Gene and alleles and equilibrium of population

Deependra Dhakal

2019-06-12

Gene and genotypes

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.

Mechanisms by which alleles at same locus changes

1. By origin: Same locus but different chromosome.
2. By state: It is indicative of the context they are put in. i.e. DNA sequence or amino acid sequence. Same amino acid sequence in some alleles may arise due to different DNA sequences (Redundancy of genetic code).
3. By descent: In practice, we are often concerned with relatively short time in past and are content to say that two alleles differ by descent if they do not share common ancestor in say, the past 10 generations. Two alleles different by descent may or may not be different by state because of mutation.

Converse of the mechanisms which cause differences in alleles are termed as identical by origin, stage or descent. Diploid individuals are said to be heterozygous at a locus if two alleles at that locus are different by state. If we are studying proteins, we may call an individual homozygous at a locus when the protein sequence of the two alleles are identical, even if their DNA sequences differ.

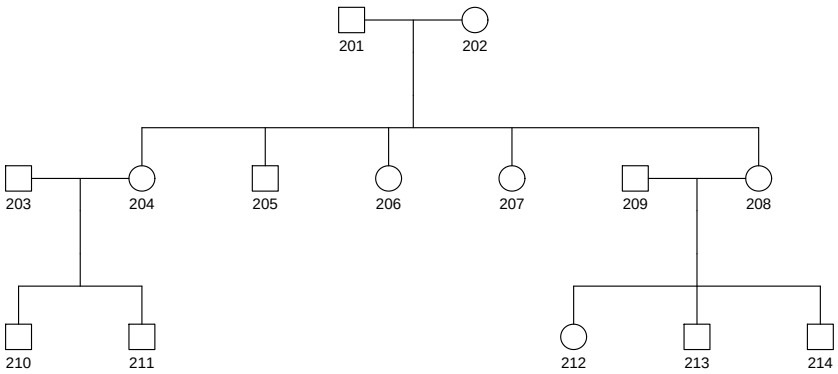

Figure 1: An example of pedigree tree showing single pedigree family with 14 subjects

Kreitman's sample

Kreitman's sample contain II alleles that differ by origin. How many alleles differ by state? If we were interested in the full DNA sequence, the sample

contains six alleles that differ by state. If we were interested in the proteins, then sample contains only two alleles that differ by state. Of the two protein alleles, the one with a lysine at position 192 makes up $\frac{6}{10} = 0.55$ of the alleles. The usual way to say this is that the allele frequency is an estimate of the population allele frequency. It's not a particularly precise estimate because of the small sample size. A rough approximation to the 95% confidence interval for a proportion is,

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where $\hat{p}$ is the estimate of proportion, 0.55 in our case and n is the sample size. Thus, the probability that the populatoin allele frequency falls within the interval (0.26, 0.84) is 0.95. If more precise estimate is needed, the sample size would have to be increased.

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,

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

The frequency of i^th allele is

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

Problems

1. How many different genotypes are there at a locus with n alleles that differ by state?

When there are n alleles, there are n homozygous genotypes, $A_iA_i, i = 1, 2, ..., n$. If we first view an A_iA_j heterozygote as distinct from an A_jA_i heterozygote, there are $n(n-1)$ such heterozygotes. The actual number of heterozygotes will be one half this number, or $\frac{n(n-1)}{2}$. Thus, the total number of genotypes is $\frac{n+n(n-1)}{2} = \frac{n(n+1)}{2}$.

1. Derive the hardy weinberg law for a sex-linked locus. Let the initial frequency A, in female be p_f and in males be p_m. Follow the two allele frequencies in successive generations untill you understand the allele frequency dynamics. Then, jump ahead and find the equilibrium genotype frequencies in females and males. Finally, graph the male and female allele frequencies over several generations for a population that is started with all A_1A_1 females ($p_f = 1$) and A_2 males ($p_m = 0$)

As males get their X-chromosomes from their mother, the frequency of A_1 in males is always equal to the frequency in females in the previous generation. As a female gets one X from her mother and one from her father, the allele freq in females is always the average of the male and female frequencies in the previous generation. Thus, the allele frequencies over the first three generation are as follows.

Generation	Females	Males	Female-male
1	p_f	p_m	$p_f - p_m$
2	$\frac{p_f+p_m}{2}$	p_m	$-\frac{p_f-p_m}{2}$
3	$\frac{p_f+\frac{p_f+p_m}{2}}{2}$	$\frac{p_f+p_m}{2}$	$\frac{p_f-p_m}{2}$

Two important things emerge from the table. First, the overall allele frequecy,

$$p = \frac{2}{3}p_f + \frac{1}{3}p_m$$

does not change over time. (Convince yourself that this is so by calculating p in generations 2 and 3). Second, the difference between the allele frequencies in females and males is halved each generation, as recorded in table. Taken together, these two observations show that eventually the allele frequencies in male and females will converge to p. At that time, the genotype frequencies in females will be at Hardy-Weinberg frequencies.

Example case

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.

$\rightarrow$ 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 1.

Table 1: 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 = 2.020202 with probability of 0.3641822 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.