Population genetics

Deependra Dhakal
Gokuleshwor Agriculture and Animal Science College
Tribhuwan University
ddhakal.rookie@gmail.com
`https://rookie.rbind.io`

Academic year 2019-2020

Outline

Outline

What is allele/gene ?

An allele/gene is the bit of DNA at the place called locus (the place on a chromosome where an allele resides). An allele is instantiation of a locus. But by orthology, a locus is not template for an allele. Similarly, a locus is not tangible, rather a map describing where to find a tangible thing, an allele on a chromosome. A diploid individual has two alleles at a particular autosomal locus.

Genotype and allele frequencies

In A loci, suppose, two alleles A_1 and A_2 are present in a diploid organism the genotype and genotypic frequency of segregating population will be;

Genotype	A_1A_1	A_1A_2	A_2A_2
Relative frequency	x_{11}	x_{12}	x_{22}

As, relative frequencies must add to 1,

$$x_{11} + x_{12} + x_{22} = 1$$

The order of subscripting heterozygous is arbitrary.

Frequency of A_1 allele in the population is,

$$p = x_{11} + \frac{1}{2}x_{12}$$

and frequency of A_2 allele is,

$$q = 1 - p = x_{22} + \frac{1}{2}x_{12}$$

Measure of each allele frequency can be thought of as independent events. For e.g., for allele p to be selected;

$$p = \left(x_{11} \times \frac{1}{P(p_{A_1A_1})}\right) + \left(x_{12} \times \frac{1}{2}\right) + (x_{22} \times 0)$$

Where, $P(p, A_1A_1)$ is the probability of getting p allele from A_1A_1 genotype, for loci with more than two alleles, frequency of i^{th} allele will be called p_i. Frequency of A_iA_j genotype will be called x_ij for heterozygotes, $i \neq j$ and, by convention, $i < j$.

If there are n alleles,

$$1 = x_{11} + x_{22} + x_{33} + \ldots + x_{nn} + x_{12} + x_{13} + x_{(n-1)n}$$
$$= \sum_{i=1}^{n} \sum_{j \geq i}^{n} x_{ij}$$

The frequency of i^th allele is

$$p_i = x_{ii} + \frac{1}{2} \sum_{j=1}^{i-1} x_{ji} + \frac{1}{2} \sum_{j=i+1}^{n} x_{ij}$$

Outline

- In a random mating population (in which each male gamete has an equal chance of mating with any female gamete), when taken a loci with differences in alleles (A/a), following genotypes are possible: *AA*, *Aa* and *aa*.
- With the corresponding frequencies of p^2, $2pq$ and q^2, respectively for each of the genotypes, and bearing that gene frequencies must add up to unity, $p^2 + 2pq + q^2 = 1$. This mathematical relationship is called the **Hardy-Weinberg equilibrium**.
- Two scientists showed that the frequency of genotypes in a population depends on the frequency of genes in the preceding generation, not on the frequency of the genotypes.

- In each subsequent generation following thereafter, however, gene and genotypic frequencies will remain unchanged, provided that:

1. Random mating occurs in a very large diploid population;
2. Allele A and allele a are equally fit (one does not confer a superior trait than the other);
3. There is no differential migration of one allele into or out of the population;
4. The mutation rate of allele A is equal to that of allele a.

Conservation of gene frequencies

- Let us presume that among humans, the difference between those who can and those who cannot taste the chemical phenyl-thiocarbamate (PTC) resides in a single gene difference with two alleles, *T* and *t*.
- The allele for tasting, *T*, is dominant over *t*, so that heterozygotes, *Tt* are tasters, and the only nontasters are *tt*.
- If we were to choose an initial population composed of an arbitary number of each genotype, we may ask what will be the frequency of these genes after many generations.
- Let us, for example, place upon an island a group of children in the ratio .40*TT* : .40*Tt* : .20*tt*. The gene frequencies in this newly formed population are therefore, .4 + .2 = .6*T*, and .2 + .2 = .4*t*.
- Let us also assume that the number of individuals in the population is large, and the tasting or nontasting has no effect upon survival (viability), fertility, or attraction between the sexes.

- As these children mature, they will choose their mates at random from those of the opposite sex regardless of their tasting abilities.
- Matings between any two genotypes can then be predicted solely on the basis of the frequency of those genotypes in the population.
- Table 1 shows matings in all possible combination.

Table 1: Types of random-mating combinations and their relative frequencies in a population containing .40TT, .40Tt, and .20tt genotypes

gametes	TT	Tt	tt
TT	0.16	0.16	0.08
Tt	0.16	0.16	0.08
tt	0.08	0.08	0.04

Table 2: Relative frequencies of the different kinds of offspring produced by the matings

Parents		Offspring ratio		
Type of mating	Frequency of mating	TT	Tt	tt
$TT \times TT$	0.16	all(.16)		
$TT \times Tt$	0.32	1/2(.32)	+1/2(.32)	
$TT \times tt$	0.16		all (.16)	
$Tt \times Tt$	0.16	1/4(.16)	+1/2 (.16)	+1/4(.16)
$Tt \times tt$	0.16		1/2(.16)	+1/2(.16)
$tt \times tt$	0.04			all (.04)

- Note that although the frequencies of genotypes have been altered by random mating, the gene frequencies have not changed.
- For the *T* gene frequency is equal to $.36 + 1/2(.48) = .60$, and the frequency of *t* is $.16 + 1/2(.48) = .40$.
- No matter what the initial frequencies of the three genotypes, the gene frequencies of the next generation will be the same as those of parental generation.

Assertion

1. Under conditions of random mating (*panmixis*) in a large population where all genotypes are equally viable, gene frequencies of a particular generation depend upon the gene frequencies of the previous generation and not upon the *genotype frequencies*.
2. The frequencies of different genotypes produced through random mating depend only upon the gene frequencies.

- After the first generation of random mating, genotype frequencies also remain stable. i.e., equilibrium.

Formal proof

Table 3: Mating combinations and frequencies of offspring produced under conditions of random mating when genotypic frequencies are p^2TT, $2pqTt$, and q^2tt

Parents		Offspring ratio		
Type of mating	Frequency of mating	TT	Tt	tt
$TT \times TT$	$p^2 \times p^2 = p^4$	p^4		
$TT \times Tt$	$2 \times p^2 \times 2pq = 4p^3q$	$2p^3q$	$2p^3q$	
$TT \times tt$	$2 \times p^2 \times q^2 = 2p^2q^2$		$2p^2q^2$	
$Tt \times Tt$	$2pq \times 2pq = 4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
$Tt \times tt$	$2 \times 2pq \times q^2 = 4pq^3$		$2pq^3$	$2pq^3$
$tt \times tt$	$q^2 \times q^2 = q^4$			q^4
	$p^2(p^2 + 2pq + q^2) + 2pq(p^2 + 2pq + q^2) + q^2(p^2 + 2pq + q^2) = p^2 + 2pq + q^2 = (p + q)^2 = 1$	$p^4 + 2p^3q + p^2q^2 = p^2(p^2 + 2pq + q^2) = p^2$	$2p^3q + 4p^2q^2 + 2pq^3 = 2pq(p^2 + 2pq + q^2) = 2pq$	$p^2q^2 + 2pq^3 + q^4 = q^2(p^2 + 2pq + q^2) = q^2$

Problem

3 A population is consisted of 200 plants. Out of them, 100 plants are of Aa, 50 plants are of AA and 50 plants are of aa genotypes. This is a random mating population and in this population the frequencies of these three genotypes are at H-W equilibrium state. After 5^{th} generations of random mating, plants having genotypes AA, Aa and aa are found in 500, 300 and 200 numbers respectively. Are they still in H-W equilibrium? Test the result with the help of χ^2 goodness of fit test.

Solution

3 Here, the population of 200 plants is stated to be in H-W equilibrium; we already have equilibrium frequencies. Hence a χ^2 test for would show whether or not both the populations are same or have diverged from H-W equilibrium state (i.e. observed frequncy of population after 5th generation is same or different than expected population frequency at initial condition). For facilitating comparison, we convert the given frequencies of observed genotypes (that of 5^{th} generation) to the add upto current population count (200 individual).

Thus observed frequencies are AA: 100; Aa: 60 and aa: 40.

Note, however, we commonly compute the expected frequency based on the expected ratios. Therefore it also imperative to show the expected frequency as the proportion of total count of observed frequency.

Now we construct contingency table, as shown in Table 4.

Table 4: 2x3 contingency table of frequency of genotypes at equilibrium generation and at 5th generation of mating

		Genotype frequency		
		Dominant (AA)	Homozygous dominant (Aa)	Recessive (aa)
Generation	1^{st}	100	50	50
	5^{th}	100	60	40

Here since the number of df is 2, we do not apply the Yate's correction. After computation, we find χ^2 = 2.02 with probability of 0.36 which is well within the confidence band of 0.95 to 0.05. We fail to reject the null hypothesis that two observations were taken from same populations. Thus, we conclude that even after 5^{th} generation of mating the population continues to be in HW equilibrium state.

Outline

- Two major types of process identified:
  1. Systematic: Predictable in both direction and in amount
  2. Dispersive: Predictable only in amount

Migration

Migration is important in small populations. It entails the entry of individuals into an existing population from outside. Because plants are sedentary, migration, when it occurs naturally, is via pollen transfer (gamete migration). The impact that this immigration will have on the recipient population will depend on the immigration rate and the difference in gene frequency between the immigrants and natives. Mathematically, $\Delta q = m(q_m - q_o)$, where Δq = the change in frequency of genes in the new mixed population, m = number of immigrants, q_m = the gene frequency of the immigrants, and q_o = the gene frequency of the bost. Plant breeders employ this process to change frequencies when they undertake introgression of genes into their breeding populations. The breeding implication is that for open-pollinated (outbreeding) species, the frequency of the immigrant gene may be low, but its effect on the host gene and genotypes could be significant.

Mutation

Natural mutations are generally rare. A unique mutation (non-recurrent mutation) would have little impact on gene frequencies. Mutations are generally recessive in gene action, but the dominant condition may also be observed. Recurrent mutation (occurs repeatedly at a constant frequency) may impact gene frequency of the population. Natural mutations are of little importance to practical plant breeding. However, breeders may artificially induce mutation to generate new variability for plant breeding.

Selection

- Selection is the most important process by which plant breeders alter population gene frequencies. Its effect is to change the mean value of the progeny population from that of the parental population. This change may be greater or lesser than the population mean, depending on the trait of interest. For example, breeders aim for higher yield but may accept and select for less of a chemical factor in the plant that may be toxic in addition to the high yield. For selection to succeed:

1. there must be phenotypic variation for the trait to allow differences between genotypes to be observed;
2. the phenotypic variation must at least be partly genetic.

Random genetic drift

- Nondirectional forces that arises from variable sampling of the gene pool each generation is known as random genetic drift.
- It is caused by the fact that real population are limited in size rather than infinite, so that gene-frequency changes occur because of sampling errors.
- For instance, if only a few parents are chosen to begin a new generation, such a small sample of genes may deviate widely from the gene frequency of the previous generation.
- The extent of the deviation in both cases can be measured mathematically by the standard deviation of a proportion ($\sigma = \sqrt{\frac{pq}{N}}$). Here p is the frequency of one allele, q of the other, and N the number of genes sampled.
- For diploid parents, each carrying two genes, $\sigma = \sqrt{\frac{pq}{2N}}$, where N is the number of actual parents.
- For example, if we begin with a large diploid population, where $p = q = .5$, and continue this population each generation by using 5000 parents, then $\sigma = \sqrt{(.5)(.5)/10000} = \sqrt{.000025} = 0.05$. The values of such populations will therefore fluctuate mostly (68% of the time), around $.5 \pm .005$, or between 0.495 and 0.505. On the other hand, a choice of only two parents as “founders” will produce a standard deviation of $\sqrt{(.5)(.5)/4} = 0.25$ or values of $.5 \pm .25$ (.25 to .75).

Effect of mating system on selection

1. Random mating
2. Non random mating:

- Genetic assortative mating: mating occurs such that the mating pair has the same phenotype more often than would occur by chance.
- Phenotypic assortative mating: the breeder selects and mates individuals on the basis of their phenotypic resemblance to each other compared to the rest of the population.
- Disassortative mating: may be genetic or phenotypic; entails mating individuals that are less closely related than they would under random mating (genotypic) or breeder may select individuals with contrasting phenotypes for mating (phenotypic).

Outline

Heterosis (Hybrid vigor)

- Hybrid vigor may be defined as the increase in size, vigor, fertility, and overall productivity of a hybrid plant over the mid-parent value (average performance of the two parents).
- It is calculated as the difference between the crossbred and inbred means:

$$\text{Hybrid vigour} = \frac{F_1 - \frac{(P_1+P_2)}{2}}{\frac{(P_1+P_2)}{2}}$$

- The estimate is usually calculated as a percentage.
- The synonymous term, heterosis, was coined by G.H. Shull.
- Advantageous hybrid vigor is observed more frequently when breeders cross parents that are genetically diverse; When two inbred lines of outbred species are crossed.
- The practical definition of heterosis is hybrid vigor that greatly exceeds the better or higher parent in a cross.
- Hybrid breeding in maize quadrupled yields of maize in US between 1930s and 1970s.

Inbreeding depression

- Inbreeding depression is reduction in fitness as a direct result of inbreeding.
- In theory, the heterosis observed on crossing is expected to be equal to the depression upon inbreeding, considering a large number of crosses between lines derived from a single base population.
- In practice, plant breeders are interested in heterosis expressed by specific crosses between selected parents, or between populations that have no known common origin.
- Reduction in fitness is usually manifested as a reduction in vigor, fertility, and productivity.
- The effect is more severe in the early generations (5-8).
- Plants including onions, sunflower, cucurbits, and rye are more tolerant of inbreeding with minimal consequences of inbreeding depression.
- Plants such as alfalfa and carrot are highly intolerant of inbreeding.

- Inbreeding is measured by the coefficient of inbreeding (F), which is the probability of identity of alleles by descent. The range of F is zero (no inbreeding; random mating) to one (prolonged selfing).
- An unfit (deleterious) recessive allele is fairly quickly reduced in frequency but declines slowly thereafter.
- On the other hand, an unfit dominant allele is rapidly eliminated from the population, while an intermediate allele is reduced more rapidly than a recessive allele because the former is open to selection in the heterozygote.
- The consequence of these outcomes is that unfit dominant or intermediate alleles are rare in cross-breeding populations, while unfit recessive alleles persist because they are protected by their recessiveness.

- A measure of Inbreeding Depression is obtained through generational mean analysis. It is calculated as:

$$\text{Inbreeding depression (ID)} = \frac{F_1 - F_2}{F_1} \times 100\%$$

- In Figure 1 (a) there is no inbreeding because there is no common ancestral pathway to the individual, A (i.e., all parents are different).
- However, in Figure 1 (b) inbreeding exists because B and C have common parents (D and E), that is, they are full sibs.
- To calculate the amount of inbreeding, the standard pedigree is converted to an arrow diagram, as shown in 1 (c).
- Each individual contributes 1/2 of its genotype to its offspring. The *coefficient of relationship* (R) is calculated by summing up all the pathways between two individuals through a common ancestor as: $R_{BC} = \sum \left(\frac{1}{2}\right)^{s}$, where s is the number of steps (arrows) from B to the common ancestor and back to C. For example, B and C probably inherited $(1/2)(1/2) = 1/4$ of their genes in common through ancestor D. Similarly, B and C probably inherited 1/4 of their genes in common through ancestor E.
- The coefficient of relationship between B and C, as a result of common ancestry, is hence $R_{BC} = 1/4 + 1/4 = 1/2 = 50\%$

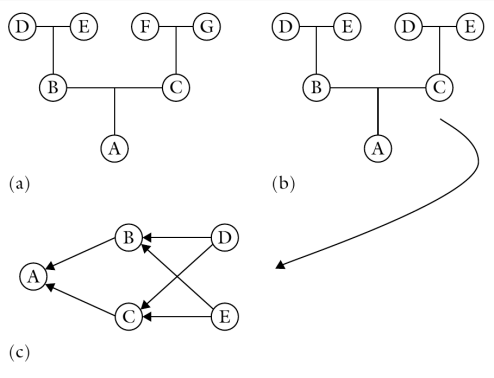

Figure 1: Pedigree diagrams can be drawn in the standard form (a, b) or converted to into an arrow diagram (c).

Genetic basis of heterosis

- To explain the genetic basis for why fitness lost on inbreeding tends to be restored upon crossing, two theories have been proposed.
  - Dominance theory: C.G. Davenport in 1908 and later by I.M. Lerner,
  - Overdominance theory: Shull in 1908 and later by K. Mather and J.L. Jinks.
- A third theory, the mechanism of epistasis (non-allelic gene interactions), has also been proposed.

Biometrics of heterosis

1. Better parent heterosis (Heterobeltiosis)

$$Hybrid\ vigour = \frac{F_1 - Better\ parent}{Better\ parent}$$

2. Mid parent heterosis

$$Hybrid\ vigour = \frac{F_1 - \frac{(P_1+P_2)}{2}}{\frac{(P_1+P_2)}{2}}$$

3. Commercial heterosis

$$Hybrid\ vigour = \frac{F_1 - Commercial\ Hybrid}{Commercial\ Hybrid}$$

Types of hybrids

- Commercial applications of hybrid breeding started with a cross of two inbred lines (a single cross - AxB) and later shifted to the more economic double cross, ([AxB]x[CxD]) and then back to a single cross.
- Other parent combinations in hybrid development have been proposed, including the three-way cross ([AxB]xC) and modified versions of the single cross, in which closely related crosses showed that the single cross was superior in performance to the other two in terms of average yield.
- However, it was noted also that the genotype x environment interaction (hybrid x environment) variability was more than twice that for the double crosses, while the mean variability for the three-way cross being intermediate.

- This indicated that the single crosses were more sensitive or responsive to environmental conditions than the other crosses.
- Whereas high average yield is important to the producer, consistency in performance across years and locations (i.e., yield stability) is also important.
- Double and three-way crosses have a more genetically divergent population for achieving buffering.
- Today commercial hybrids are predominantly single cross, of best combining parental inbred lines.
- For outline of mating scheme, See Lecture 7 on “Hybridization techniques and its consequences” (Course: Introductory plant breeding, 4^{th} semester, BScAg).

Outline

Problem

In blackgram, grain yield of parents (P_1 and P_2) their F_1 and F_2 progenies are given below:

Parent 1	Parent 2	F1 hybrid	F2 progeny
19	23	29	15

Calculate average heterosis, heterobeltiosis and inbreeding depression.

Solution

$$\text{Mid parent heterosis} = \frac{F_1 - MP}{MP} \times 100\%$$

$$\text{Here, Value of } F_1 = 29.38$$

$$\text{Mean of parents (MP)} = \frac{18.9 + 22.69}{2} = 20.81$$

$$\text{Mid parent heterosis} = \frac{29.38 - 20.81}{20.81} \times 100\% = 41.12\%$$

$$\text{Heterobeltiosis} = \frac{F_1 - BP}{BP} \times 100\% = \frac{29.38 - 22.69}{22.69} \times 100\% = 29.48\%$$

$$\text{Inbreeding depression} = \frac{F_1 - F_2}{F_1} \times 100\% = \frac{29.38 - 15.18}{15.18} \times 100\% = 48.33\%$$

Outline

References