****

FREQUENCY OF RECESSIVE TRAITS IN A POPULATION IN WHICH RECESSIVE HOMOZYGOTES ARE NOT ALLOWED TO REPRODUCE

Contributor: Richard F. Baum*

****

Mathematics: elementary probability, proof by mathematical induction (Math. 0)

Abstract: A theorem is proved about the decrease in frequency of recessive genes under selection against the recessive homozygote.

*Based upon: Dahlberg, Gunner. Mathematical Methods for Population Genetics. New York: Interscience Publishers, 1948.

Suppose that we are studying a characteristic of a species, which is determined by a single pair of genes at a given locus. If we assume complete dominance and neglect the effect of the environment, we may expect two phenotypes, corresponding to the recessive homozygote and the heterozygote or dominant homozygote. Let a denote the recessive gene and A, the dominant gene. Furthermore, let C_a represent the phenotype corresponding to the recessive homozygote and C_A, the phenotype associated with the dominant homozygote or heterozygote.

Suppose that an individual with phenotype C_a is prevented from entering into reproduction, as would be the case if the individual died before sexual maturity. Assume that:

(a) At a given moment genotypes aa, aA, and AA have respective proportions s_1, t_1, u_1;

(b) these proportions are the same for males and females;

(c) mating is random among persons of genotypes aA and AA, but persons of genotype aa do not mate.

Then the probability that one partner in a mating will donate gene a to an offspring is

$$r_1 = \left(\frac{1}{2}\right)\left(\frac{t_1}{t_1 + u_1}\right) ,$$

where $t_1(t_1 + u_1)$ is the probability that the partner is aA and 1/2 is the probability that such a parent will donate gene a. Then for the proportions in the succeeding generation, we have

$$(1) \qquad s_2 = r_1^2 , \qquad t_2 = 2r_1(1 - r_1) , \qquad u_2 = (1 - r_1)^2 ,$$

where the factor 2 in the expression for t_2 accounts for the fact that either parent can contribute gene a. Observe that in generation 2 the probability that a parent will transmit gene a is

$$r_2 = \frac{t_2}{2(t_2 + u_2)} = \frac{2r_1(1 - r_1)}{2[2r_1(1 - r_1) + (1 - r_1)^2]}$$

which can be simplified to yield

(2) $$r_2 = \frac{r_1}{r_1 + 1} .$$

Claim:

The probability r_N that a parent in the Nth generation will transmit gene a is

(3) $$r_N = \frac{r_1}{1 + (N - 1)r_1} ,$$

where r is the corresponding probability in the first generation.

Proof:

Proof will be by mathematical induction.

For N = 1, equation (3) reduces to

$$r_1 = \frac{r_1}{1 + or_1} ;$$

and for N = 2, equation (3) reduces to equation (2). For the induction step, assume that equation (3) is valid for N = k so that

(4) $$r_k = \frac{r_1}{1 + (k - 1)r_1} .$$

Next, we observe that equation (2) holds for any two consecutive generations, so that, in particular,

(5) $$r_{k+1} = \frac{r_k}{1 + r_k} .$$

Now, combining (4) and (5), we get

$$r_{k+1} = \frac{\dfrac{r_1}{1 + (k - 1)r_1}}{\dfrac{1 + r_1}{1 + (k - 1)r_1}}$$

$$= \frac{r_1}{1 + (k - 1)r_1 + r_1}$$

$$= \frac{r_1}{1 + kr_1} \, .$$

This is equation (3) for $N = k + 1$. Therefore, the induction step is verified and the proof is complete.

As a consequence of equation (3) and the fact that equation (1) is valid for any two consecutive generations, we see that

(6) $$s_N = (r_{N-1})^2 = \frac{r_1^2}{(1 + (N - 2)r_1)^2} \, ,$$

(7) $$t_N = 2r_{N-1}(1 - r_{N-1}) = \frac{2r_1(1 + (N - 3)r_1}{(1 + (N - 2)r_1)^2} \, ,$$

and

(8) $$u_N = (1 - r_{N-1})^2 = \frac{(1 + (N - 3)r_1)^2}{(1 + (N - 2)r_1)^2} \, .$$

Consider as an example a case where $s_1 = .25$, $t_1 = .5$, and $u_1 = .25$. Then

$$r_1 = \frac{t_1}{2(t_1 + u_1)} = \frac{.5}{2(.5 + .25)} = \frac{1}{3} \, .$$

Now using equations (6), (7), and (8) we derive Table 1 below. Note that:

Table 1

Taken from: Dahlberg, Gunner: Mathematical Methods for Population Genetics. New York: Interscience, 1948.

Generation	aa	aA	AA
1	0.2500	0.5000	0.2500
2	0.1111	0.4444	0.4444
3	0.0625	0.3750	0.5625
4	0.0400	0.3200	0.6400
5	0.0278	0.2778	0.6944
6	0.0204	0.2449	0.7347
7	0.0156	0.2188	0.7656
8	0.0123	0.1975	0.7901
9	0.0100	0.1800	0.8100
10	0.0083	0.1653	0.8264
20	0.0023	0.0907	0.9070
30	0.0010	0.0624	0.9365
40	0.0006	0.0476	0.9518
50	0.0004	0.0384	0.9612
100	0.0001	0.0196	0.9803

(A) The individuals of type aa are eliminated very rapidly. By the fourth generation their frequency has dropped to 4%. However, if we begin with $r_1 = 0.1\%$ (30th generation), in ten generations, their frequency will have dropped only about half, to 0.06%.

(B) The effect on heterozygotes, i.e., individuals of type aA, is considerably weaker. Between the fourth and ninth generation, the individuals of type aa are decreased from 4% to 1%, while the aA individuals are decreased only from 32% to 18%. After 100 generations, there are only 0.01 individuals of type aA, i.e., 1.96%.

It is clear from equations (6) and (7) that s_N and t_N both tend to zero as N becomes large and hence that eventually gene a will disappear from the population. The fact that s_N tends to zero much more rapidly than t_N means that overt appearance of gene a (i.e., persons of type aa) will probably cease long before the gene itself is entirely gone.