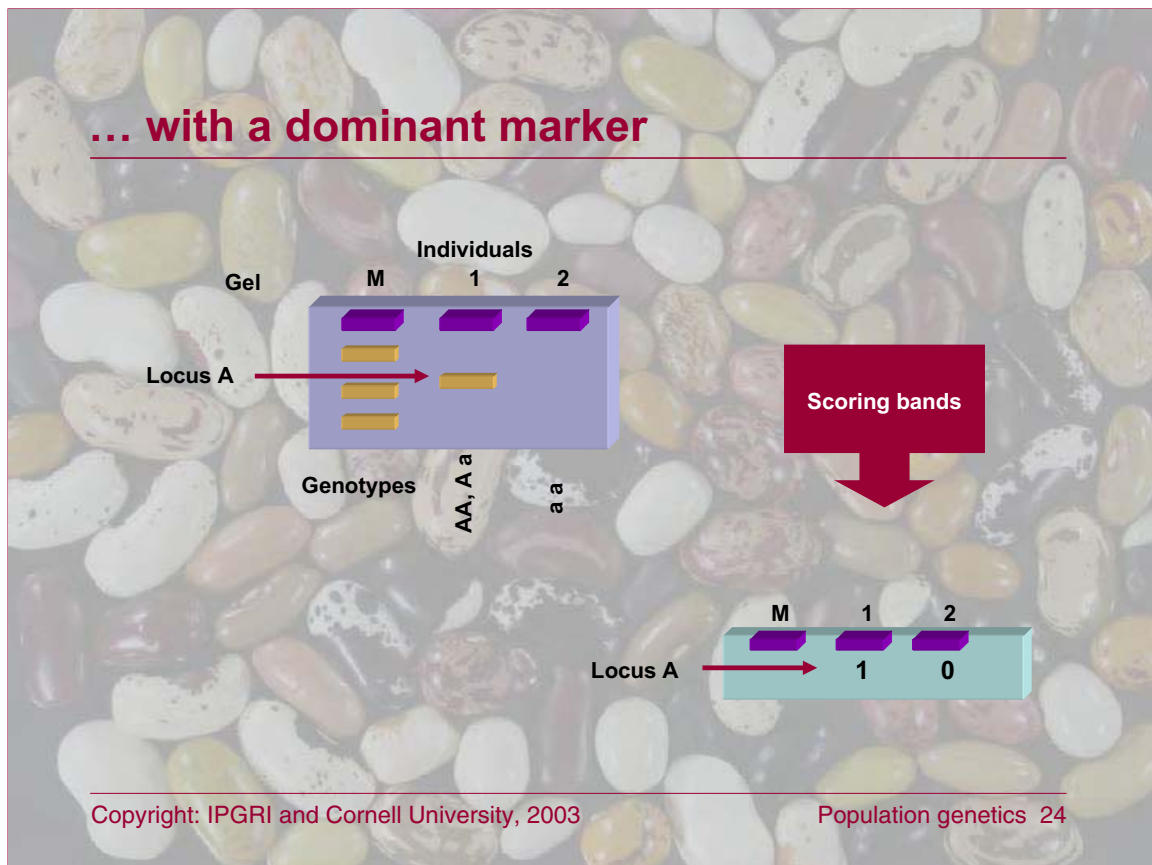
The example in this slide is similar to that in the previous slide, but with 10 individuals and three segregating loci (A, B and D). For ease of presentation, only one scoring method is used (bottom of slide). (M = size marker.)

Note that a gel, like the one in the example, can be obtained only by multiplexing, that is, loading different marker reactions into the same well.

Below, we see a table with the calculations of genotype and allele frequencies. (exp. = expected values; obs. = observed values.)

Locus	Data analysis					Allele freq.	
	Genotypes	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	Total	p	q
A	Genotype freq. (exp.)	$p^2$	$2pq$	$q^2$	1	p	q
	Number of indivs.	2	3	5	10		
	Genotype freq. (obs.)	$P_{11} = 0.2$	$P_{12} = 0.3$	$P_{22} = 0.5$	1	0.35	0.65
	Genotypes	$B_1 B_1$	$B_1 B_2$	$B_2 B_2$	Total	p	q
Genotype freq. (exp.)	$p^2$	$2pq$	$q^2$	1			
B	Number of indivs.	8	1	1	10		
	Genotype freq. (obs.)	$P_{11} = 0.8$	$P_{12} = 0.1$	$P_{22} = 0.1$	1	0.85	0.15
	Genotypes	$D_1 D_1$	$D_1 D_2$	$D_2 D_2$	Total	p	q
	Genotype freq. (exp.)	$p^2$	$2pq$	$q^2$	1		
D	Number of indivs.	2	4	4	10		
	Genotype freq. (obs.)	$P_{11} = 0.2$	$P_{12} = 0.4$	$P_{22} = 0.4$	1	0.40	0.60

## ... with a dominant marker



(continued on next slide)

With a dominant marker, only two genotypic classes can be observed: AA + Aa and aa, that is, one of the homozygote classes is confounded with the heterozygote. The gel image with the banding pattern of a dominant marker for a single locus will show either one band or no band for each individual. The bands are scored in a way similar to that for the codominant marker, where bands are converted to a score of 1 if present or 0 if not. (M = size marker.)

The calculations of frequencies are performed as shown in the table below. (p, q = allele frequencies.)

Phenotypes	A <sub>-</sub>		aa	Total
	AA	Aa	aa	
Genotypes	AA	Aa	aa	
Phenotype frequencies (expected)	$p^2 + 2pq$		$q^2$	1
Number of individuals	$n_1 = 84$		$n_2 = 16$	$n = 100$
Phenotype frequencies (observed)	$P_1 = n_1/n = 0.84$		$P_2 = n_2/n = 0.16$	1

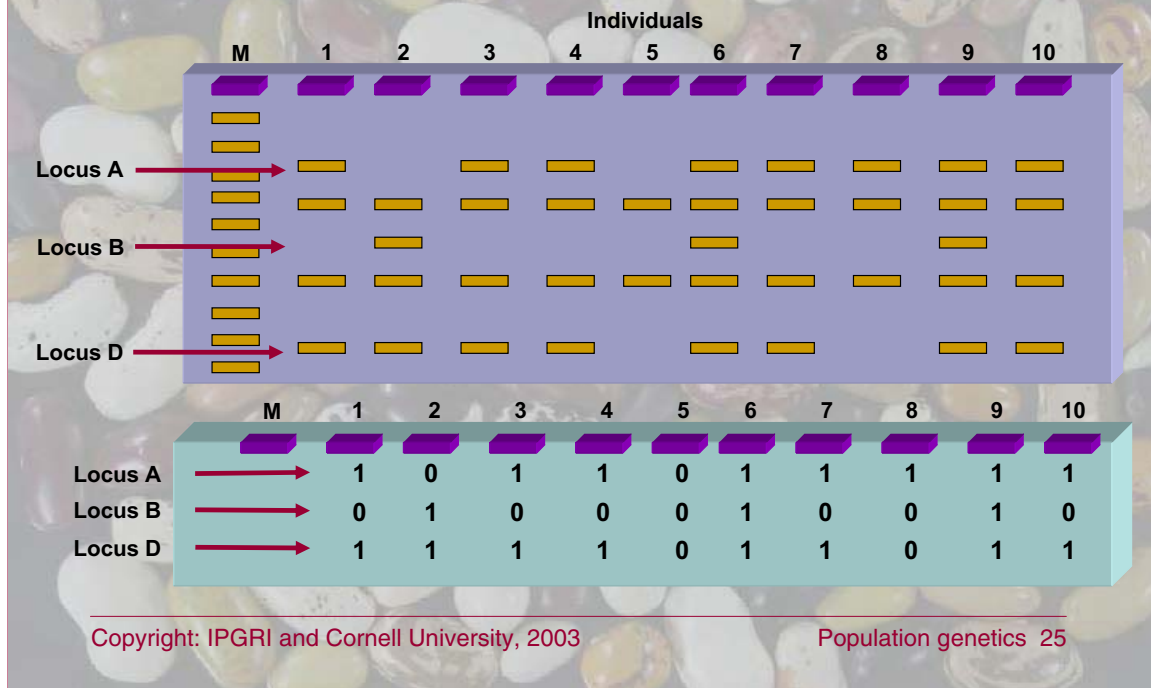
$$q = \sqrt{(n_2/n)} = \sqrt{(P_2)} = \sqrt{(0.16)} = 0.4$$

$$p = (1 - q) = 0.6$$

This is a biased estimate because it does not take into account the recessive alleles in the heterozygotes.



## ... with a dominant marker (continued)



Here we have an example similar to that in the previous slide but with 10 individuals and three segregating loci (A, B, and D). (M = size marker.)

Non-segregating bands (monomorphic) are not scored and thus are not included in the analysis.

Below, we see the table with calculations of genotype and allele frequencies:

Locus	Data analysis				Allele freq.	
	Genotypes	$A_1 \_$	$A_2 A_2$	Total	p	q
A	Genotype freq. (exp.)	$p^2 + 2pq$	$q^2$	1	p	q
	Number of indivs.	8	2	10		
	Genotype freq. (obs.)	$P_1 = 0.8$	$P_2 = 0.2$	1	0.55	0.45
	Genotypes	$B_1 \_$	$B_2 B_2$	Total	p	q
Genotype freq. (exp.)	$p^2 + 2pq$	$q^2$	1			
B	Number of indivs.	3	7	10	p	q
	Genotype freq. (obs.)	$P_1 = 0.3$	$P_2 = 0.7$	1		
	Genotypes	$D_1 \_$	$D_2 D_2$	Total	p	q
	Genotype freq. (exp.)	$p^2 + 2pq$	$q^2$	1		
D	Number of indivs.	8	2	10	p	q
	Genotype freq. (obs.)	$P_1 = 0.8$	$P_2 = 0.2$	1		

We cannot distinguish heterozygotes, but we can estimate the expected number of heterozygotes in a population. For example, if sample size = 1000, then:

For locus A, no. expected heterozygotes =  $2pqN = 2(0.55)(0.45)(1000) = 495$

For locus B, no. expected heterozygotes =  $2pqN = 2(0.16)(0.84)(1000) = 269$

and so on ...

## ... with a codominant gene having multiple alleles

Genotypes	A <sub>1</sub> A <sub>1</sub>	A <sub>1</sub> A <sub>2</sub>	A <sub>1</sub> A <sub>3</sub>	A <sub>2</sub> A <sub>2</sub>	A <sub>2</sub> A <sub>3</sub>	...	A <sub>n</sub> A <sub>n</sub>	Total
Genotype frequencies (expected)	$p_1^2$	$2p_1p_2$	$2p_1p_3$	$p_2^2$	$2p_2p_3$	...	$p_n^2$	1
Number of individuals	$n_{11}$	$n_{12}$	$n_{13}$	$n_{22}$	$n_{23}$	...	$n_{nn}$	$n$
Genotype frequencies (observed)	$P_{11} = \frac{n_{11}}{n}$	$P_{12} = \frac{n_{12}}{n}$	$P_{13} = \frac{n_{13}}{n}$	$P_{22} = \frac{n_{22}}{n}$	$P_{23} = \frac{n_{23}}{n}$	...	$P_{nn} = \frac{n_{nn}}{n}$	1

$$p_1 = P_{11} + \frac{1}{2} \sum_{i \neq 1} P_{1i}$$

$$p_2 = P_{22} + \frac{1}{2} \sum_{j \neq 2} P_{2j}$$

$$p_3 = P_{33} + \frac{1}{2} \sum_{j \neq 3} P_{3j}$$

$$p_4 = P_{44} + \frac{1}{2} \sum_{j \neq 4} P_{4j}$$

$$p_n = P_{nn} + \frac{1}{2} \sum_{j \neq n} P_{nj}$$

Copyright: IPGRI and Cornell University, 2003

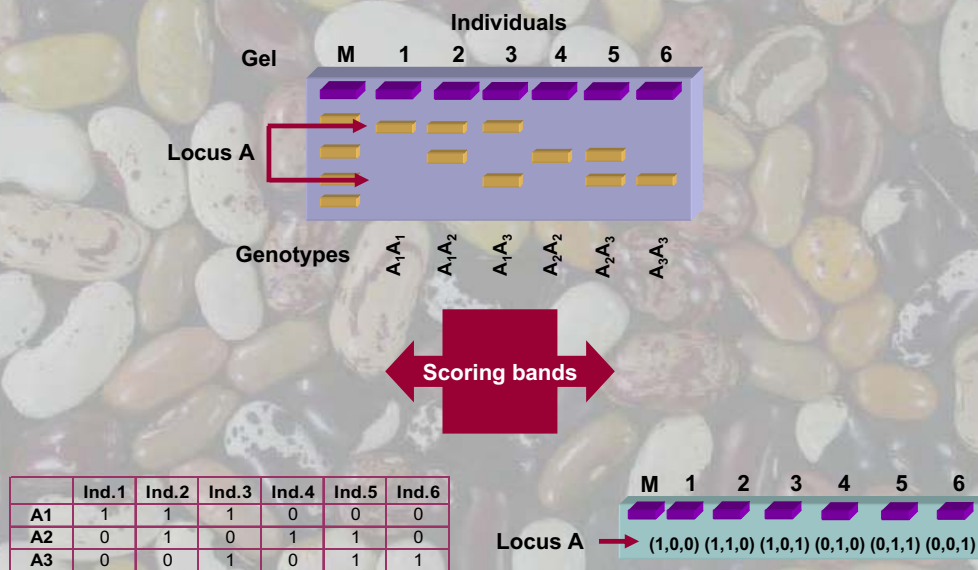
Population genetics 26

(continued on next slide)

This is the situation encountered when markers such as microsatellites are used.

We have a locus A with  $n$  alleles  $A_1, A_2, A_3, \dots, A_n$  and allele frequencies  $p_1, p_2, p_3, \dots, p_n$ , respectively, such that  $A_1 = A_2 = A_3 = \dots = A_n$

## ... with a codominant gene having multiple alleles (continued)



Copyright: IPGRI and Cornell University, 2003

Population genetics 27

Again, with a codominant marker, the genotypes of the three genotypic classes can be observed. In the drawing above, top centre, we see a gel image with the banding pattern of a codominant marker with three alleles ( $A_1$ ,  $A_2$  and  $A_3$ ) in a diploid sample. We score each band (each row) independently, and transform them to a score of 1 if present or a score of 0 if not. We can do it by band (bottom left of slide) or by genotype (bottom right corner). In the table below, we can see the calculations of the expected and observed genotype frequencies, as well as the allele frequencies ( $p_1$ ,  $p_2$  and  $p_3$ ). (M = size marker.)

Genotypes	$A_1 A_1$	$A_1 A_2$	$A_1 A_3$	$A_2 A_2$	$A_2 A_3$	$A_3 A_3$	Total
Genotype frequency (exp.)	$p_1^2$	$2p_1p_2$	$2p_1p_3$	$p_2^2$	$2p_2p_3$	$p_3^2$	1
Number of individuals	$n_{11} = 4$	$n_{12} = 6$	$n_{13} = 0$	$n_{22} = 10$	$n_{23} = 2$	$n_{33} = 2$	$n = 24$
Genotype frequency (obs.)	$P_{11} = n_{11}/n = 0.17$	$P_{12} = n_{12}/n = 0.25$	$P_{13} = n_{13}/n = 0$	$P_{22} = n_{22}/n = 0.42$	$P_{23} = n_{23}/n = 0.08$	$P_{33} = n_{33}/n = 0.08$	1

$$p_1 = P_{11} + \frac{1}{2}P_{12} + \frac{1}{2}P_{13} = P_{11} + \frac{1}{2}\sum_{j \neq 1} P_{1j} = 0.17 + \frac{1}{2}(0.25 + 0.00) = 0.30$$

$$p_2 = P_{22} + \frac{1}{2}P_{21} + \frac{1}{2}P_{23} = P_{22} + \frac{1}{2}\sum_{j \neq 2} P_{2j} = 0.42 + \frac{1}{2}(0.25 + 0.08) = 0.59$$

$$p_3 = P_{33} + \frac{1}{2}P_{31} + \frac{1}{2}P_{32} = P_{33} + \frac{1}{2}\sum_{j \neq 3} P_{3j} = 0.08 + \frac{1}{2}(0.00 + 0.08) = 0.12$$

## Reproduction and mating systems

- ▶ Outbreeding, inbreeding or asexual reproduction
- ▶ They influence:
  - The degree of genetic relatedness among mates
  - The organization of genes into genotypes

Copyright: IPGRI and Cornell University, 2003

Population genetics 28

In principle, outbreeding occurs as random mating, and inbreeding and asexual reproduction both involve nonrandom mating.

Outcrossing species, compared with inbreeders, may retain large numbers of deleterious recessives because the dominance situation hides them. The recessives undergo frequent recombination, resulting in new gametic types.

Dominance refers to situations where, under heterozygous conditions, one allele has a phenotypic effect that is strong enough to conceal the presence of the other (recessive) allele. In a situation of dominance, only two phenotypes can be observed: the dominant phenotype that is a mixture of the homozygous dominant and the heterozygous, and the recessive phenotype.

In cross-pollinating species, selfing leads to inbreeding depression because it increases the proportion of homozygotes, thus permitting rare recessive alleles to become visible. Heterozygotes in cross-pollinating species have a more favourable effect.

In organisms with asexual reproduction, a variety of consequences result, depending on its type. Although asexual reproduction may be a constant mode of reproduction, it may be combined with cycles of sexual reproduction, which allow recombination of the current variation and, as such, the generation of new forms or combinations. If only asexual reproduction occurs in the population, genotypic frequencies cannot be changed.

## Random mating

---

- ▶ Mating that takes place at random, that is, the chances of individual A mating with individual B do not depend on the genotypes of either
- ▶ If random mating occurs, the chance that an individual mates with a given genotype is equal to the frequency of that genotype in the population

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 29

Random mating is typically found in many outbreeding populations. For example, we may have a population in which genotype AA is present 10% of times, Aa 58% of times and aa 32% of times. If mating is random, then the chances of an individual AA mating with another AA is 10/100, Aa 58/100, or aa 32/100.

## Nonrandom mating

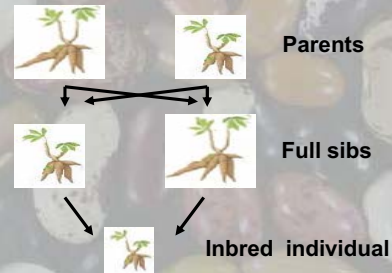
### ▶ Classified mating:

- Positive: mating among individuals with similar phenotypes
- Negative: mating among individuals with dissimilar phenotypes



### ▶ Inbreeding:

- Mating among relatives



Copyright: IPGRI and Cornell University, 2003

Population genetics 30

Nonrandom mating occurs when individuals that are more closely (inbreeding) or less closely related mate more often than would be expected by chance for the population.

Self-pollination or inbreeding is similar to mating between relatives. It increases the homozygosity of a population and its effect is generalized for all alleles. Inbreeding per se does not change the allelic frequencies but, over time, it leads to homozygosity by slowly increasing the two homozygous classes.

## Inbreeding coefficient

---

- ▶ It compares the actual proportion of heterozygous genotypes with those expected under random mating

$$F = \frac{(H_0 - H)}{H_0}$$

- ▶ F is the inbreeding coefficient and measures the reduction of heterozygosity

---

Copyright: IPGRI and Cornell University, 2003

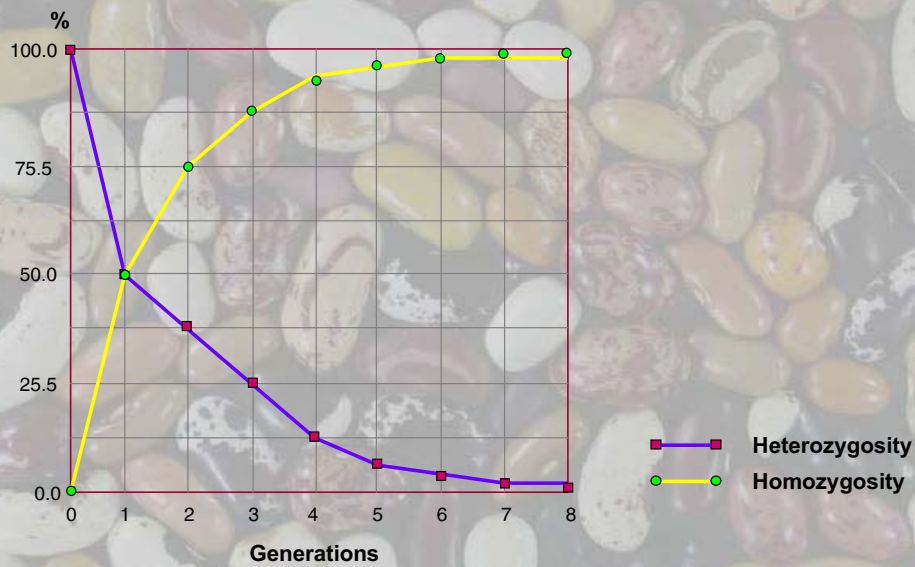
Population genetics 31

H = actual frequency of heterozygotes in the population.

H<sub>0</sub> = expected number of heterozygotes under random mating.

The inbreeding coefficient indicates the degree of inbreeding in a population.

## What happens in selfing?



Copyright: IPGRI and Cornell University, 2003

Population genetics 32

Selfing is a powerful inbreeding system that allows the attainment of high levels of homozygosity within few generations. Simultaneously, heterozygosity decreases. The chart on the slide shows the phenomenon and, in the table below, we can see the changing values of homozygosity and heterozygosity in 9 generations ( $G_0$  to  $G_8$ ).

Gener- ation	Selfed genotypes ratio/generation	Homozygosit y (%)	Heterozygosit y (%)
$G_0$	Aa	0	100
$G_1$	1AA, 2Aa, 1aa	50	50
$G_2$	6AA, 4Aa, 6aa	75	25
$G_3$	28AA, 8Aa, 28aa	87.5	12.5
$G_4$	120AA, 16Aa, 120aa	93.75	6.25
$G_5$	496AA, 32Aa, 496aa	96.875	3.125
$G_6$	2016AA, 64Aa, 2016aa	98.4375	1.5625
$G_7$	8128AA, 128Aa, 8128aa	99.21875	0.78125
$G_8$	32640AA, 256Aa, 32640aa	99.60938	0.390625





## Forces shaping genetic diversity

---

- ▶ Mutation
- ▶ Migration
- ▶ Recombination
- ▶ Selection
- ▶ Drift

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 33

If, for any reason, a population becomes homogeneous, evolution would not occur. Thus, constant change essentially depends on new variation.

A genetic population is the aggregate of allelic frequencies of all genes in that population. Populations change or evolve because their gene frequencies experience change. Several factors can produce changes in fitness—the capability of an individual to survive until reproduction is achieved. If the fitness of an individual in a population changes, the genotypes in the subsequent generation will not be directly related to the gene frequencies of the former population, thus driving the population to evolve.

Because changes in populations require changes in gene frequencies, we must understand how these frequencies can change. In the next slides, we discuss the primary causes of change: mutation, migration, recombination, selection and drift.

## Mutation

---

- ▶ It is the ultimate source of variation and may be caused by:
  - Errors in DNA replication
  - Damage by radiation
- ▶ Mutation increases diversity but, because spontaneous mutations are rare, the rate of change in gene frequency is very low
- ▶ Consequently, mutation alone does not drive the evolution of populations and species

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 34

The simplest mutation is that which produces a change in a single nucleotide in the DNA sequence of a gene. A mutation can cause one allele to change to another allele already present in the population (from a dominant to a recessive) or it could create a completely new allele. Mutations may be adverse or favourable. Many will be adverse and disappear. But, if they are good for the individual, then frequencies of that allele will increase from generation to generation. Moreover, this mutation may migrate to other populations and spread.

*Note:* Genomes may go through a process known as gene duplication. This event helps the individual to endure an unfavourable mutation in a copy of the gene without major difficulties, because the other copy of the gene still may function appropriately. More changes may affect the mutated gene and determine different types of adaptation for the individual.

## Migration

---

- ▶ It is the movement of individuals or any form of introduction of genes from one population to another
- ▶ Migration increases diversity and its rate can be large, causing significant changes in frequency
- ▶ The change in gene frequency is proportional to the difference in frequency between the recipient population and the average of the donor populations

Copyright: IPGRI and Cornell University, 2003

Population genetics 35

From a genetic perspective, migration implies not only the movement of individuals into new populations but that this movement introduces new alleles into the population (gene flow). Changes in gene frequencies will occur through migration either because more copies of an allele already present will be brought in or because a new allele arrives.

Several factors affect migration in crop species:

- Breeding system
- Sympatry with wild and/or weedy relatives
- Pollinators
- Seed dispersal

The immediate effect of migration is to increase a population's genetic variability and, as such, helps increase the possibilities of that population to withstand environmental changes. Migration also helps blend populations and prevent their divergence.

## Recombination

---

- ▶ It is the process whereby a cell generates new chromosomal combinations, compared with that cell or with those of its progenitors
- ▶ It does not create new diversity but generates new combinations of the existing diversity
- ▶ Genetic variance through recombination, given that segregating alleles exist at different loci, can occur very quickly

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 36

Genetic diversity through recombination is the result of reshuffling the genetic components of the original type. Mechanisms exist to generate allelic diversity (intragenic recombination) and genomic diversity (new multigene combinations).



## Selection

---

It is the inherited ability of organisms to survive and reproduce. It acts in such a manner that, with time, superior genotypes increase their frequency in the population

---

Copyright: IPGRI and Cornell University, 2003

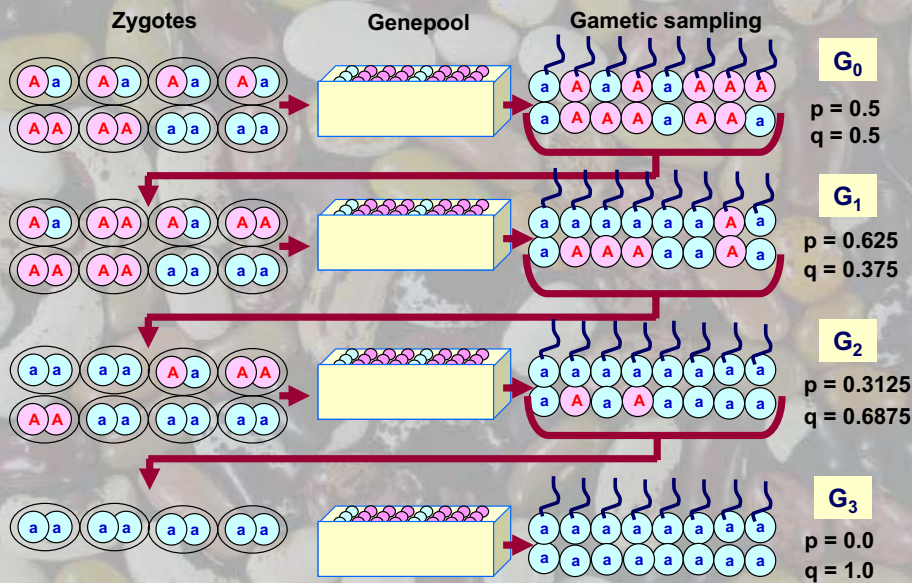
Population genetics 37

As a result of mutation, new forms develop. These forms, as explained, may be favourable or deleterious to the individual's ability to survive. If changes are advantageous, then the new alleles will tend to prevail by being selected in the population.

The effect of selection on diversity may be:

- Directional, where it decreases diversity
- Balancing, where it increases diversity. Heterozygotes have the highest fitness, so selection favours the maintenance of multiple alleles
- Frequency dependent, where it increases diversity. Fitness is a function of allele or genotype frequency and changes over time

## Drift



Copyright: IPGRI and Cornell University, 2003

Population genetics 38

Genetic drift refers to fluctuations in allele frequencies that occur by chance (particularly in small populations) as a result of random sampling among gametes.

Drift decreases diversity within a population because it tends to cause the loss of rare alleles, reducing the overall number of alleles.

In the example above, population size is constant in each generation (8 individuals). Each individual may produce thousands of gametes but, from the total gene pool in each generation, only  $2N$  gametes per individual are needed (16 in our example). This situation is similar to extracting small samples from two boxes, one containing a million white balls and the other a million red balls. In each extracting experiment, we may take a different number of white and red balls. We simulate that, in generation G<sub>0</sub>, 10 gametes of the thousands possible carried allele A and only 6 carried allele a. In G<sub>1</sub>, from those gametes that participated in the constitution of the zygotes for next generation, 5 carried allele A and 11 allele a, and so on. These values vary randomly. In G<sub>3</sub> all individuals are formed by allele a (homozygotes) and allele A is lost.

## Effective population size

---

- ▶  $N_e$  is the number of parents responsible for the genetic composition of the next generation
- ▶  $N_e$  is generally less than  $N$  because of
  - Variation of population size from generation to generation
  - Unequal sex ratio
  - Overlapping generations
  - Geographic dispersion of populations

---

Copyright: IPGRI and Cornell University, 2003

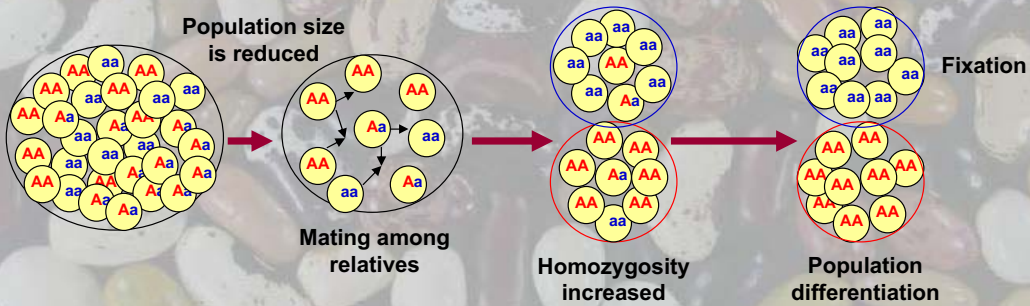
Population genetics 39

How big is the population?

- The actual number of individuals in a population is called the census number ( $N$ ). This number is often an imprecise representation of the population size from a genetic standpoint.
- The effective population size ( $N_e$ ) describes the size of an ideal population that shows the same rate of loss of genetic variation due to genetic drift as for the population of interest.

## Consequences of a decreasing population size

- ▶ Genetic drift, with random variation of allele frequencies
- ▶ Inbreeding
- ▶ Homozygosity: fixation and loss of alleles
- ▶ Subpopulation differentiation



Copyright: IPGRI and Cornell University, 2003

Population genetics 40

(continued on next slide)

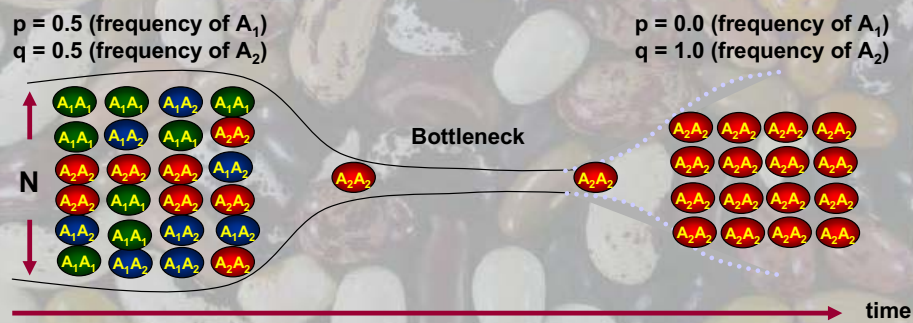
Events that reduce population size:

- Domestication
- Subpopulations (inbreeding, clonal reproduction)
- Long-range dispersal (founder effect)
- Regeneration of genetic resources collections



## Consequences ... (continued)

- ▶ A bottleneck develops when the population size drops sharply
- ▶ The founder effect occurs when a few individuals colonize and become established in a new environment



Copyright: IPGRI and Cornell University, 2003

Population genetics 41

Small populations are highly vulnerable to extinction because the surviving sample may not be representative of the pre-crash gene pool.

Both effects depend on the number of survivors (or colonizers) and the rate of population growth.

The graphic in the slide shows a bottleneck effect. To the left is a population in Hardy-Weinberg equilibrium with allele frequencies of 0.5. If a sudden reduction occurs, then the original size restored, the result may be a loss of alleles and the fixation of others. In the example (right), only  $A_2 A_2$  survived and allele  $A_1$  was lost.

## With bottleneck and founder effects ...

- ▶ Heterozygosity declines at the rate of:

$$H_1 = (1 - 1/2N)H_0$$

- ▶ Alleles are lost at the rate of:

$$P = p^{2N} + q^{2N}$$

Copyright: IPGRI and Cornell University, 2003

Population genetics 42

- Heterozygosity decline

Where,

$H_1$  = final heterozygosity

$H_0$  = initial heterozygosity

$N$  = population size

Then:

If  $N = 100$  and  $H_0 = 0.25$ , then  $H_1 = 0.24875$

If  $N = 40$  and  $H_0 = 0.25$ , then  $H_1 = 0.24685$

- Allele loss

Where,

$P$  = allele loss

$p$  and  $q$  = allele frequencies

$2N$  = total number of alleles in the population

Then:

If  $N = 100$ ,  $p = 0.90$  and  $q = 0.1$ , then  $P = 7.05508 \times 10^{-10}$

If  $N = 15$ ,  $p = 0.90$  and  $q = 0.1$ , then  $P = 0.0423911$

If  $N = 10$ ,  $p = 0.90$  and  $q = 0.1$ , then  $P = 0.12157665$

## Appendix 1

---

Copyright: IPGRI and Cornell University, 2003

Population genetics 43

Appendix 1. Critical values of the chi-square distribution



## In summary

---

To analyse and interpret genetic diversity data we must be acquainted with:

- A few basic definitions of population genetics
- The Hardy-Weinberg principle
- Calculating allele and genotype frequencies
- The primary causes of changes in genetic diversity: mutation, migration, recombination, selection and drift

## By now you should know ...

---

- ▶ The basic definitions used in population genetics
- ▶ The Hardy-Weinberg principle
- ▶ How to calculate allele frequencies with marker data
- ▶ The mating system's effect on a population's diversity
- ▶ The main sources of variation and their consequences



If you would like to read more ...

Copyright: IPGRI and Cornell University, 2003

Population genetics 46

de Vicente, M.C. and T. Fulton. 2003. Using Molecular Marker Technology in Studies on Plant Genetic Diversity.

[www.ipgri.cgiar.org/publications/pubfile.asp?ID\\_PUB=912](http://www.ipgri.cgiar.org/publications/pubfile.asp?ID_PUB=912) >

Doolittle, D.P. 1987. Population Genetics: Basic Principles. Springer-Verlag, Berlin.

Falconer, D.S. and T.F.C. Mackay (eds.). 1996. Introduction to quantitative genetics (4th edn.). Longman Group, London.

Griffiths, A.J.F., J.H. Miller, D.T. Suzuki, R.C. Lewontin and W.M. Gelbart (eds.). 1996. An Introduction to Genetic Analysis (6th edn.). Freeman and Co., NY.

Hartl, D.L. 1988. A Primer of Population Genetics (2nd edn.). Sinauer Associates, Sunderland, MA.

Hedrick, P.W. 1985. Genetics of Populations. Jones and Barlett Publishers, Boston, MA.

Snedecor, G.W. and W.S. Cochran (eds.). 1980. Statistical Methods (7th edn.). Iowa State University Press, Ames, IO.



## Next

---

- ▶ Measures of genetic diversity
- ▶ Software programs for analysing genetic diversity
- ▶ Glossary