Genet. Res., Camb. (1979), **34**, *pp.* 11–17
Printed in Great Britain

The time of detection of sex-linked recessives in small populations

By J. W. JAMES*

Station de Génétique Quantitative et Appliquée, Centre National de Recherches Zootechniques, I.N.R.A., Jouy-en-Josas, France

(*Received* 7 *November* 1978)

SUMMARY

A sex-linked recessive gene with visible effect will first be detected in the hemizygous sex (male). In lines with equal numbers of males and females, when the gene is initially present in a single female the probability of detection falls from 2/3 in single pair lines to 0·54 in large lines. The mean and standard deviation of time to detection are almost independent of population size, being about 4/3 and 2/3 respectively. About 98% of all detections occur within three generations, so a gene detected much later than this after the foundation of a selection line is likely to be a new mutant. Higher initial frequencies and selection favouring heterozygotes increase the chance of detection. The time taken is decreased with higher initial frequencies and increased slightly by selection favouring heterozygotes.

1. INTRODUCTION

Genes with visible effects are often found in selection lines, and may sometimes be shown to be favoured by the selection practised. If a gene is first detected many generations after foundation of the line, it may be suspected that it has arisen through mutation or recombination during the selection programme. To assist in interpreting such events, Robertson (1978) has derived the distribution of time to detection of an autosomal recessive gene in a small population. The time scale was proportional to the cube root of population size rather than directly proportional to population size as for many processes. These results are also relevant to the age of a recessive mutant detected for the first time in a finite population. Although I know of no reported instance of a sex-linked recessive visible gene being detected in a selection line, such events may well occur. The purpose of this paper is to obtain the distribution of time to detection for a sex-linked recessive.

2. METHODS

For convenience it will be assumed that males are heterokaryotypic. Thus the first detection must always occur in a male, since a female can only be homozygous if she has inherited the gene from both parents, in which case it will have been

* Permanent address: School of Wool and Pastoral Sciences, The University of New South Wales, P.O. Box 1, Kensington, NSW, 2033, Australia.

0016-6723/79/2828-7490 $01.00

detected in her sire. It is this feature of sex-linked recessives which greatly simplifies the analysis. We shall not consider here the fact that in practice a recessive may not be recognized on its first occurrence. Since we are dealing only with events prior to first detection, all male parents carry the dominant allele, and we need consider only the distribution of the number of heterozygous (carrier) females, together with the probability of finding at least one recessive male.

Suppose that M males are examined each generation, and that N females are used for breeding. We need not concern ourselves with the number of breeding males since they are all dominants, nor with the number of females examined, since detection will not occur in them. Let j be the number of carrier females so that $x = j/2N$ is the frequency of the recessive allele in females. The next generation of females is formed by pairing N dominant alleles from males with N randomly chosen alleles from females. Thus the chance that there are k carriers in the next generation is

$$P_{kj} = \binom{N}{k} x^k (1-x)^{N-k}.$$

The chance that no male examined receives a recessive allele is $(1-x)^M$, so that the chance of detection in the next generation is $1-(1-x)^M$. If $q_{j,t}$ is the probability that the recessive has not been detected by generation t and the number of carrier females is j, while d_t is the probability of first detection in generation t,

$$d_{t+1} = \sum_{j=0}^{N} q_{j,t}(1-(1-j/2N)^M).$$

We also have

$$q_{k,t+1} = \sum_{j=0}^{N} q_{j,t} P_{kj}(1-j/2N)^M.$$

If Q_t is the chance that detection has occurred by generation t,

$$Q_{t+1} = Q_t + d_{t+1}.$$

Thus for any initial condition we can compute the distribution of detection times. For a single initial carrier, $q_{1,0} = 1$ with all other $q_{j,0}$ being zero. Other initial conditions are specified in the same way. The effect of selection can be studied by letting s denote the selective advantage of carrier females and then setting P_{kj} equal to

$$\binom{N}{k} (x+sx\ (1-2x))^k (1-x-sx(1-2x))^{N-k}$$

as shown by Robertson (1978).

In lines maintained with only one female per generation, explicit solutions can be found. The chance that the gene is not detected in the first generation, given an initial carrier female, is $0{\cdot}5^M$. The chance that the female chosen as a dam is a carrier is 0·5, and given this the chance of not detecting the gene in the next generation is $0{\cdot}5^M$. Thus the overall chance of detecting the gene for the first time in generation 2 is $0{\cdot}5^{M+1}(1-0{\cdot}5^M)$. Repetition of the argument shows that each

https://doi.org/10.1017/S0016672300019248 Published online by Cambridge University Press

succeeding generation has a chance of first detection $0{\cdot}5^{M+1}$ times that of the previous generation. The total probability of detection is found by summation to be

$$D = (1-0{\cdot}5^{M})/(1-0{\cdot}5^{M+1}).$$

The distribution of detection times, assuming the recessive is detected, is geometric, with ratio of successive terms $0{\cdot}5^{M+1}$, and from the properties of the geometric distribution, the mean is

$$T = 1/(1-0{\cdot}5^{M+1})$$

and the standard deviation is

$$\sigma = 0{\cdot}5^{(M+1)/2}/(1-0{\cdot}5^{M+1}).$$

For large populations we may approximate by assuming that the numbers of sons examined and daughters mated from each carrier female have Poisson distributions with means m and n respectively. If there are j carrier females in generation t, the expected total numbers of recessive males and carrier females in generation $t+1$ are $0{\cdot}5mj$ and $0{\cdot}5nj$ respectively. Then the chance that the gene is detected in the next generation is $1-\exp(-0{\cdot}5mj)$, while the chance that it is not detected and there are k carrier females is,

$$f_{kj} = \exp(-0{\cdot}5mj)\exp(-0{\cdot}5nj)(0{\cdot}5nj)^{k}/k!$$

Summing over different values of j

$$q_{k,t+1} = \sum_{j=0}^{\infty} q_{j,t} f_{kj}.$$

We define the function

$$F_t(z) = \sum_{j=0}^{\infty} q_{j,t} z^{j}$$

from which the probability of detection by time t is

$$Q_t = 1 - F_t(1).$$

$$\text{Also, } F_{t+1}(z) = \sum_{k=0}^{\infty} q_{k,t+1}\, z^{k},$$

and on inserting the expression for $q_{k,t+1}$, reversing the order of summation and simplifying

$$F_{t+1}(z) = \sum_{j=0}^{\infty} q_{j,t} \exp(-0{\cdot}5(m+n-nz)j).$$

Thus, writing $\exp(-0{\cdot}5)$ as A, we have

$$F_{t+1}(z) = F_t(A^{m+n-nz}).$$

In particular when one male and one female offspring are expected per dam and there is no selection, $m = n = 1$ and then

$$F_{t+1}(z) = F_t(A^{2-z}).$$

If there is initially a single recessive, $F_0(z) = z$. It is readily verified from the recurrence relation for $F_t(z)$ that

$$F_{t+1}(1) = A^{m+n-nF_t(1)}.$$

https://doi.org/10.1017/S0016672300019248 Published online by Cambridge University Press

Thus the distribution of detection times is readily computed,while the probability of failing to detect the recessive is the limiting value of $F_t(1)$, the solution of the equation

$$-2\ln F = m+n-nF.$$

The effect of having initially C carriers can be studied by setting $F_0(z) = z^C$, and then the values of $F_t(1)$ are found to be the Cth powers of those when $C = 1$. This is because we are essentially dealing with the combination of C independent branching processes each begun by a single carrier.

3. RESULTS

The transition matrix method was programmed, the correctness of programming being checked by noting that the correct results were obtained with $N = 1$, and that as N increased the results approached those for the branching process method.

To begin, the case of a single initial carrier, with equal numbers of males and females and no selection, will be considered. Table 1 gives some figures for different values of N. The probability of detection falls from 2/3 to 0·5361 as N goes from 1 to infinity, but the mean and standard deviation of detection time are virtually the same at all values of N. The probability of detection in the first generation is

Table 1. *Time to detection for a gene initially occurring only once, with no selection and equal numbers of males and females*

		Detection time		Fractions of detections by	
N	Proportion detected	Mean	S.D.	Gen. 1	Gen. 3
1	0·667	1·333	0·667	0·750	0·984
2	0·594	1·345	0·666	0·737	0·985
10	0·547	1·349	0·667	0·734	0·985
20	0·541	1·349	0·667	0·734	0·985
40	0·539	1·349	0·667	0·734	0·985
∞	0·536	1·350	0·667	0·734	0·985

about $\frac{3}{4}$ that of eventual detection for all population sizes, while only about 1·5% of all detections occur later than three generations. The ultimate probability of detection can be closely approximated by $0{\cdot}5361+0{\cdot}115/N$ for N from 2 to infinity, the approximation giving 0·6511 for $N = 1$. Thus a sex-linked recessive initially present in a single carrier under these conditions will be detected, if at all, almost certainly in the first three generations.

The effects of variation in the number of carriers initially present are shown in Table 2 for $N = 10$ and $N = \infty$, results for $N = 20$ or 40 being intermediate. Obviously, beginning with more carriers increases the probability of detection and reduces the time taken. Comparing cases with different population sizes, it is clearly the initial number of carriers rather than the initial gene frequency which is important, in contrast to the case of autosomal recessives where Robertson found that gene frequency was the determining factor. But, as can be seen in

Table 2, the distributions of detection times for a given initial number of carriers are very little affected by N over the range 10 to infinity, although this involves very marked gene frequency changes.

Table 2. *The effect of differing initial numbers of carriers, with no selection and equal numbers of males and females*

N	Initial carriers	Proportion detected	Detection time		Fraction of detections by	
			Mean	S.D.	Gen. 1	Gen. 3
10	1	0·547	1·349	0·667	0·734	0·985
	2	0·802	1·240	0·563	0·812	0·991
	4	0·967	1·093	0·354	0·923	0·997
∞	1	0·536	1·350	0·667	0·734	0·985
	2	0·785	1·249	0·573	0·806	0·990
	4	0·954	1·114	0·392	0·907	0·997

So far it has been assumed that the numbers of males and females are equal. There are two opposing factors which should be examined. In selection lines more animals are examined than are used for breeding. Here only the number of males observed is relevant, so we should examine the case where $M > N$. On the other hand, the number of sires is often much smaller than the number of dams, so we should examine the case $M < N$. In Table 3 the effect of varying M from 1 to

Table 3. *The effect of differing numbers of males examined with no selection, one initial carrier and* 20 *dams*

Males	Proportion detected	Detection time		Fraction of detections by	
		Mean	S.D.	Gen. 1	Gen. 3
1	0·048	1·927	1·325	0·516	0·889
2	0·093	1·861	1·249	0·533	0·902
4	0·171	1·753	1·127	0·565	0·922
10	0·349	1·539	0·887	0·642	0·960
20	0·541	1·349	0·667	0·734	0·985
40	0·752	1·175	0·442	0·847	0·998
100	0·951	1·033	0·182	0·968	1·000

100 when $N = 20$ and $C = 1$ is shown. The number of males examined has as expected a very marked effect on the probability of detection, which varies from 5 to 95 %. The average time to detection is rather greater when M is small, but the proportion of detections occurring in the first 3 generations is still as high as 89 % when $M = 1$. In effect, when M is small the chance of discovering the gene in a given generation is small, but if it is not detected early it will probably be lost.

The effect of selection acting on carrier females is illustrated for the case of a single initial carrier with $M = N = 20$ in Table 4. Positive selection is likely in artificial selection experiments, while if there is no artificial selection a selective disadvantage for females heterozygous for a newly arisen mutant may be more

likely. In any case, it can be seen from Table 4 that the effect of selection is primarily on the probability of detection, the distribution of detection times being only slightly altered, though if the gene is selected against it has a slightly lower mean detection time because it will be eliminated if not detected early.

Table 4. *The effect of selection with a single initial carrier and* 20 *males and* 20 *females*

Selection coefficient	Proportion detected	Detection time		Fraction of detections by	
		Mean	S.D.	Gen. 1.	Gen. 3
0·4	0·695	1·355	0·659	0·726	0·986
0·2	0·624	1·359	0·671	0·726	0·985
0·0	0·541	1·349	0·667	0·734	0·985
−0·2	0·446	1·322	0·640	0·752	0·987
−0·4	0·343	1·274	0·584	0·783	0·991

As will be clear from the results above, the rate of decline of the distribution is very similar for all values of N, changing from 0·25 when $N = 1$ to 0·232 when $N = \infty$. This is of limited relevance, however, since virtually all detection has occurred before the limiting rate of decline is established.

4. DISCUSSION

From the viewpoint of selection experiments the results of this study are clear cut. If a sex-linked recessive is found in a selection line for the first time more than three generations after the foundation of the line, then it is virtually certain to have arisen since the line was established. Similarly, a new sex-linked recessive mutant, if it is detected at all, is virtually sure to be detected within very few generations. Thus in no case can a sex-linked recessive which has just been discovered be plausibly regarded as having been present for many generations unless it is likely that previous occurrences in the hemizygous or homozygous condition could have passed unnoticed. The likelihood of this obviously depends on the magnitude of the visible effect of the gene and on the acuity of the observer.

The pattern of the results can be simply explained by considering the branching process model, which gives reasonably accurate results even for quite small population sizes. The relevant parameters for the development of this model are the number of female progeny expected to be mated per carrier female, the expected number of male progeny examined per carrier female, and the initial number of carrier females. Using appropriate values of these parameters in the model enables any case, including those involving selection, to be treated with reasonable accuracy, except that the probability of detection is somewhat underestimated for small population sizes. From this viewpoint, each carrier female in each generation either passes the recessive to at least one son so that it is discovered, or does not do so and passes it on to a variable number of daughters who then repeat this process. It is this independence which leads to the recurrence relation for $F_t(z)$ and also allows the initial presence of several carriers to be treated

so simply. Obviously the basic unit of the process is determined only by the expected numbers of sons and daughters per carrier, so that provided the population is not too small for the independence assumption to hold approximately, we may use this method. The reason why the method applies at such low population sizes is that if the number of carriers becomes even moderately large the gene is almost sure to occur in a male and the population is removed from further consideration. Thus only populations with small numbers of carriers are involved in the argument, so that N need not be large for the argument to be valid.

I wish to thank Mm. L. Ollivier and B. Vissac for providing facilities, the French Ministry of Foreign Affairs for a Fellowship and the University of New South Wales for a grant of study leave.

REFERENCE

ROBERTSON, A. (1978). The time of detection of recessive visible genes in small populations. *Genetical Research* **31**, 255–264.