LOSS OF HETEROZYGOSITY IN POPULATIONS UNDER MIXED RANDOM MATING AND SELFING

By G.L. Ghai

Institute of Agricultural Research Statistics, New Delhi

In a population being inbred with one system of mating, the coefficient of inbreeding depends upon the number of factors or loci, recombination values and the system of mating employed. The expression for the coefficient of inbreeding does not present any difficulty when the various gene pairs involved are segregating independently. It can be immediately worked out from the inbreeding coefficient obtained for the single locus case *i e*. if F is the inbreeding coefficient in the latter case, then it would be F^k when k independently segregating loci are considered. But in plant populations of interest to the breeder complete cross fertilization at random or complete inbreeding such as selfing seldom obtains. There are crops like wheat and rice which are highly but not completely self-fertilized, crops like cotton which are moderately self-fertilized, and crops like various 'brassica' species or maize which are largely cross fertilized. It would, therefore, be of interest to study populations where a mixture of breeding systems obtains in varying proportions. The present investigation deals with the study of loss in heterozygosity in a population when in a certain proportion random mating obtains and the remaining proportion is self-fertilized.

The initial population is assumed to be panmictic *i.e.*

$$II\,[p^2(A)\,.\,AA + 2p(A)p(a)\,.\,Aa + p^2(a)\,.aa] \qquad \ldots(1)$$

where $p(A)$ and $p(a)$ etc. are the frequencies of the alleles A and a respectively of the gene pair A-a. In each generation a constant proportion, say y, of the population is mated at random and the remaining proportion, say $x(x+y=1)$, self-fertilized. The various gene pairs involved are further assumed to be segregating independently and viability and fertility differences are absent.

Let $H_i^{(n)}(i=0,1,2, \ldots k)$ denote the frequency of i— factor heterozygotes *i.e.* sum of the frequencies of genotypes in which i

factors of the k factors A-a, B-b, C-c...considered are in heterozygous state and the remaining (k-i) factors are in homozygous, in the nth generation of mixed random mating and selfing. Thus $H_1^{(n)}$, when only two gene pairs A-a and B-b are involved, will be the sum of frequencies of *AaBB, Aabb, AABb and aaBb* genotypes each of which is heterozygous with respect to one locus and homozygous for the other. $H_1^{(o)}$ are the frequencies of the corresponding heterozygotes in the initial population.

Single locus A-a : The frequency of single heterozygotes, $H_1^{(n)}$, in the nth generation of mixed random mating and selfing is obtained as

$$H_1^{(n)} = y \,.\, 2p(A)p(a) + (x/2)H_1^{(n-1)}$$

$$= \left[\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right]H_1^{(o)},$$

since $\quad 2p(A)p(a) = H_1^{(0)}$

Defining the coefficient of inbreeding, F_n, in the nth generation as the loss in heterozygosity relative to that in the initial population, we have

$$F_n = 1 - \frac{H_1^{(n)}}{H_1^{(0)}}$$

$$= \frac{x}{2-x}\left\{1-\left(\frac{x}{2}\right)^n\right\} \qquad \dots (2)$$

In the limiting case when $n \to \infty$

$$F_\infty = \frac{x}{2-x}$$

Two loci A-a, B-b : The following relations can be easily obtained for the frequencies of double heterozygotes and single heterozygotes in the nth generation.

$$H_2^{(n)} = y.\ 4p(A)p(a)p(B)p(b) + \left(\frac{x}{4}\right)H_2^{(n-1)}$$

$$= \left[\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)+\frac{4y}{4-x}\right]H_2^{(0)},$$

Since $4p(A)p(a)p(B)p(b) = H_2^{(0)}$,

and $H_1^{(n)} = y.\ R_1^{(n)} + \left(\frac{x}{2}\right) H_1^{(n-1)} + \left(\frac{x}{2}\right) H_2^{(n-1)}$

where $R_1^{(n)}$ is the frequency of single heterozygotes in the nth generation when there is only random mating in the previous generation and is given by

$$R_1^{(n)} = 2\begin{bmatrix} p(A)p(a)\{p^2(B)+p^2(b)\} + \\ p(B)p(b)\{p^2(A)+p^2(a)\} \end{bmatrix}$$

$$= H_1^{(0)}, \text{ for all } n.$$

Therefore,

$$H_1^{(n)} = y\ R_1^{(n)} + \left(\frac{x}{2}\right)H_1^{(n-1)} + \left(\frac{x}{2}\right)H_2^{(n-1)}$$

$$= \left[\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right]\left(H_1^{(0)} + 2H_2^{(0)}\right)$$

$$-2\left[\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)+\frac{4y}{4-x}\right]H_2^{(0)}$$

Let $H^{(n)}$ denote the frequency of all heterozygotes in the nth generation, then

$$H^{(n)} = H_2^{(n)} + H_1^{(n)}$$

$$= \left[\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right]\left(H^{(0)}+H_2^{(0)}\right)$$

$$-\left[\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)+\frac{4y}{4-x}\right]H_2^{(0)}$$

Hence, the coefficient of inbreeding, F_n, will be given by

$$F_n = 1 - \frac{H^{(n)}}{H^{(0)}}$$

$$= 1 - \left[\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right]$$

$$-\left[\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right.$$

$$\left.-\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)-\frac{4y}{4-x}\right]\frac{H_2^{(0)}}{H^{(0)}} \qquad \ldots(3)$$

It will be seen from (3) that under mixed random mating and selfing, the inbreeding coefficient depends upon the constitution of the initial population.

When the initial population consists of double heterozygotes only, (3) reduces to

$$F_n = 1 - 2\left[\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right]$$

$$+\left[\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)+\frac{4y}{4-x}\right]$$

$$= 2\left(1-\frac{2y}{2-x}\right)\left\{1-\left(\frac{x}{2}\right)^n\right\}$$

$$-\left(1-\frac{4y}{4-x}\right)\left\{1-\left(\frac{x}{4}\right)^n\right\} \qquad \ldots(4)$$

In the limiting case when $n \to \infty$

$$F_\infty = \frac{x\,(2+x)}{(2-x)\,(4-x)}.$$

When $x=1$, *i.e.* complete selfing, (4) reduces to the known result, *viz.*,

$$F_n\,(x=1) = \left(1-\frac{1}{2^n}\right)^2.$$

Thus under mixed systems of breeding namely random mating and selfing, the coefficient of inbreeding when two gene pairs involved are segregating independently, cannot be expressed in terms of the inbreeding coefficient obtained for a single locus as is the case when population is completely selfed in successive generations. Therefore, the expression for the inbreeding coefficient when more than one locus is involved, cannot be predicted from that obtained for single locus even though the loci may be segregating independently and hence requires further study.

Three loci A-a, B-b, C c : With three independently segregating loci, the following relations hold for the frequencies of various types of heterozygotes in the nth generation.

$$H_3^{(n)} = y.\,8\,p(A)\,p(a)\,p(B)\,p(b)\,p(C)\,p(c) + \frac{x}{8}H_3^{(n-1)}$$

$$= \left[\left(\frac{x}{8}\right)^n\left(1-\frac{8y}{8-x}\right)+\frac{8y}{8-x}\right]H_3^{(0)},$$

since $8\,p(A)\,p(a)\,p(B)\,p(b)\,p(C)\,p(c) = H_3^{(0)}$;

$$H_2^{(n)} = y\,.\,R_2^{(n)} + \left(\frac{x}{4}\right)H_2^{(n-1)} + 6\left(\frac{x}{16}\right)H_3^{(n-1)}$$

and $$H_1^{(n)} = y\,.\,R_1^{(n)} + \left(\frac{x}{2}\right)H_1^{(n-1)} + \left(\frac{x}{2}\right)H_2^{(n-1)} + 12\left(\frac{x}{32}\right)H_3^{(n-1)}$$

where $R_2^{(n)}$ and $R_1^{(n)}$ are the frequencies of double heterozygotes and single heterozygotes in the nth generation obtained by random mating in the $(n-1)$th generation and are given by

$$R_2^{(n)} = \sum 4\,p(A)\,p(a)\,p(B)\,p(b)\left\{p^2\,(C)+p^2\,(c)\right\}$$

$$= H_2^{(0)}, \text{ for all } n$$

$$R_1^{(n)} = \sum 2\,p(A)\,p(a)\,p^2(B)\,p^2(C)$$

$$= H_1^{(0)}, \text{ for all } n.$$

Therefore,

$$H_2^{(n)}=y_2R^{(n)}+\left(\frac{x}{4}\right)H_2^{(n-1)}+6\left(\frac{x}{16}\right)H_3^{(n-1)}$$

$$=\left[\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)+\frac{4y}{4-x}\right]H_2^{(0)}$$

$$+3\left[\left\{\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)+\frac{4y}{4-x}\right\}\right.$$

$$\left.-\left\{\left(\frac{x}{8}\right)^n\left(1-\frac{8y}{8-x}\right)+\frac{8y}{8-x}\right\}\right]H_3^{(0)}$$

and

$$H_1^{(n)}=y\,R_1^{(n)}+\left(\frac{x}{2}\right)H_1^{(n-1)}+\left(\frac{x}{2}\right)H_2^{(n-1)}+12\left(\frac{x}{32}\right)H_3^{(n-1)}$$

$$=\left[\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right]H_1^{(0)}$$

$$+2\left[\left\{\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right\}\right.$$

$$\left.-\left\{\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)+\frac{4y}{4-x}\right\}\right]H_2^{(0)}$$

$$+3\left[\left\{\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right\}\right.$$

$$-2\left\{\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)+\frac{4y}{4-x}\right\}+\left\{\left(\frac{x}{8}\right)^n\left(1-\frac{8y}{8-x}\right)\right.$$

$$\left.\left.+\frac{8y}{8-x}\right\}\right]H_3^{(0)}.$$

Therefore $H^{(n)}$, the frequency of heterozygotes in the nth generation will be given by

$$H^{(n)}=H_3^{(n)}+H_2^{(n)}+H_1^{(n)}$$

$$=\left[\left(\frac{x}{2}\right)^n\left(1-\frac{2y}{2-x}\right)+\frac{2y}{2-x}\right](H^{(0)}+H_2^{(0)}+2H_3^{(0)})$$

$$-\left[\left(\frac{x}{4}\right)^n\left(1-\frac{4y}{4-x}\right)+\frac{4y}{4-x}\right](H_2^{(0)}+3H_3^{(0)})$$

$$+\left[\left(\frac{x}{8}\right)^n\left(1-\frac{8y}{8-x}\right)+\frac{8y}{8-x}\right]H_3^{(0)}.$$

Hence,

$$F_n = 1 - \frac{H^{(n)}}{H^{(0)}}$$

$$= \left\{1 - \left(\frac{x}{2}\right)^n\right\}\left(1 - \frac{2y}{2-x}\right) - \left[\left\{\left(\frac{x}{2}\right)^n\left(1 - \frac{2y}{2-x}\right) + \frac{2y}{2-x}\right\}\right.$$

$$\left. - \left\{\left(\frac{x}{4}\right)^n\left(1 - \frac{4y}{4-x}\right) + \frac{4y}{4-x}\right\}\right]\frac{H_2^{(0)}}{H^{(0)}}$$

$$- \left[2\left\{\left(\frac{x}{2}\right)^n\left(1 - \frac{2y}{2-x}\right) + \frac{2y}{2-x}\right\} - 3\left\{\left(\frac{x}{4}\right)^n\left(1 - \frac{4y}{4-x}\right)\right.\right.$$

$$\left.\left. + \frac{4y}{4-x}\right\} + \left\{\left(\frac{x}{8}\right)^n\left(1 - \frac{8y}{8-x}\right) + \frac{8y}{8-x}\right\}\right]\frac{H_3^{(0)}}{H^{(0)}} \qquad ...(5)$$

Considering the initial population to be constituted of triple heterozygotes only, (5) reduces to

$$F_n = 3\left(1 - \frac{2y}{2-x}\right)\left\{1 - \left(\frac{x}{2}\right)^n\right\} - 3\left(1 - \frac{4y}{4-x}\right) \times$$

$$\left\{1 - \left(\frac{x}{4}\right)^n\right\} + \left(1 - \frac{8y}{8-x}\right)\left\{1 - \left(\frac{x}{8}\right)^n\right\} \qquad ...(6)$$

In the limiting case, when $n \to \infty$

$$F_\infty = \frac{x\,(8 + 12x + x^2)}{(2-x)\,(4-x)\,(8-x)}.$$

When $x=1$, *i.e.* complete selfing, (6) reduces to the already known result, *viz.*,

$$F_n\,(x=1) = \left(1 - \frac{1}{2^n}\right)^3$$

k-loci. We have so far obtained the inbreeding coefficient under mixed random mating and selfing for one, two, and three independently segregating loci. These are given in equations (2), (4) and (6). From these equations it is possible to obtain an expression for inbreeding coefficient in general when more than one locus is involved. Hence, the coefficient of inbreeding in the general case of several independently segregating loci, say k, would be

$$F_n = \sum_{i=1}^{k} {}^kC_i\,(-1)^{i+1}\left(1 - \frac{y}{1-c_{ii}}\right)(1 - c_{ii}{}^n) \qquad ...(7)$$

where $\quad c_{ii} = (x/2^i)$.

This can also be written as

$$F_n = \sum_{i=0}^{k} {}^kC_i \, (-1)^i \left[\frac{y}{1-c_{ii}} + c_{ii}^n \left(1 - \frac{y}{1-c_{ii}}\right) \right]$$

$$= x^n \left(1 - \frac{1}{2^n}\right)^k + \sum_{i=0}^{k} {}^kC_i \, (-1)^i \, y \left(\frac{1-c_{ii}^n}{1-c_{ii}} \right)$$

In the limiting case, when $n \rightarrow \infty$

$$F_\infty = \sum_{i=0}^{k} {}^kC_i \, (-1)^i \, \frac{y}{1-c_{ii}}$$

$$= \sum_{i=0}^{k} {}^kC_i \, (-1)^i \, \frac{2^i y}{2^i - x}.$$

The loss in heterozygosity (F_n) in different generations of mixed random mating and selfing has been examined upto ten factors in the attached table for various amounts of self-fertilization. The values for $x = 1{\cdot}0$ correspond to the case of complete selfing. It would be seen from the table that loss in the heterozygosity increases with increasing amount of self-fertilization in the population, with the increase in the number of generations, and decreases with the increase in the number of segregating loci. But heterozygosity is not completely lost except under complete selfing. The maximum loss in heterozygosity that can occur under continued mixed random mating and selfing in the case of single factor segregation is of the order of 5 per cent when selfing percentage is 10 which rises to 82 per cent when selfing percentage is 90. These values are of the order of 3% and 76% in the case of two factors, ·9% and 70% in the case of four factors, ·2% and 64% in the case of eight factors, and ·1% and 62% in the case of ten factors. It may also be observed from the table that for a given amount of self-fertilization the maximum value is reached very rapidly and there is relatively little increase thereafter.

SUMMARY

In the present investigation a study has been made with respect to loss in heterozygosity in populations under mixture of breeding systems namely random mating and selfing. When the population is completely selfed the loss in heterozygosity in the case of several

Table : Loss of heterozygosity relative to that in the initial population (F_n) in different generations (n) of mixed random mating and selfing upto ten factors.

$x=$	·10	·30	·50	·70	·90	1·00
$k=1$						
$n=1$	·0500	·1500	·2500	·3500	·4500	·5000
$n=2$	·0525	·1725	·3125	·4725	·6525	·7500
$n=3$	·0526	·1758	·3281	·5153	·7436	·8750
$n=4$	·0526	·1763	·3320	·5303	·7846	·9375
$n\to\infty$	·0526	·1764	·3333	·5384	·8181	1·0000
$k=2$						
$n=1$	·0250	·0750	·1250	·1750	·2250	·2500
$n=2$	·0281	·1031	·2031	·3281	·4781	·5625
$n=3$	·0283	·1086	·2285	·3977	·6262	·7656
$n=4$	·0283	·1095	·2355	·4249	·7005	·8789
$n\to\infty$	·0283	·1096	·2380	·4405	·7653	1·0000
$k=4$						
$n=1$	·0062	·0187	·0312	·0437	·0562	·0625
$n=2$	·0087	·0416	·0947	·1681	·2619	·3164
$n=3$	·0090	·0488	·1284	·2606	·4585	·5861
$n=4$	·0090	·0503	·1400	·3054	·5808	·7724
$n\to\infty$	·0090	·0507	·1447	·3343	·7021	1·0000
$k=8$						
$n=1$	·0003	·0011	·0019	·0027	·0035	·0039
$n=2$	·0013	·0098	·0260	·0498	·0814	·1001
$n=3$	·0015	·0164	·0564	·1333	·2589	·3436
$n=4$	·0016	·0184	·0722	·1941	·4250	·5967
$n\to\infty$	·0016	·0189	·0801	·2441	·6382	1·0000
$k=10$						
$n=1$	·0000	·0002	·0004	·0006	·0008	·0009
$n=2$	·0006	·0052	·0143	0278	·0456	·0563
$n=3$	·0008	·0108	·0401	·0987	·1964	·2630
$n=4$	·0008	·0129	·0565	·1614	·3679	·5244
$n\to\infty$	·0008	·0135	·0655	·2196	·6182	1·0000

independently segregating loci can be easily worked out from that obtained for a single locus segregation. The present study in populations under mixed random mating and selfing reveals that this loss cannot be predicted from the study of a single locus when more than one locus is considered. The general expression for the loss in heterozygosity relative to that in the initial population in the nth generation

(F_n) due to mixed random mating and selfing has been obtained in the general case of several independently segregating loci.

ACKNOWLEDGEMENT

The author wishes to thank Dr. V.G. Panse, Statistical Adviser, Institute of Agricultural Research Statistics (I.C.A.R.), for his encouragement and guidance in the investigation.

REFERENCES

1. Bennet, J.H. and Binet, F.E. ... "Association between Mendelian factors with mixed selfing and random mating," *Heredity*, 1956, **10**, 51-55.

2. Fisher, R.A. ... *The Theory of Inbreeding*, Oliver and Boyd, Edinburgh, 1949.

3. Garber, M.J. ... "Approach to genotypic equilibrium with varying percentage of self-fertilization," *J. Heredity*, 1951, **42**, 299-300.

4. Ghai, G.L. ... "The genotypic composition and variability in plant populations under mixed self-fertilization and radom mating," *Iour. Ind. Socy. Agri. Stat.*, 1964, **16**, 93-125.

5. Kempthorne, O. ... *An Introduction to Genetic Statistics*, John Wiley and Sons, New York, 1957.

6. Li, C.C. ... *Population Genetics*, The University Press, Chicago, 1955.