Theor Appl Genet (1989) 77:867–872

Time of detection of recessive genes: effects of system of mating and number of examined individuals

E. Santiago

Departamento de Biología Funcional (Area de Genética), Universidad de Oviedo, E-33071 Oviedo, Spain

Received November 1, 1988; Accepted December 14, 1988
Communicated by L. D. Van Vleck

Summary. The elapsed time from the appearance of a single copy of a recessive gene in a population with separate sexes until the first detection of a recessive homozygote has been analyzed using simulation techniques. Several systems of mating frequently used in laboratory maintenance and artificial selection processes have been the subject of this study. The expected time of detection (T) in each system of mating is described by a function of number of parents (N) and the factor K that when multiplied by N gives the number of examined individuals per generation (e.g. $T=1.6+1.8\ (N/K)^{1/3}$ for the mass-mating system). Although the expected time of detection changes greatly depending on the system of mating and on the K value (the causes are discussed), both the time scale ($N^{1/3}$) and the coefficient of variation of the distribution of the times of detection (⅔) seem to be unaffected. The additional reduction of the effective population number caused by artificial selection on a heritable trait modifies the expected detection time in a minor way unless both high heritability and high selection intensity are involved.

Key words: Recessive genes – Mating systems – Detection time – Genetic hypotheses

Introduction

The discovery of homozygotic individuals for a recessive visible gene that was not previously detected in the population seems to be a common event in artificial selection processes. To determine whether the gene was present in the base population or if it is a product of a more recent mutation, the distribution of first detection times of initially single copy recessive alleles must be analyzed. The elapsed time until the detection of homozygotes of new mutant types is relevant from the point of view of the population genetics and the contribution of new recessive mutations with visible effects on the response to artificial selection. The origins of the genes responsible for rapid responses observed in selection lines have been frequently discussed (Hollingdale 1971; Frankham 1980; Yoo 1980).

This topic was examined by Robertson (1978), who studied the relationship between population size and the first occurrence of a homozygous recessive individual in a monoecious random-mating population, using transition matrix methods. If initially there is a single copy of the gene and only the N individuals that reproduce are examined, the distribution of times of first detection takes a near geometric form with a mean very close to $2N^{1/3}$ generations, a coefficient of variation of ⅔ and the 95% upper confidence limit near to 2.5 times the mean. Differential selection upon the heterozygote seems to have very little effect on the detection time of the gene. Karlin and Tavaré (1980, 1981 a, b) confirmed the time scale $N^{1/3}$ proposed by Robertson and examined a number of variations on the basic model by means of diffusion approximations. They tried to assess the effects of examining a greater number of individuals than those used as parents as to probability and time of detection. Direct selection on heterozygotes together with partial penetrance in heterozygotes were also dealt with.

A wide spectrum of systems of mating are used to maintain artificial populations. The different systems produce different rates of inbreeding, which may lead to differences in the time of detection as compared to the random mass-mating model studied by Robertson (1978). Also the examination of a greater number of individuals than those used as parents affects the time of detection in different ways, depending on the system of mating; this is because each system splits the population forming mating-groups or couples in a particular way. The purpose of this paper is to analyze the better known mating systems to provide assistance in testing the two alternative hypotheses dealing with the origin of the recessive gene. Mean time of detection will be expressed as a function of number of parents and number of examined individuals for each system of mating. Another goal was to try to assess the effect of the reduction of the effective population number due to artificial selection on a heritable trait (Robertson 1961).

Models and methods

This work was carried out using simulation techniques. Initially, one single copy of the recessive gene was present in a population with separated sexes. A constant number of individuals was examined in each generation to look for a recessive homozygote, and the parents that form the following generation were chosen from these (equal numbers of males and females were used as parents). The relative fitness of the heterozygote was considered equal to the dominant homozygote. The following mating systems were analyzed:

(1) Random mass-mating. The two gametes that form each individual were randomly sampled from the two gene pools of the population (male and female).

(2) Individual random mating and random examination. The parents were previously paired at random. The offspring were obtained at random from the different prefixed couples. Thus, the contribution of each family to all examined individuals follows a binomial distribution. The parents that formed the next generation were randomly sampled from the pool of examined individuals.

(3) Individual random mating and equal number of examined individuals per family. The uniform contribution of all of the families to all examined individuals was the only difference from mating system 2.

(4) Limited random mating. The population was split into groups before mating and the gametes were randomly combined within each group to obtain the offspring that were examined (equal number of individuals were examined per group). The parents that formed the next generation were sampled among the offspring of each group (equal number of parents was sampled from each group). These parents were randomly allocated to form new mating groups. The extreme case of two parents per group is equivalent to the mating system used to carry out the within-family selection.

(5) Circular mating. As in mating system 4, the population was split to mate. The groups were arranged in a sequential and circular form. The combination of gametes and the choice of parents were made in the same way as in mating system 4. The chosen males of a group were mated with the chosen females from the next group. This system is used in selection experiments to reduce the genetic drift of the population, although it produces a rapid decrease in average heterozygosity in the first generations (Kimura and Crow 1963; Robertson 1964).

For the first three mating systems, combinations of the following values were considered:

- Number of parents (N): 2, 4, 6, 8, 12, 16, 23, 32, 48, 64, 96, 128, 192, 256, 384, 512, 768.
- The factor (K) that when multiplied by N gives the number of examined individuals per generation: 1, 2, 3, 4, 5 and infinite times the number of parents.

In the limited and circular mating systems, the parameters were:

- Number of parents per group (Ng): 2, 4, 8, 16, 32 (this last value only in circular mating).
- Number of groups (G): 3, 6, 12, 24. For the limited system of mating with 2 parents per group, these cases were extended to the 17 G values from 1 to 384 corresponding to 17 N values, which are equivalent to those considered in the first three mating systems.
- The factor (K) that when multiplied by N gives the number of examined individuals per generation: 1, 4 and infinite times the number of parents.

The consequences of the additional reduction of effective population number due to artificial selection on heritable trait were analyzed with the third mating system. A genetic system of nine additive loci with identical effects and two alleles per locus controlled the trait. In the first generation, frequencies from 0.1 to 0.9 were assigned to the nine alleles with positive effects. One copy of a recessive gene on the tenth locus is assigned to a single individual. There was no linkage among the ten loci. In each generation, an equal number of individuals per family was measured for the trait and examined to detect recessive homozygotic individuals for the tenth locus. Combinations of the following parameters were considered:

- Number of parents: 10, 20, 40.
- Proportion of selected individuals: 0.5, 0.25, 0.1. These values are equivalent to the K values 2, 4 and 10, respectively.
- Heritability of the base population: 0, 0.2, 0.4, 0.6.

For each combination of mating system, number of parents and K value, 500 simulations were made from the appearance of the recessive gene until it was lost or until a recessive homozygote was detected. The mean detection time (in generations) was estimated only from those simulations where the gene was detected as homozygote. This implies that each mean is based on a different number of data (ranging from 28 to 250), which decreases as the population size increases. This, together with the increase in the variance of the detection times, also associated with population size, implies that the standard error of the estimate of the mean increases considerably as population size increases within the same regression. Because of this, the linear regressions were carried out giving each point a weight equal to the inverse of its error variance. Statistical tests of deviations from the regression lines were made combining the probabilities of the deviation tests from all points of the same line as described by Fisher (1958).

Results

Robertson (1978) suggested the order of magnitude of $N^{1/3}$ for the expected time of detection of a recessive gene in a random-mating population when only the N reproducing individuals are examined. To check the generality of this relation, the linear regressions of the mean detection times on $N^{1/3}$ values were calculated for each mating system and ratio of examined individuals (K). Figure 1 shows graphically the lines of regression for the mass-mating system separately for the different K values. The fitting to the linear regression form $a+bN^{1/3}$ is satisfactory for the mass-mating and for both individual mating systems when N is the number of parents of the whole population: deviations from regression were not significant (Table 1). However, it was not possible to describe

Table 1. The a and b coefficients ± standard errors (S.E.) of the regression lines ($a+b\,Ng^{1/3}$ for the circular mating system and $a+b\,N^{1/3}$ for the other systems), fitted to the observed means of detection times. Probabilities from test of significance of deviations from the regression lines are also given

Mating system	K	b ±S.E.	a ±S.E.	n of means	P
Mass-mating	1	1.81±0.09	1.62±0.21	17	0.52
	2	1.46±0.06	1.58±0.16	17	0.59
	3	1.26±0.05	1.56±0.13	17	0.36
	4	1.12±0.04	1.63±0.11	17	0.66
	5	1.04±0.03	1.69±0.10	17	0.92
	Inf	0.01±0.02	3.02±0.05	17	0.19
Individual random mating and random examination	1	2.08±0.08	1.07±0.19	17	0.46
	2	1.68±0.07	1.30±0.14	17	0.24
	3	1.56±0.06	1.30±0.12	17	0.29
	4	1.50±0.05	1.31±0.11	17	0.28
	5	1.50±0.04	1.02±0.12	17	0.31
	Inf	1.56±0.06	1.02±0.12	17	0.30
Individual random mating and equal examination	1	3.49±0.12	−0.22±0.26	17	0.18
	2	2.42±0.09	0.32±0.18	17	0.76
	3	1.88±0.08	0.97±0.16	17	0.94
	4	1.82±0.07	0.89±0.15	17	0.75
	5	1.84±0.07	0.67±0.14	17	0.31
	Inf	1.56±0.06	1.02±0.12	17	0.30
Limited 2 parents/group	1	3.49±0.12	−0.22±0.26	17	0.18
	4	2.91±0.08	−0.47±0.16	17	0.82
	Inf	2.82±0.08	−0.49±0.16	17	0.16
4 p/g	1	2.09±0.30	1.83±0.81	4	0.40
	4	1.70±0.21	1.50±0.53	4	0.20
	Inf	1.91±0.14	−0.41±0.36	4	0.11
8 p/g	1	1.86±0.31	1.61±1.02	4	0.80
	4	1.37±0.19	1.63±0.71	4	0.53
	Inf	0.98±0.12	1.19±0.38	4	0.89
16 p/g	1	2.21±0.39	0.50±1.13	4	0.77
	4	1.32±0.23	1.33±0.86	4	0.68
	Inf	0.64±0.13	1.59±0.40	4	0.31
Circular	1	2.23±0.26	4.80±0.49	5	0.78
	4	1.26±0.25	4.63±0.44	5	0.25
	Inf	0.01±0.02	4.86±0.19	5	0.80

Inf – infinite

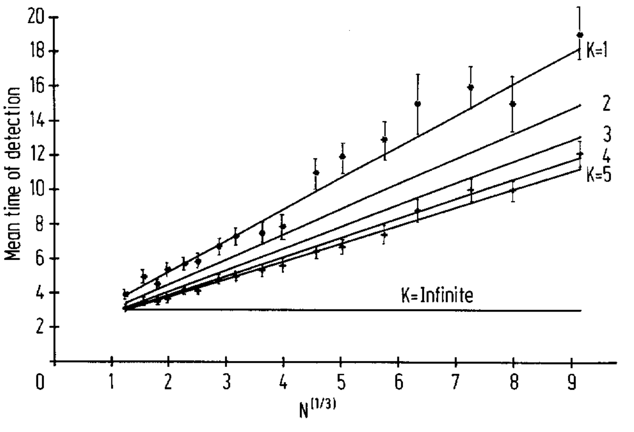

Fig. 1. Regression lines fitted to the means of detection times for the six K values of the mass-mating system. The means ± standard errors for $K=1$ (*) and $K=5$ (+) are also plotted

mean time of detection as a function of the entire number of parents in the other two mating systems. This failure was expected because of the variable degree of splitting into mating-groups of the populations with similar N values.

Time of detection is dependent on the decrease in heterozygosity. In the circular mating system, the inbreeding coefficient is mainly determined by the number of parents per group (Ng). The number of groups may have an influence only if the number is very small. Since there were no differences among the mean times of detection with simulations for the same value of Ng even though they had different numbers of groups, they were pooled by Ng values. Surprisingly, the fits for linear regressions of mean times on $Ng^{1/3}$ values were excellent. Conversely, the limited random-mating system showed differences among simulations with the same Ng value;

Table 2. The expected time of detection as a function of the number of parents (N) and the factor of examination (K) for the different systems of mating. Chi-square tests of deviation of the estimated b coefficients (Table 1) from those expected are also given

Mating system			
Mass-mating	$1.6+1.80\,(N/K)^{1/3}$	$\chi^2_{5df}=0.6$	$P=0.98$
Individual random mating and random examination	$1.2+1.44\,N^{1/3}+0.59\,(N/K^4)^{1/3}$	$\chi^2_{4df}=5.1$	$P=0.28$
Individual random mating and equal number of examined individuals	$0.7+1.54\,N^{1/3}+1.94\,(N/K^4)^{1/3}$	$\chi^2_{4df}=4.2$	$P=0.38$
Limited			
2 p/g	$-0.4+2.81\,N^{1/3}+0.67\,(N/K^4)^{1/3}$	$\chi^2_{1df}=0.1$	$P=0.82$
4 p/g	$1.0+1.83\,N^{1/3}+0.21\,(N/K^4)^{1/3}$	$\chi^2_{1df}=1.0$	$P=0.33$
8 p/g	$1.5+1.04\,N^{1/3}+0.89\,(N/K^4)^{1/3}$	$\chi^2_{1df}=1.1$	$P=0.29$
16 p/g	$1.1+0.70\,N^{1/3}+1.65\,(N/K^4)^{1/3}$	$\chi^2_{1df}=1.8$	$P=0.19$
Circular	$4.8+2.17\,(Ng/K)^{1/3}$	$\chi^2_{2df}=0.3$	$P=0.86$

Table 3. Coefficients of variation (CV) for some combinations of K and N values from the five systems of mating. Standard errors must be used with care because the detection times are non-normally distributed

Mating system	K	N			
		4	12	24	96
Mass-mating	1	0.78±0.06	0.71±0.07	0.68±0.07	0.75±0.10
	4	0.72±0.05	0.76±0.06	0.64±0.06	0.70±0.07
	Inf	0.64±0.05	0.63±0.05	0.79±0.05	0.69±0.05
Individual random and random examination	1	0.75±0.06	0.71±0.07	0.80±0.08	0.68±0.07
	4	0.68±0.05	0.62±0.06	0.67±0.07	0.69±0.06
	Inf	0.70±0.05	0.64±0.06	0.65±0.07	0.63±0.09
Individual random and equal examination	1	0.65±0.06	0.63±0.07	0.71±0.07	0.70±0.09
	4	0.72±0.05	0.68±0.05	0.75±0.07	0.78±0.08
	Inf	0.70±0.05	0.64±0.06	0.65±0.07	0.63±0.09
Limited, 2 parents per group	1	0.73±0.05	0.75±0.07	0.81±0.08	0.63±0.07
	4	0.70±0.05	0.62±0.06	0.64±0.07	0.60±0.08
	Inf	0.68±0.04	0.60±0.05	0.60±0.07	0.57±0.08
		Ng			
		2	4	8	32
Circular	1	0.63±0.04	0.66±0.04	0.58±0.05	0.53±0.06
	4	0.64±0.03	0.59±0.04	0.60±0.04	0.61±0.05
	Inf	0.72±0.06	0.67±0.06	0.68±0.06	0.91±0.07

$$CV=\frac{\text{Standard Deviation}}{\text{Mean}\ -1} \qquad \text{S.E.}=\frac{CV}{\sqrt{2n}}\sqrt{1+2\,CV^2}$$

Inf – infinite

so the regression fitting was made on $N^{1/3}$ values separately for the different mating-group sizes.

These results show that the cube root of the number of parents is the order of magnitude adequate for a wide spectrum of mating systems. Examination of more individuals than those that are used as parents does not alter the linear relation. This result allows the construction of a formula to estimate the mean time of detection that also includes the ratio of examined individuals K. In this way, the linear equations for each mating system were pooled as follows: the a coefficients were averaged (in the mass-mating system, the a coefficient corresponding to K= infinite was excluded) and the coefficients or regression b were adjusted to the function $cK^{-1/3}$ for the mass-mating and circular systems, and to $d+cK^{-4/3}$ for the other systems (other functions had been tested previous-

Table 4. Mean times of detection in generations (±S.E.) for combinations of heritability, population size and intensity of selection. The percentages of reduction of the mean times of detection in relation to $h^2 = 0$ are given in brackets

K	*N*	Heritability			
		0.0	0.2	0.4	0.6
2	10	5.6±0.2	5.3±0.2 (5.4)	5.1±0.2 (8.9)	5.1±0.2 (8.9)
	20	6.8±0.3	6.9±0.4 (−1.5)	6.8±0.4 (0)	6.4±0.4 (5.9)
	40	8.5±0.5	7.7±0.5 (9.4)	8.1±0.5 (4.7)	8.2±0.5 (3.5)
% Reduction averaged . .			(4.4)	(5.5)	(6.1)
4	10	4.9±0.2	4.6±0.2 (6.1)	4.3±0.2 (12.2)	4.3±0.2 (12.2)
	20	6.0±0.3	5.6±0.3 (6.7)	5.7±0.3 (5.0)	4.9±0.2 (18.3)
	40	7.4±0.5	7.0±0.4 (5.4)	6.6±0.4 (10.8)	6.6±0.4 (10.8)
% Reduction averaged . .			(6.1)	(9.4)	(13.8)
10	10	4.4±0.2	4.4±0.2 (0)	4.1±0.2 (6.8)	3.7±0.2 (15.9)
	20	5.4±0.3	5.4±0.3 (0)	4.9±0.3 (9.3)	4.8±0.3 (11.1)
	40	6.9±0.4	6.6±0.4 (4.3)	6.6±0.4 (4.3)	5.4±0.3 (21.7)
% Reduction averaged . .			(1.4)	(6.8)	(16.2)

ly). The expected mean times of detection are shown in Table 2 as a function of *N* and *K*. The deviations of *b* coefficients from that estimated by the function were always non-significant.

The distribution of times of detection in a monoecious random-mating population when only the *N* parents are examined follows a quasi-geometrical distribution, with a coefficient of variation near to ⅔ (Robertson 1978). The effect of excluding selfing is to increase the mean time of detection by one generation, as the allele cannot be detected in the first generation. To check the constancy of the variation of the distribution of detection times for other systems of mating and number of examined individuals, the coefficients of variation were calculated after subtracting one generation from the mean time of detection. Table 3 shows the coefficients of variation of a sample of several combinations of *N* (*Ng* in the circular mating system) and *K* from all the mating systems considered. Although some significant deviations from the ⅔ value were detected (e.g. the combination $Ng = 32$, $K =$ infinite of the circular mating system in Table 3), the value of ⅔ seems to be a good approximation to the coefficients of variation for all the systems of mating.

In artificial selection processes there is an actual reduction in effective population number. This reduction is a function of the intensity of selection and of the value of heritability (Robertson 1961). Selection for a character with a high heritability may cause an important reduction in the effective number (up to around 50% for $h^2 = 0.6$ and selected proportion $= 0.10$). Obviously, the effective number must affect the distribution of times of detection. Table 4 shows the mean times of detection for different combinations of heritability, intensity of selection and population size. The intensity of selection is generally related to the observation of a larger number of individuals. In this way, the effect of a reduction in the effective number on the detection time must be analyzed through the effect produced by an increase in heritability. As expected by the reduction in the effective population number, the reduction in mean times across heritability is greater with more intense selection. However, the reduction seems to be rather small. From low to moderately high heritability values, the reductions range up to a maximum of 10%. Very high values of heritability and intense selection are required to reduce the mean time of detection to about 20%.

Discussion

Differences in times of detection among the systems of mating may be considered to be caused by differences in the inbreeding increment and the particular split of the population into couples or groups for mating. However, these differences do not modify either the time scale $N^{1/3}$ or the coefficient of variation, ⅔, proposed by Robertson (1978). The time scale for the circular mating system is related to group size ($Ng^{1/3}$), because the rate of inbreeding increase for this system depends on the size of the group (Kimura and Crow 1963). In the remaining systems of mating, the increase in inbreeding is dependent on the effective number for the entire population.

The effective population number may be estimated from the variance of the family size (Wright 1940; Crow 1954). Random mass-mating and individual random-mating with random examination have the same variance of family size. The number of individuals chosen to be parents from each family follows a binomial distribution, with variance $2(N\text{-}2)/N$. Both systems of mating show similar regression lines of mean detection times on $N^{1/3}$ values when population size and number of individuals examined were the same ($K=1$). However, there is an increasing difference between the regressions for the two systems as K increases. Although in both systems the mean detection time decreases as K increases, only detection times in the random mass-mating system tend to be independent of the population size. Obviously, this is a consequence of the splitting of the population to mate. When groups or couples are previously determined to mate, the attainment of a given genotype is conditioned. Increasing the number of examined individuals reveals the differences with respect to a non-split population.

In the limited random-mating system, family size also follows a binomial distribution, but the variance is related to group size, $(Ng-2)/Ng$. Therefore, effective number is determined by group size as well as by number of groups. When number of parents per group is 2, the ratio between the effective population number (Ne) and actual size tends to 2. This increase in effective population number seems to be responsible for the high mean detection times. As number of reproducers per group increases, the ratio Ne/N decreases rapidly and differences with the mass-mating system disappear.

The effective population number of the individual mating system, when an equal number of individuals per family is examined, is determined by population size and by number of individuals examined. Because reproducers are sampled from among the individuals examined, the contribution of each family to the next generation follows a hypergeometric distribution. The variance of this distribution is $2(N-2)(K-1)/(KN-1)$. When only the N individuals that are to reproduce are examined ($K=1$), effective population number is the same as for the limited random-mating system with two parents per group (in fact, a common simulation was carried out for both cases). As the value of K increases with values 2, 3, 4, 5 and infinite, the ratio Ne/N rapidly decreases to values 1.33, 1.21, 1.14, 1.10 and 1.00, respectively. When number of individuals examined is infinite, the two individual mating systems considered are really the same.

Acknowledgements. I thank Prof. C. López-Fanjul for valuable discussions and comments. This work was supported by the Ministry of Education and Science of Spain within the project PA86-0007-CO2-02.

References

Crow JF (1954) Breeding structure of populations. II. Effective population number. In: Kemthorne O (ed) Statistics and mathematics in biology. Iowa State College Press, Ames, pp 543–556

Fisher RA (1958) Statistical methods for research workers. Oliver and Boyd, London

Frankham R (1980) Origin of genetic variation in selected lines. In: Robertson A (ed) Selection experiments in laboratory and domestics animals. Commonwealth Agricultural Bureaux, Slough, pp 56–68

Hollingdale B (1971) Analyses of some genes from abdominal bristle number selection lines in *D. melanogaster*. Theor Appl Genet 41:292–301

Karlin S, Tavaré S (1980) The detection of a recessive visible gene in finite populations. Genet Res Camb 37:33–46

Karlin S, Tavaré S (1981a) The detection of particular genotypes in finite populations. I. Natural selection effects. Theor Popul Biol 19:187–214

Karlin S, Tavaré S (1981b) The detection of particular genotypes in finite populations. II. The effects of partial penetrance and family structure. Theor Popul Biol 19:215–229

Kimura M, Crow JF (1963) On the maximum avoidance of inbreeding. Genet Res Camb 4:339–415

Robertson A (1961) Inbreeding in artificial selection programmes. Genet Res Camb 2:189–194

Robertson A (1964) The effect of non-random mating within inbred lines on rate of inbreeding. Genet Res Camb 5: 164–167

Robertson A (1978) The time of detection of recessive visible genes in small populations. Genet Res Camb 31:255–264

Wright S (1940) Breeding structure of populations in relation to speciation. Am Nat 74:232–248

Yoo BH (1980) Long-term selection for a quantitative character in large replicate populations of *D. melanogaster*. I. Genet Res 35:1–17