THE EFFECTS OF LINKED LOCI ON ESTIMATES OF RELATIVE VIABILITY DERIVED FROM PEDIGREE ANALYSIS

D. J. COLGAN

Department of Population Biology, Research School of Biological Sciences, The Australian National University, PO Box 475, Canberra City, A.C.T. 2601, Australia

SUMMARY

The effect of viability selection at linked loci on segregation ratios at a specified studied locus is analysed in this paper for the crosses of a heterozygote to a homozygote and of two heterozygotes. Both autosomal and sex-linked loci are treated. The analysis indicates that it would be unusual for the change in frequency of a given allele at the studied locus to be distorted by as much as one per cent by the combined effects (assumed to be additive) of selection at all linked loci. This result encourages the use of pedigree analysis in tests of the null hypothesis of selective neutrality of the genotypes at a particular locus since the inevitable cases where selection at linked loci does lead to incorrect rejection of the null hypothesis should be relatively few.

1. INTRODUCTION

Pedigree analysis has often been used in investigation of the relative viability of different genotypes at a locus (Haldane, 1962; Sing *et al.*, 1971). Deviations from expected segregation ratios are taken as indicating the possibility of differential fitnesses. But estimates of the relative viability of the genotypes at a given locus are affected by selection at linked loci (Sved, 1968; Thompson, 1977; Asmussen and Clegg, 1981). Before selection is postulated as the cause of a significant deviation from expected segregation ratios it is necessary to account for the effect of linked loci on the locus actually being studied.

Colgan (1981) approached the complication of linkage by comparing, for a variety of human traits, the numbers of heterozygous and homozygous progeny of the cross of a heterozygote and a homozygote. In that study, it was assumed that selection at linked loci would favour one class of genotype at the studied locus (*i.e.*, heterozygotes) as often as it would favour the other (the homozygotes). There was, however, a significant excess over one half in the fraction of the numbers of traits where more heterozygous than homozygous offspring were found. This result was interpreted as suggesting the possibility of a general heterozygous advantage. But the method is not usually directly applicable to the study of single loci. The purpose of this paper is to investigate the theory of the effects of linkage on estimates of relative viability derived from pedigree analysis of specified loci.

The combined effects of a number of linked loci are so complex that it is not possible to determine exactly what distortion in segregation ratios is actually due to selection at the studied locus. So the determination of the relative fitnesses of genotypes at the studied locus can not be precise. The

theory of linkage effects here developed is intended only to enable discrimination between cases where there is or is not some significant level of selection at a studied locus. The maximal change in the genotypic frequencies at a studied locus which can be caused by selection at linked loci is estimated and subtracted from an observed change to leave a residual which is compared with the alterations in frequency allowed by the hypothesis of selective neutrality at the studied locus.

2. Crosses between a heterozygote and a homozygote

(i) *Development of the model*

The locus under consideration is designated β. It is assumed to be autosomal. Two alleles at the locus, denoted B and b are used in the analysis. It is not initially known whether selection discriminates between the genotypes at the locus. The frequencies of the genotypes in the general population from which the parental generation is taken are f_1 bb, f_2 Bb and $f_3 BB$. Segregation ratios in pedigrees are affected by viability selection at a closely linked locus, α, at which there are assumed to be two alleles, A and a. The genetic symbols, A, a, B and b have been chosen to distinguish the alleles and do not imply dominance. The relative fitnesses of the genotypes at locus α are defined as $1 - s$ for AA, 1 for Aa and $1 - t$ for aa. The values of s and t are unknown, in general, because locus α is not scored in the study. The amount of recombination between the α and β loci in the parental generation is symbolised as r.

The gametic frequencies are treated as being conditioned on the β-locus gametotype. Of the b-carrying chromosomes in the parental generation p_1 carry the A allele and p_2 carry the a allele. Similarly, q_1 of the B-carrying chromosomes carry A and q_2 carry a. So the frequencies of the chromosomal types, conditioned on the β locus gametotype, are p_1 Ab, p_2 ab, q_1 AB and $q_2 aB$ where $p_1 + p_2 = q_1 + q_2 = 1$. The chromosomal types are, initially, assumed to be combined at random in the actual genotypes of the parents. It is assumed that mating is random with respect to the α locus, that there are no differences between the sexes in gametic frequencies or relative genotypic fitnesses and that there are no differences between the fertility of the parental genotypes.

The frequencies of gametes produced by the bb homozygotes will be p_1 for Ab and p_2 for ab, whatever the degree of recombination since crossing over in such individuals does not lead to recombinant gametic classes. The situation is more complex for heterozygous Bb parents. The four possible strand constitutions of such individuals are shown in table 1. The table indicates the total frequencies of the gametic classes produced by all of the types of heterozygotes, assuming equal fertility. The parameter, D^c, $(= p_1 q_2 - p_2 q_1 = p_1 - q_1)$ which is adopted in the table as the "conditional linkage disequilibrium" may be considered as the analogue of the linkage disequilibrium parameter, D, defined for the gametic frequencies in the population as a whole (Crow and Kimura, 1970).

The relative viabilities of the α locus genotypes can be combined with the zygotic frequencies of the genotypes which are expected under random mating to obtain the adult frequencies expected in the filial generation

TABLE 1

Calculation of the frequencies of the various gametic classes produced by Bb heterozygotes

Heterozygote's genotype	Frequency of heterozygote†	Production of gametes			
		AB	aB	Ab	ab
AB/Ab	p_1q_1	$p_1q_1/2$		$p_1q_1/2$	
aB/Ab	p_1q_2	$p_1q_2r/2$	$p_1q_2(1-r)/2$	$p_1q_2(1-r)/2$	$p_1q_2r/2$
AB/ab	p_2q_1	$p_2q_1(1-r)/2$	$p_2q_1r/2$	$p_2q_1(1-r)/2$	$p_2q_1(1-r)/2$
aB/ab	p_2q_2		$p_2q_2/2$		$p_2q_2/2$
Frequency of gametes*		$(q_1+rD^c)/2$	$(q_2-rD^c)/2$	$(p_1-rD^c)/2$	$(p_2+rD^c)/2$

* D^c, the conditional linkage disequilibrium, equals $(p_1q_2-p_2q_1)=(p_1-q_1)$.
† Assuming the random combination of chromosomal types.

under the null hypothesis that selection is not acting at the β locus, *e.g.*,

$$f(AABb) = (1-s)(p_1q_1 + rp_1D^c)/2\bar{W}$$

where, by summation, $\bar{W} = 1 - 0.5(sp_1q_1 + tp_2q_2 + sp_1^2 + tp_2^2)$.

The change in the frequency of the B allele can be computed by subtracting its frequency in the parental generation (0·25) from half of the frequency of the Bb genotype in the filial generation. That is, the change in the frequency of B due to selection at locus α is

$$\Delta fr(B) = [(1 - sp_1q_1 - tp_2q_2 - r(sp_1 - tp_2)D^c)/4\bar{W}] - 0.25 \qquad (1)$$

$$= [(s+t)p_1 - t](1-r)/8\bar{W}. \qquad (2)$$

There is an inverse relation between the potential effect—be it positive or negative—of locus α and the strength of the linkage of this locus to the studied locus β. When there is no recombination the change in the frequency of B due to selection at the α locus is

$$\Delta fr(B) = ((s+t)p_1 - t)D^c/8\bar{W}. \qquad (3)$$

(ii) *The effect of associative overdominance*

It is not possible to predict whether the effects of selection at locus α will tend to increase or decrease the frequency of allele B unless further information is available as to the values of the parameters s, t, p_1 and q_1. And nothing can be said about general tendencies at a number of loci unless there are general constraints on the values. The simplest example of a set of such constraints is the restriction that s and t are both greater than zero and the assumption that locus α is at an equilibrium in the population as a whole so that the overall frequency of allele A, written as $f(\hat{A})$, is equal to $t/(s+t)$. From equation 2, with $r=0$ for maximal effect, it can be seen that the change in the frequency of B due to selection at locus α is

$$\Delta fr(B) = (s+t)(p_1 - q_1)(p_1 - f(\hat{A}))/8\bar{W}. \qquad (4)$$

It will be assumed, for the moment, that the number of pedigrees examined is large enough to allow the frequency of A in the population at equilibrium to be written as

$$f(\hat{A}) = (f_1 + f_2/2)p_1 + (f_2/2 + f_3)q_1.$$

Then

$$p_1 - f(\hat{A}) = (p_1 - q_1)(f_2/2 + f_3). \tag{5}$$

So

$$\Delta fr(B) = (s + t)(p_1 - q_1)^2 (f_2/2 + f_3)/8\,\bar{W}. \tag{6}$$

Since $(s + t)$ must be positive, the effect of a linked locus α in heterotic equilibrium must be to increase the frequency of allele B.

This type of effect has been called "associative overdominance" (Sved, 1968). Estimates of the values derived from equation (6) for $\Delta fr(B)$ may be useful in deriving limitations on the amount of segregaton ratio distortion caused by linked loci in heterotic equilibrium, even though the effect of a single locus cannot be exactly specified. Now $(f_2/2 + f_3)/8\,\bar{W}$ will be about 0·05 (or less if allele B is rare). And it seems unlikely that the value of $(s + t)$ would average more than 0·05 although there is no information available on this point. Finally, in the absence of restrictions on recombination the average value of $(p_1 - q_1)$ would probably be low, say less than 0·2. These broad estimates suggest that the change in the frequency of B due to selection at a single linked locus is unlikely to average as much as 0·0001. There could be a large number of linked loci in heterotic equilibrium exerting an effect on locus β. Changes in the frequency of the B allele due to these effects might be assumed additive. Under this assumption, one hundred loci causing an average change of 0·0001 would be required to alter the frequency of B by 1 per cent.

The situation where allele B is rare in the population deserves further comment. If the frequency of the allele is less than 0·01 and if the term $(s + t)/8\,\bar{W}$ is about 0·01, then, irrespective of the value of $(p_1 - q_1)$, the maximal change due to selection at a linked locus in heterotic equilibrium would be of the order of 0·0001. Even if a number of loci exert an additive effect on the β locus by associative overdominance it seems unlikely that the frequency of the (rare) B allele can be altered by as much as 1 per cent by this type of interaction.

(iii) *The alteration of linkage effects by sampling*

The values of p_1 and q_1 found in the parents in the pedigrees available for analysis will differ from those obtaining for the population as a whole because of sampling effects. The populational values of these parameters are written as $\bar{p}_1$ (or $\bar{p}_2$). The observed frequency of Ab gametes in the sample of bb homozygous parents is defined as $\bar{p}_1 + \delta_1$, the observed frequency of Ab gametes in the sample of Bb heterozygous parents as $\bar{p}_1 + \delta_2$ and the observed frequency of AB gametes in the sample of Bb heterozygous parents as $\bar{q}_1 + \gamma$.

By following the procedure of the deterministic case, with r assumed to be zero, it can be shown that the change in the frequency of B due to selection at the linked locus, when allowance is made for sampling error in the parental genotypes, is

$$\Delta fr(B) = (\bar{p}_1 + \delta_2 - \bar{q}_1 - \gamma)[(\bar{p}_1 + \delta_1)(s + t) - t]/8\,\bar{W} \tag{7}$$

where

$$\bar{W} = 1 - 0{\cdot}5[s(\bar{p}_1 + \delta_1)(\bar{p}_1 + \delta_2 + \bar{q}_1 + \gamma) + t(\bar{p}_2 - \delta_1)(\bar{q}_2 - \gamma + \bar{p}_2 - \delta_2)].$$

The distribution of δ_1 may be approximated as a normal variate with a mean of 0 and a variance of $\bar{p}_1(1-\bar{p}_1)/2n_1$ where n_1 is the number of bb individuals examined in the parental generation. If n_2 is the number of Bb individuals examined then the distribution of δ_2 can be approximated as $N(0, \bar{p}_1(1-\bar{p}_1)/n_2)$, and that of γ as $N(0, \bar{q}_1(1-q_1)/n_2)$. Now $\bar{p}_1(1-\bar{p}_1)<\frac{1}{4}$ so about 200 homozygotes would need to be examined to be at least 95 per cent sure that $\delta_1<0.05$. If 200 heterozygotes were used and if δ_2 and γ are independent then there would be at least 95 per cent confidence that $\delta_2-\gamma<0.1$. Assuming the upper limits on the averages of parameters to be 0.05 for $(s+t)$, 0.10 for $\bar{W}$ and 0.2 for $(\bar{p}_1-\bar{q}_1)$, equation (7) shows that the likely maximal change in the frequency of B caused by selection at a single linked locus when sampling is accommodated would be less than 0.0015.

By inspection of equations (4), (5) and (7) it can be seen that the change in the frequency of B caused by selection at a single linked locus in heterotic equilibrium when sampling effects are considered is given by the equation:

$$\Delta fr(B) = (s+t)(\bar{p}_1 + \delta_2 - \bar{q}_1 - \gamma)[(\bar{p}_1 - \bar{q}_1)(f_2/2 + f_3) + \delta_1]/8\,\bar{W}. \qquad (8)$$

If $(f_2/2 + f_3)$ is about $\frac{1}{2}$ and the other parameters take the values assumed in the previous paragraph then it is unlikely that the value of equation 8 would be above 0.00025. The majority of samples would give values nearer to the deterministic estimate of 0.0001.

3. CROSSES BETWEEN TWO HETEROZYGOTES

(i) *Deterministic analyses*

The second situation in which segregation ratios may be used to obtain an estimate of fitness is the cross between two heterozygous parents. The frequency of allele B in the parental generation of such crosses is 0.5. The relative frequencies of the gametes produced by each sex in this cross will be the same as those in table 1. The various α locus genotypes of the parents are incorporated in table 2 which allows calculation of the genotypic frequencies in the zygotes. Again it is assumed that there are no differences in the genotypic frequencies in the two sexes, that there are no fertility differences between genotypes and that mating is random with respect to the α locus.

The frequency of the B allele in the offspring after selection is, as usual, the sum of the frequency of BB homozygotes and half of the frequency of Bb heterozygotes. This can be shown to equal $[2 - s(q_1 + rD^c)(q_1 + p_1) - t(q_2 - rD^c)(p_2 + q_2)]/4\,\bar{W}$, where $\bar{W}$ is equal to $1 - s(p_1 + q_1)^2/4 - t(p_2 + q_2)^2/4$.

So the change in the frequency of the B allele between the parental and offspring generations is given by the equation:

$$\Delta fr(B) = (1 - 2r)D^c[(p_1 + q_1)(s+t) - 2t]/8\,\bar{W}. \qquad (9)$$

In general, the conclusions of sections 2(i) and 2(ii) regarding the effect of linked loci on changes in the frequency of B remain valid. The conclusion that selection at linked loci in heterotic equilibrium will not have a major effect on the frequency of B is reinforced by the consideration that $(p_1 - f(\hat{A}))$ and $(q_1 - f(\hat{A}))$ will be opposite in sign when locus α is at equilibrium. In

TABLE 2

Genotypic frequencies in the zygotes produced by the Bb × Bb cross

Female gametes	Frequency	Male gametes			
		AB $(q_1+rD^c)/2$	aB $(q_2-rD^c)/2$	Ab $(p_1-rD^c)/2$	ab $(p_2+rD^c)/2$
AB	$(q_1+rD^c)/2$	$(q_1+rD^c)^2/4$	(q_2-rD^c) $\times$ $(q_1+rD^c)/4$	(p_1-rD^c) $\times$ $(q_1+rD^c)/4$	(p_2+rD^c) $\times$ $(q_1+rD^c)/4$
aB	$(q_2-rD^c)/2$	(q_1+rD^c) $\times$ $(q_2-rD^c)/4$	$(q_2-rD^c)^2/4$	(p_1-rD^c) $\times$ $(q_2-rD^c)/4$	(p_2+rD^c) $\times$ $(q_2-rD^c)/4$
Ab	$(p_1-rD^c)/2$	(q_1+rD^c) $\times$ $(p_1-rD^c)/4$	(q_2-rD^c) $\times$ $(p_1-rD^c)/4$	$(p_1-rD^c)^2/4$	(p_2+rD^c) $\times$ $(p_1-rD^c)/4$
ab	$(p_2+rD^c)/2$	(q_1+rD^c) $\times$ $(p_2+rD^c)/4$	(q_2-rD^c) $\times$ $(p_2+rD^c)/4$	$(p_1-rD^c)/4$ $\times$ $(p_2+rD^c)/4$	$(p_2+rD^c)^2/4$

such a case the change in the frequency of B is

$$(1-2r)D^c(s+t)[(p_1-f(\hat{A})) + (q_1-f(\hat{A}))]/8\,\bar{W}. \tag{10}$$

The value of $(q_1-f(\hat{A}))$ is $(q_1-p_1)(f_1+f_2/2)$ and that of $(p_1-f(\hat{A}))$ is $(p_1-q_1)(f_2/2+f_3)$. So

$$\Delta fr(B) = (1-2r)(s+t)(p_1-q_1)^2(f_3-f_1)/8\,\bar{W}. \tag{11}$$

The factor (f_3-f_1) in equation (11) suggests that loci to be studied should be chosen on the basis of having approximately equal frequencies of the homozygotes. The analysis of the cross of heterozygotes is complicated in cases where allele B is rare by the probability that BB offspring are inbred at a large number of linked loci. Such cases are better treated in homozygote x heterozygote crosses.

(ii) *Sampling theory*

To obtain estimates of the maximal effects of sampling r is assumed to be 0. The values of p_1 and q_1 in the overall population at the time of sampling of the parental generation are again written as $\bar{p}_1$ and $\bar{q}_1$. The observed frequency of Ab gametes is written as $\bar{p}_1 + \varepsilon_1$ for the male parents and as $\bar{p}_1 + \varepsilon_2$ for the female parents. The observed frequency of AB gametes is written as $\bar{q}_1 + \eta_1$ for males and as $\bar{q}_1 + \eta_2$ for females. The change in the frequency of B caused by selection on locus α which is superimposed on the deterministic change by sampling effects is equal to

$$[s(p_1(\varepsilon_1+\varepsilon_2) - q_1(\eta_1+\eta_2) + \varepsilon_1\varepsilon_2 - \eta_1\eta_2)$$
$$- t(p_2(\varepsilon_1+\varepsilon_2) - q_2(\eta_1+\eta_2) - \varepsilon_1\varepsilon_2 + \eta_1\eta_2)]/8\,\bar{W}. \tag{12}$$

The sampling deviations in equation (12) are expected to follow the binomial distribution. For approximations, $\bar{W}$ may be treated as 1 and the terms in $\varepsilon_1\varepsilon_2$ and $\eta_1\eta_2$ ignored. Under the normal approximation to

the binomial, the examination of 400 parental males and 400 parental females would allow at least 95 per cent confidence that the terms $(p_1(\varepsilon_1 + \varepsilon_2) - q_1(\eta_1 + \eta_2))$ and $(p_2(\varepsilon_1 + \varepsilon_2) - q_2(\eta_1 + \eta_2))$ are in the range $(-0\cdot1, 0\cdot1)$. If, again, the value of $(s + t)$ is about $0\cdot05$ then the maximal change in the frequency of B according to equation (12) is unlikely to be above $0\cdot05 \times 0\cdot1/8$ ($0\cdot0006$). There would be instances where extreme values of s, t or the sampling deviations cause such large changes in the frequency of B that the null hypothesis of selective neutrality is incorrectly rejected but in the majority of cases the superimposed effect of sampling error would be small.

4. Sex linkage

Separate analyses are required for males and females if a locus is sex-linked. Genotypic frequencies in female offspring may be computed in the same way as frequencies in the offspring of the cross of a heterozygote to a homozygote. The (hemizygous) males are grouped according to the β locus allele which they carry. Only two types of X-carrying gamete are produced by each group of males. Heterozygous females produce the same types of gametes in the same frequencies as heterozygotes in the case of autosomal loci (see table 1). The frequencies of the various genotypes in females are the same as those derived from all offspring in the autosomal situation. The analysis for the sex-linked locus proceeds to equation (3) and, if sampling effects are taken into account, to equation (7). However, the equilibrium frequency under selection at a sex-linked locus is not necessarily the same as the equilibrium frequency at an autosomal locus (Crow and Kimura, 1970). So it is not possible to proceed from equation (3) to equation (4).

The expected genotypic frequencies in male zygotes are equal to the frequencies of the gametes produced by the female parents. These are given in table 1. If the relative fitnesses of the A and a genotypes are written as $1 - u$ and 1 then it can be shown that the change in the frequency of B caused by selection at the linked locus α is equal to:

$$u(p_1 - q_1)(1 - 2r)/4\bar{W} \tag{13}$$

where $\bar{W} = 1 - u(p_1 + q_1)/2$.

5. Discussion and conclusions

The analysis in this paper has been directed at evaluation of the selective significance of changes in the frequency of an allele at a specified locus. The maximal likely effect on the frequency of the allele caused by selection at linked loci is estimated for the type of mating under consideration. As a rule of thumb, the value of $0\cdot01$ might be chosen for the maximal effect since order-of-magnitude calculations suggest that selection at linked loci will not often cause such levels of distortion in allelic frequencies if samples of several hundred parents of each sex are scored. The maximal likely effect is subtracted from any actual change which is found. The residual is tested for significance by use of the binomial distribution or the normal approximation to the binomial. This approach can only detect selective differences which are large enough to be noted after compensation for estimated linkage effects. Large samples of the offspring are necessary to be confident of

detecting such differences. Loci with hgh selective differentials would be in the minority but the approach does allow estimation of the actual proportion of such loci.

There will inevitably be cases where selection at linked loci has such an effect on segregation ratios at a studied locus that the null hypothesis of selective neutrality is incorrectly dismissed. Such cases might be due to the conjunction of very large values of the selective coefficients s and t at the linked loci with high values of the conditional linkage disequilibrium parameter, D^c. The frequency of these cases can be minimised by appropriate choice of the loci for study. Rare alleles are suitable for the cross of a heterozygote with a homozygote (equation (6)). For crosses between two heterozygotes, loci that have approximately equal frequencies of homozygotes in the population at large should be chosen (equation (11)).

A locus may be affected by selection at a large number of linked loci in such a manner that the changes in the frequency of the studied allele due to the separate effects of the linked loci are almost all in the same direction. This might be the case if the linked loci are in heterotic equilibrium since, then, the frequency of the allele designated as B is always increased. The association of the studied locus with a large number of selected loci may arise by chance or it may be due to linkage to an inversion or with a block of genes that is protected from recombination by some such means as the localisation of chiasmata. Further study of a locus should be neglected if preliminary investigation reveals a high degree of linkage disequilibrium between loci in the genetic background.

Other factors that must be considered when dealing with specific loci are non-random mating at the α locus, complications caused by multiple alleles at the α locus and the possible distortions in β locus gametic frequencies caused by such processes as gene conversion (Lamb and Helmi, 1982), meiotic drive (Sandler and Novitski, 1957) or gametic selection. But with adequate experimental design, incorrect rejections of the null hypothesis of selective neutrality would not be much increased over the usual rate of 5 per cent if a maximal value of $0 \cdot 01$ is assumed for the effect of linked loci on the frequency of B. So pedigree analysis would seem to provide a good opportunity for dealing with the problems caused by linkage in the demonstration of selective differentials.

6. REFERENCES

ASMUSSEN, M. A. AND CLEGG, M. T. 1981. Dynamics of the linkage disequilibrium function under models of gene frequency hitch-hiking. *Genetics, 99*, 337–356.

COLGAN, D. J. 1981. The relative viabilities of alternative human genotypes. *Human Heredity, 31*, 172–179.

CROW, J. F. AND KIMURA, M. 1970. *An Introduction to Population Genetics Theory.* Harper and Row, New York.

HALDANE, J. B. S. 1962. Evidence for heterosis in woodlice. *J. Genet., 58*, 39–45.

LAMB, B. C. AND HELMI, S. 1982. The extent to which gene conversion can change allele frequencies in populations. *Genet. Res., 39*, 199–217.

SANDLER, L. AND NOVITSKI, E. 1957. Meiotic drive as an evolutionary force. *Amer. Nat., 91*, 105–110.

SING, C. F., SHREFFLER, D. C. NEEL, J. V. AND NAPIER, J. A. 1971. Studies on genetic selection in a completely ascertained Caucasian population. II. Family analyses of eleven blood group systems. *Am. J. Hum. Genet.*, *23*, 164–198.

SVED, J. A. 1968. The stability of linked systems of loci with a small population size. *Genetics*, *59*, 543–563.