Genet. Res., Camb. (1984), **44**, *pp.* 321–341
With 1 text-figure
Printed in Great Britain

The population genetics of haplo-diploids and X-linked genes

By P. J. AVERY

Department of Applied Statistics, University of Reading, Reading, Berks

(*Received 7 December 1983 and in revised form 8 August 1984*)

SUMMARY

From the available electrophoretic data, it is clear that haplodiploid insects have a much lower level of genetic variability than diploid insects, a difference that is only partially explained by the social structure of some haplodiploid species. The data comparing X-linked genes and autosomal genes in the same species is much more sparse and little can be inferred from it. This data is compared with theoretical analyses of X-linked genes and genes in haplodiploids. (The theoretical population genetics of X-linked genes and genes in haplodiploids are identical.) X-linked genes under directional selection will be lost or fixed more quickly than autosomal genes as selection acts more directly on X-linked genes and the effective population size is smaller. However, deleterious disease genes, maintained by mutation pressure, will give higher disease incidences at X-linked loci and hence rare mutants are easier to detect at X-linked loci. Considering the forces which can maintain balanced polymorphisms, there are much stronger restrictions on the fitness parameters at X-linked loci than at autosomal loci if genetic variability is to be maintained, and thus fewer polymorphic loci are to be expected on the X-chromosome and in haplodiploids. However, the mutation-random drift hypothesis also leads to the expectation of lower heterozygosity due to the decrease in effective population size. Thus the theoretical results fit in with the data but it is still subject to argument whether selection or mutation-random drift are maintaining most of the genetic variability at X-linked genes and genes in haplodiploids.

1. INTRODUCTION

The study of X-linked genes and genes in haplodiploid insects has been a neglected area until relatively recently. However, much more data is now being collected, particularly on haplodiploids, on levels of genetic variability as measured by electrophoresis (Berkelhammer (1983); Lester & Selander (1979); Cooper *et al.* (1979)). Haplodiploid insects are of particular interest due to the occurrence of a complex social system in certain species and the sex-ratio distortion that can also occur in them. (Virtually all haplodiploid insects are hymenoptera, e.g. bees, wasps and ants.) Further, in dioecious diploid species with a chromosomal sex-determining mechanism, the X-chromosome can form a substantial part of the genome (e.g. in *Drosophila robusta*, 38% of the euchromatin is on the X-chromosome, Carson (1955)) and thus many genes may be on the X-chromosome. Thus it is of

importance to look at X-linked genes and genes in haplodiploids and to characterise the similarities and differences from autosomal genes in diploids.

In this paper, I shall first discuss the experimental results and then look at various theoretical results to see how the two fit together. However, it is first necessary to define the system under consideration. The crucial property about X-linked genes and genes in haplodiploids is that one sex, usually the males, has only one copy of the gene (received from the female parent if males) whereas the other sex has two copies, one from each parent. In haplodiploids, this is achieved because the males develop from unfertilized eggs. Most species with an X-Y sex determining system have XY males, but certain species (e.g. birds and moths) have XY females whose single copy of the X-chromosome comes from their paternal parent. For simplicity, I shall assume the males to be the heterogametic sex and thus for a gene with two alleles, A_1 and A_2, we shall have the following genotypes:

Females			Males	
A_1A_1	A_1A_2	A_2A_2	A_1	A_2

X-linked genes and genes in haplodiploids have the same method of transmission. The Y-chromosome generally does not have genes in common with the X-chromosome and there is generally no crossing over between the X and Y. Thus similar results should apply to both X-linked genes and genes in haplodiploids.

2. EXPERIMENTAL RESULTS

Most of the data on genetic variability at X-linked genes and in haplo-diploids have been obtained from looking at the electrophoretic analysis of protein variability. The simplest summary measure of genetic variability to use is H, the average proportion of heterozygous loci per female within a population. In most of the data between 15 and 20 loci have been studied for each individual. Some authors have also looked at the proportion of polymorphic loci in a population. However, this is more affected by small sample sizes or small population sizes than H and different authors have used different definitions as to what constitutes a polymorphic locus, (e.g. frequency of common allele $= P < 0{\cdot}95$ or $< 0{\cdot}99$). Thus this measure is a more difficult one to use when comparing different authors' work and I shall concentrate on H.

Some average values for H from haplodiploid insects and other species is given in Table 1 along with between species variation in H. The figures for diploid organisms are calculated from various loci, a proportion of which will be X-linked. It is clear from the table that haplodiploid insects have much lower mean heterozygosities than other insects. However, the difference may not necessarily be completely due to the haplodiploidy but to other differences between haplodiploid hymenoptera and diploid insects, as the mean heterozygosity for haplodiploids is only marginally less than the mean heterozygosity in mammals. Berkelhammer (1983) investigated whether the social nature of hymenoptera was a cause of the difference. 90% of hymenopteran species are actually solitary. However, the social species have been investigated much more so that 70% of the available data is from these species. Berkelhammer (1983) found that primitive eusocial species (i.e. those

Table 1. *Summary of heterozygosity values* (H)

Category	No. of species	Mean heterozygosity (S.E.)	S.D. of H*
Haplodiploids (all hymenoptera)†	51	0·037 (0·004)	0·027
Solitary†	18	0·039 (0·006)	0·026
Primitively eusocial†	9	0·011 (0·006)	0·018
Advanced eusocial†	24	0·045 (0·005)	0·025
Drosophila‡	40	0·137 (0·008)	0·053
Other diploid insects‡	17	0·112 (0·022)	0·089
Non-insect invertebrates‖	13	0·102 (0·021)	0·076
Vertebrates‖	71	0·050 (0·004)	0·035

* Standard deviation of H between species.
† Berkelhammer (1983).
‡ Ward (1980).
‖ Powell (1975).

social species with little morphological difference between queens and workers and relatively little behavioural domination of queens over workers) had significantly smaller mean heterozygosity than the solitary species or advanced eusocial species (e.g. ants). He argued that the primitive eusocial species are likely to have evolved a social system relatively recently and that a high level of inbreeding considerably helps the evolution of eusociality. However, the solitary species still have much lower mean heterozygosity than other diploid insects, suggesting that the haplodiploid system is causing a difference separately from the social structure. Table 1 also shows that the between species variation in heterozygosity seems to increase with the mean value and hence is lower in haplodiploids.

The available data on *X*-linked genes are much more limited. A summary was given by Cooper *et al.* (1979). In man, they found little evidence of any difference in genetic variability between *X*-linked genes and autosomal genes but the amount of data was small and the loci studied were not necessarily randomly chosen over the genome. In McKusick's (1975) list of 1124 human gene loci, 93 (8·1 %) are on the *X*-chromosome, whereas the *X*-chromosome forms only about 5 % of the genome. Cooper *et al.* (1979) suggested that this excess of loci on the *X*-chromosome was because an *X*-linked gene is easier to detect than an autosomal recessive. Cooper *et al.* (1979) also discussed *Drosophila* data from Prakash, Lewontin & Hubby (1969) and Prakash (1973, 1977*a*, *b*). They found proportionally less polymorphic loci on the *X*-chromosome (23 %) than you would expect from the relative size of the *X*-chromosome to the whole genome (38 %) but again the selection of loci was probably far from random. The mean proportion of heterozygotes at polymorphic loci for 3 Drosophila species were also given by Cooper *et al.* (1979) (see Table 2). These showed little evidence of a difference between *X*-linked and autosomal loci.

Finally, Cooper *et al.* (1979) gave some data from ten species of Kangaroos and wallabies. Kangaroos and wallabies are of special interest due to their method of dosage compensation. Various methods of dosage compensation exist to try to equalize the effect of hemizygotes in the males with the equivalent homozygotes in the females. In most mammals, this is done by random *X*-inactivation in cells

Table 2. *Mean proportion of heterozygotes at polymorphic* X*-linked and autosomal loci in three species of Drosophila (from Cooper* et al. *(1979) using data from Prakash* et al. *(1969) and Prakash (1973, 1977* a *and* b))

Species	*X*-linked	Autosomal
D. pseudoobscura	0·216 (0·106)*	0·184 (0·033)
D. persimilis	0·350 (0·094)	0·220 (0·061)
D. robusta	0·157 (0·066)	0·211 (0·044)

* Standard errors given in parentheses.

in females whereas in female kangaroos and wallabies the paternal *X*-chromosome is invariably inactivated (this is discussed in detail in Cooper *et al.* (1977)). Cooper *et al.* (1979) took 11 loci known to be autosomal in humans and 4 loci known to be *X*-linked in humans. They then used Ohno's law of conservation of the *X*-chromosome (Ohno, 1967) to assume that these loci would be similarly sited in these ten species of kangaroos and wallabies. All loci which showed any variability, and whose position (i.e. on the *X*-chromosome or on the autosomes) could thus be checked, complied with Ohno's Law. The average proportion of polymorphic loci per species was very similar in *X*-linked loci (0·182, S.E. = 0·067) and autosomal loci (0·167, S.E. = 0·038). Similarly the mean heterozygosity at autosomal loci (0·040, S.E. = 0·015) was very similar to that for *X*-linked loci (0·042, S.E. = 0·019). (The latter was calculated assuming Hardy–Weinberg Law as heterozygotes cannot be detected.)

Thus, in summary, haplodiploid hymenoptera seem to show less genetic variability than diploid insects whereas the data are far less conclusive when comparing *X*-linked and autosomal genes in the same species. However, there is far less usable data on the latter and so on balance, we should expect to find mechanisms that might lead to less genetic variability in haplodiploids and at *X*-linked genes than at autosomal genes. In the following sections, I shall discuss various theoretical results on *X*-linked genes and see how they fit in with the available data.

3. THEORETICAL RESULTS FOR LARGE POPULATIONS

(i) *No selection*

The one well-known result on *X*-linked genes is that if you consider a very large random-mating population and assume that the effects of selection, mutation and migration are negligible, then the difference in gene frequency between males and females is reduced by a half each generation. This was first discussed by Jennings (1916). If you consider two alleles, A_1 and A_2, at an *X*-linked gene and let p_{mt} and p_{ft} be the gene frequencies of A_1 at generation t in males and females respectively, then under this simple model,

$$p_{mt} = \bar{p} + \tfrac{2}{3}(p_{m0} - p_{f0})(-\tfrac{1}{2})^t \tag{3.1}$$

and

$$p_{ft} = \bar{p} - \tfrac{1}{3}(p_{m0} - p_{f0})(-\tfrac{1}{2})^t, \tag{3.2}$$

where

$$\bar{p} = \tfrac{2}{3}p_{f0} + \tfrac{1}{3}p_{m0} = \tfrac{2}{3}p_{ft} + \tfrac{1}{3}p_{mt}.$$

https://doi.org/10.1017/S0016672300026550 Published online by Cambridge University Press

Thus the gene frequencies in the two sexes converge to the average gene frequency and the difference between them gets small quite rapidly.

Under this model the genotypic frequencies at generation $t+1$ are given by

Females			Males	
A_1A_1	A_1A_2	A_2A_2	A_1	A_2
$p_{mt}p_{ft}$	$p_{mt}(1-p_{ft})+p_{ft}(1-p_{mt})$	$(1-p_{mt})(1-p_{ft})$	p_{ft}	$(1-p_{ft})$

and so the Hardy–Weinberg Law in females only holds asymptotically when $p_m = p_f$. In general authors have found little evidence of differences in gene frequencies between sexes or departures from Hardy–Weinberg Law (e.g. Ward, 1980).

(ii) *Directional selection*

Let us assume the following fitness model:

Females			Males	
A_1A_1	A_1A_2	A_2A_2	A_1	A_2
$1-s$	$1-hs$	1	$1-s$	1

$$(0 \leqslant h \leqslant 1,\ s \leqslant 1)$$

For simplicity, I have assumed there to be dosage compensation in fitness which seems to be a reasonable assumption given the available data.

Haldane (1926) and Nagylaki (1979) have considered models such as this. Nagylaki (1979) has shown that, for small s, the gene frequencies in the two sexes converge quickly, whatever the value of h, as they do in the absence of selection. Thus assuming that $p_{mt} \simeq p_{ft} = p_t$, it can easily be shown that the change in the average gene frequency is approximately given by

$$\Delta\bar{p}_t = (2p_{f,t+1}/3 + p_{m,t+1}/3) - (2p_{ft}/3 + p_{mt}/3) \simeq -\frac{s}{3}p_tq_t[(1+2h)+(1-2h)\,2p_t], \tag{3.3}$$

if $p_{ft} \simeq p_{mt} = p_t$ and s is reasonably small. The gene frequency of A_1 in the males changes slightly quicker than in the females as selection acts more directly in the males.

The equivalent expression for autosomal genes is

$$\Delta p_t \simeq -sp_tq_t[h+(1-2h)\,p_t]. \tag{3.4}$$

Since

$$\tfrac{1}{3}[(1+2h)+(1-2h)\,2p_t] > h+(1-2h)\,p_t \quad \text{for all} \quad 0 \leqslant h \leqslant 1 \quad \text{and} \quad 0 \leqslant p_t \leqslant 1,$$

the rate of change in gene frequency, at a particular gene frequency, will be greater in the X-linked case. Thus disadvantageous genes will decline in frequency more quickly, and advantageous genes increase in frequency more quickly, at X-linked

genes than at autosomal genes as one would intuitively expect because the selection acts more directly in the hemizygous males. A similar remark without proof was made by Li (1955).

In section 2, I discussed kangaroos where paternal X-inactivation occurs. In this case, fitnesses are merely determined by the maternal gene and so for the case of directional selection, equal in the two sexes, we have:

Females		Males	
$A_1A_1, A_1^fA_2^m$	$A_1^mA_2^f, A_2A_2$	A_1	A_2
$1-s$	1	$1-s$	1

where the superscripts, m and f, give the sex of the parent from which that gene was derived.

Again the two gene frequencies converge quickly and

$$\Delta\bar{p}_t \simeq \frac{-2s}{3}p_{ft}(1-p_{ft}) \quad \text{for small } s.$$

Comparing this with the expression for autosomal genes with dominance (3.4), the decline in frequency of a deleterious gene (or the increase in an advantageous gene) is less for an X-linked gene than an autosomal gene only if the degree of dominance, h, of the autosomal gene is greater than two-thirds and the gene frequency of the disadvantageous gene, p_t, is small ($< \{(h-\frac{2}{3})/(2h-1)\}$) or if p_t is greater than two-thirds and h is small ($< \{(p_t-\frac{2}{3})/(2p_t-1)\}$). Thus again gene frequencies will generally change more quickly at X-linked genes.

Many disadvantageous genes (e.g. those causing diseases) may well be maintained in populations by directional selection being balanced by mutation. Let us consider the case of a X-linked deleterious recessive gene being maintained by recurrent mutation.

In this case, we can easily show that

$$p_{m,t+1} = (u+p_{ft}(1-s-u))/(1-sp_{ft}) \tag{3.5}$$

and

$$p_{f,t+1} = (u+\tfrac{1}{2}(p_{mt}+p_{ft})(1-u)-sp_{mt}p_{ft})/(1-sp_{mt}p_{ft}), \tag{3.6}$$

where u is the mutation rate from A_2 to A_1 per generation in both sexes and mutation is assumed to occur after selection. If we assume that at equilibrium p_m and p_f are small (effectively assuming that u is small) so that quadratic and higher order terms can be neglected then

$$p_m = u+p_f(1-s-u)$$

and

$$p_f = u+\tfrac{1}{2}(p_m+p_f)(1-u)$$

which leads to

$$p_f = \frac{u(3-u)}{s+u(3-u-s)} \tag{3.7}$$

and

$$p_m = \frac{u(3-u)-2us}{s+u(3-u-s)} \tag{3.8}$$

which are approximated by $p_f = 3u/s$ and $p_m = u(3-2s)/s$ if u is small and $u \ll s$. These expressions were first produced by Haldane (1935). For small s,

https://doi.org/10.1017/S0016672300026550 Published online by Cambridge University Press

$p_m \simeq p_f = 3u/s$ and as s increases, the difference between the equilibrium gene frequencies increases until for recessive lethals $p_f = 3u$ and $p_m = u$. Nagylaki (1977) has also considered this problem. His answers are equivalent but look different as he has calculated the gene frequencies in each sex at birth (i.e. before selection and mutation have acted) whereas Haldane and I have calculated the gene frequencies in gametes produced by each sex (i.e. after selection and mutation have acted).

For autosomal recessives the equivalent expression to (3.5) and (3.6) is

$$p_{t+1} = \frac{p_t + u(1-p_t) - sp_t^2}{1 - sp_t^2}. \tag{3.9}$$

At equilibrium when $p_{t+1} = p_t = p$, this leads to

$$p = 1 \quad \text{or} \quad \sqrt{(u/s)}, \tag{3.10}$$

the latter being the stable equilibrium point if $u < s$.

The incidence of the disease at equilibrium is approximately $p_{fl}/2 = (3u)/(2s)$ for X-linked recessive genes, all affected being male, and $p^2 = u/s$ for autosomal recessive genes. Thus similar selective disadvantage and mutation pressure will lead to a 50% larger disease incidence in X-linked recessives than in autosomal recessives whatever value is the reduction in fitness. However, the frequency of heterozygote (carrier) females will be much less for X-linked diseases than for equivalent autosomal diseases (being approximately $(6u-2s)/s$ and $2\sqrt{(u/s)}$ respectively). If the disease is dominant in the females ($h = 1$) or if there is paternal X-inactivation, there is a stable equilibrium at $p_f \simeq 3u/(2s)$ and $p_m \simeq u(3-s)/(2s)$ giving approximate disease incidences of $9u/(4s)$ and $3u/(2s)$ respectively. While for an autosomal dominant disease gene the disease incidence is approximately $2u/s$. Thus for dominant disease genes, the disease incidence is again higher for X-linked genes than for autosomal genes.

(iii) *Fitness models leading to balanced polymorphisms*

This is that part of the analysis of X-linked genes which has received most attention (e.g. Bennett (1958); Haldane & Jayakar (1964); Cannings (1967); Pamilo (1979)). There are essentially two fitness models that can produce a balanced polymorphism, one involves overdominance in the females and the other has selection acting in opposite directions in the two sexes. True overdominance has generally not been found at autosomal loci but effective overdominance can be produced by variation in fitness over time or space or by the effect of closely linked loci. Thus it seems valid to consider effective overdominance as a possible force for maintaining variability when comparing X-linked and autosomal loci. If, for simplicity, we again assume dosage compensation so that homozygous females have the same fitness as the equivalent hemizygous males then we have the following fitness model for X-linked genes,

Females			Males		
A_1A_1	A_1A_2	A_2A_2	A_1	A_2	(A)
$1-s_1$	1	$1-s_2$	$1-s_1$	$1-s_2$	

https://doi.org/10.1017/S0016672300026550 Published online by Cambridge University Press

For autosomal genes, any degree of overdominance i.e. $1 > s_1, s_2 > 0$, will lead to a balanced polymorphism while at X-linked genes two further conditions need to be satisfied, $s_2 < 3s_1 - 2s_1^2$ and $s_1 < 3s_2 - 2s_2^2$ (Pamilo, 1979), which imply that s_1 and s_2 must not differ too much (see Fig. 1). This need for symmetry is similar to the results of Robertson (1962), who showed that for autosomal genes in finite populations, very asymmetric overdominance can in fact increase the rate of fixation of alleles relative to neutral alleles rather than decrease it as symmetric overdominance does.

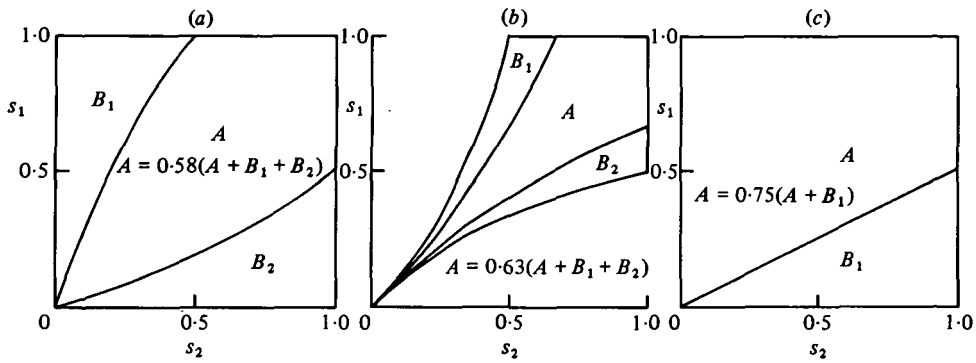

Fig. 1. Regions for the parameter values which lead to a balanced polymorphism; A for X-linked genes and $A + B_1 + B_2$ for autosomal genes. (*a*) Holds for overdominant selection (model A), (*b*) for selection without dominance in opposite directions in the two sexes (model B) and (*c*) for completely recessive selection in opposite directions in the two sexes (model C).

If selection acts in opposite directions in the two sexes then a balanced polymorphism can be produced at X-linked genes but the restrictions on the parameter values are stricter than for autosomal genes. Figure 1(*b*) gives the allowable values for selection without dominance at X-linked genes, i.e.

Females			Males		(B)
A_1A_1	A_1A_2	A_2A_2	A_1	A_2	
1	$1-s_1/2$	$1-s_1$	$1-s_2$	1	

The conditions for a balanced polymorphism with no dominance selection are $0 < s_1 < s_2/(1-s_2)$ and $0 < s_2 < s_1/(1-s_1)$ for autosomal genes and $0 < s_1 < 2s_2/(2-s_2)$ and $0 < s_2 < 2s_1/(2-s_1)$ for X-linked genes (Pamilo, 1979). Figure 1(*c*) gives the allowable values when recessive selection in the females is balanced by selection in the opposite direction in the males, i.e.

Females			Males		(C)
A_1A_1	A_1A_2	A_2A_2	A_1	A_2	
1	1	$1-s_1$	$1-s_2$	1	

https://doi.org/10.1017/S0016672300026550 Published online by Cambridge University Press

The conditions for a balanced polymorphism with completely recessive selection are $0 < s_1, s_2 < 1$ for autosomal genes and $0 < s_2 < \text{Min}\,(2s_1, 1)$ and $0 < s_1 < 1$ for X-linked genes (Pamilo, 1979). The figures given in figure 1 on the relative sizes of the parameter spaces necessary for balanced polymorphisms must be treated carefully as there is no reason why any particular pair of parameter values should be equally likely.

Thus there are fitness models which can maintain genetic variability at X-linked loci but the conditions on the fitnesses are more restrictive than for autosomal genes. Therefore if similar selective forces act at X-linked and autosomal loci proportionally fewer X-linked loci should be polymorphic. The discussion of selection acting in opposite directions in the two sexes may be rather misleading as I am unaware of a proven case of a gene being maintained by differential selection in the two sexes. One attempt to look for evidence of this was reported by Curtsinger (1980) using data from Kerr & Kerr (1952) and Wilton (1979). They used balancer chromosomes in Drosophila where females homozygous for the balancer chromosome are inviable but the balancer males are nearly normal. Curtsinger (1980) thus assumed the following fitness model:

Females			Males	
XX	XX_B	X_BX_B	X	X_B
$1+z$	$1-z$	0	$1+y$	$1-y$

He then estimated z, y and y/z for a total of 121 wild-type X-chromosomes. The average value for y/z was 1·87 (S.E. = 0·94) for Kerr & Kerr's (1952) data and 1·98 (S.E. = 1·42) for Wilton's (1979) data. Thus selection seems to act in the same direction in both males and females. Only 1 of the 121 chromosomes gave strong evidence of selection in opposite directions in the two sexes ($y/z = -6{\cdot}14$). However, as the estimated values of y/z for each chromosome have quite large standard errors (at least 2·5 for this particular chromosome), it is quite possible that this value has occurred by chance when selection is acting in the same direction in both sexes. The balancer chromosomes are however unusual in causing inviability in homozygous females and so we should not put too much weight on this data.

When there is paternal X-inactivation, overdominance is not possible and only balancing selection in the two sexes will produce a balanced polymorphism, i.e. if the fitnesses are defined as:

Females		Males	
$A_1A_1, A_1^fA_2^m$	$A_1^mA_2^f, A_2A_2$	A_1	A_2
1	$1-s_1$	$1-s_1$	1

The conditions for maintaining a stable polymorphism are as for an autosomal gene with selection without dominance (see Fig. 1 (*b*)), i.e. $0 < s_1 < s_2/(1-s_2)$ and $0 < s_2 < s_1/(1-s_1)$ (Cooper, 1976). Thus with paternal X-inactivation it seems to

be even more difficult for selection to maintain genetic variability. Crozier & Pamilo (1979) have considered two rather specific frequency-dependent selection models at paternal X-inactivated loci and have shown that, under certain conditions, a balanced polymorphism is possible. More general frequency dependent selection models may also allow balanced polymorphisms to occur at X-linked loci with or without paternal X-inactivation. However, it is difficult to judge how important frequency dependence may be at maintaining genetic variability.

4. THEORETICAL RESULTS FOR FINITE POPULATIONS

(i) *Neutral models*

An exact analysis of X-linked genes in finite populations is quite complex because of the necessity to consider the gene frequencies in the two sexes separately. Hence relatively little work has been published in this area. Wright (1933) considered the case of a random mating finite population of M males and F females with no selection, mutation and migration. He used path coefficients to find equations for the rate of decrease of heterozygosity and hence developed a measure of effective population size. A slightly more straightforward and informative approach is to use sampling theory.

Given the frequencies at generation $t-1$, the gene frequencies of an allele, A_1, at generation t are $p_{mt} = R_1/M$ where R_1 is a Binomially distributed random variable, sample size M and probability of 'success' $p_{f,t-1}$ and $p_{ft} = (R_2+R_3)/(2F)$ where R_2 is Binomially distributed, sample size F and probability $p_{f,t-1}$ and R_3 is Binomially distributed, sample size F and probability $p_{m,t-1}$. Similarly, the numbers of each genotype in the females have a trinomial distribution conditional on the frequencies in generation $t-1$.

It follows very easily from sampling theory that

$$E(p_{mt}-p_{ft}) = -\tfrac{1}{2}E(p_{m,t-1}-p_{f,t-1}) = (-\tfrac{1}{2})^t(p_{m0}-p_{f0}) \tag{4.1}$$

and

$$E(2p_{ft}/3+p_{mt}/3) = E(\bar{p}_t) = E(\bar{p}_{t-1}) = \bar{p}_0, \tag{4.2}$$

i.e. the deterministic results hold on average. As fixation of one gene must eventually occur and $E(\bar{p}_t) = \bar{p}_0$ for all t, $p(p_{mt} = p_{ft} = 1) = \bar{p}_0$ and $p(p_{mt} = p_{ft} = 0) = (1-\bar{p}_0)$ as $t \to \infty$. Thus probabilities of ultimate fixation depend for X-linked genes on the initial average gene frequencies.

If H_{t+1} is the frequency of heterozygotes in the F females at generation, $t+1$,

$$E(H_{t+1}) = E(p_{mt}(1-p_{ft})+p_{ft}(1-p_{mt})), \tag{4.3}$$

and it is then straightforward to show that

$$\begin{bmatrix} E(2p_{ft}(1-p_{ft})) \\ E(2p_{mt}(1-p_{mt})) \\ E(H_{t+1}) \end{bmatrix} = \begin{bmatrix} (1-1/F)/4 & (1-1/F)/4 & \tfrac{1}{2} \\ (1-1/M) & 0 & 0 \\ \tfrac{1}{2} & 0 & \tfrac{1}{2} \end{bmatrix} \begin{bmatrix} E(2p_{f,t-1}(1-p_{f,t-1})) \\ E(2p_{m,t-1}(1-p_{m,t-1})) \\ E(H_t) \end{bmatrix}. \tag{4.4}$$

Explicit expressions for these three quantities ($E(H_{t+1})$, $E(2p_{ft}(1-p_{ft}))$ and $E(2p_{mt}(1-p_{mt}))$) can be produced in terms of the eigenvalues of the above matrix. As the eigenvalues are all less than one in modulus, these quantities tend to zero

as $t \to \infty$, and the asymptotic rate of decline of each is given by the largest eigenvalue of the matrix. From the above matrix, you can show that the eigenvalues, λ, satisfy the equation,

$$\lambda^3 - \lambda^2(1-2c_1) - \lambda(c_1+2c_2) + c_2 = 0 \tag{4.5}$$

where $c_1 = (1+1/F)/8$ and $c_2 = (\frac{1}{8})(1-1/M)(1-1/F)$. If you replace λ by $(1+x)$ the resultant equation is identical to that of Wright (1933). For reasonably large values for M and F, this quantity, x, becomes small and is approximately given by $-(2/F+1/M)/9$. A definition of inbreeding effective population size (N_{ex}) can then be obtained by equating this expression to $-1/(2N_{ex})$. This gives

$$N_{ex} = \frac{9MF}{2(2M+F)} \tag{4.6}$$

(as obtained by Nagylaki (1981) using a different method) as compared with $4MF/(M+F)$ for autosomal loci. When $M = F = N/2$, $N_{ex} = 3N/4$ for X-linked loci and N for autosomal loci. Thus the effective population size for X-linked loci is less than for autosomal loci unless the number of females is much larger than the number of males ($F > 7M$). Also from the above calculations, it is straightforward to show that

$$\begin{aligned}\text{Var}\,(2p_{ft}/3 + p_{mt}/3) &= \text{Var}\,(\bar{p}_t) \\ &= \bar{p}_0(1-\bar{p}_0) - (2E(2p_{ft}(1-p_{ft})) + E(2p_{mt}(1-p_{mt}))/2 + 2E(H_{t+1}))/9\end{aligned} \tag{4.7}$$

and

$$\begin{aligned}\text{Var}\,(p_{mt}-p_{ft}) = -(-\tfrac{1}{2})^t(p_{m_0}-p_{f_0}) + E(H_{t+1}) \\ -\tfrac{1}{2}E(2p_{ft}(1-p_{ft})) - E(2p_{mt}(1-p_{mt}))\end{aligned}, \tag{4.8}$$

i.e. are related to the three quantities used before and thus asymptotically change at the same rate as the expected frequency of heterozygotes. Var $(\bar{p}_t) \to \bar{p}_0(1-\bar{p}_0)$ and Var $(p_{mt}-p_{ft}) \to 0$ as $t \to \infty$. An alternative definition of effective population size can be obtained from the variance of the change in gene frequency per generation, i.e. by equating Var $(\Delta\bar{p}_t | p_{mt}, p_{ft})$ to $\bar{p}_t(1-\bar{p}_t)/(2N_{eX})$.

Using sampling theory,

$$\begin{aligned}\text{Var}\,(\Delta\bar{p}_t) &= (\tfrac{1}{9})\left\{\frac{p_{ft}(1-p_{ft})}{F} + \frac{p_{mt}(1-p_{mt})}{F} + \frac{p_{ft}(1-p_{ft})}{M}\right\} \\ &\simeq \tfrac{1}{9}\bar{p}_t(1-\bar{p}_t)\left(\frac{2}{F}+\frac{1}{M}\right)\end{aligned} \tag{4.9}$$

if $p_{mt} \simeq p_{ft} = \bar{p}_t$ which as we have seen from deterministic and stochastic results is generally true. Thus $1/(2N_{ex}) = (2/F+1/M)/9$ and the same value of N_{eX} is obtained as before.

In principle, sex-linked genes can be studied by modelling the problem as a Markov chain as suggested by Yang, Lee & Chee (1966). However, the number of possible states is $(M+1)\times(2F+1)$ and thus it is only a feasible approach when M and F are small. Its advantage is that it can be adapted to include the effects of selection as we shall see later.

Assuming an infinite allele model, i.e. each new mutation produces a novel allele,

https://doi.org/10.1017/S0016672300026550 Published online by Cambridge University Press

Nagylaki (1981) showed that for X-linked genes the expected frequency of heterozygotes in females was, for weak mutation and large population size, approximately given by

$$E(H) = \frac{4N_{eX}u}{1+4N_{eX}u}, \tag{4.10}$$

where u is the mutation rate per generation and N_{eX} is the effective population size as defined by equation (4.6). This is the same expression as produced for autosomal loci (Kimura & Crow, 1964) except that there is a different definition for effective population size. When $M = F = N/2$, $E(H)$ for sex linked genes $= 3Nu/(1+3Nu)$ (Mayo(1976)) and the expected proportion of heterozygotes in females at X-linked loci, H_X, satisfies $\frac{3}{4}H_a < H_X < H_a$ where H_a is the expected proportion of heterozygotes at autosomal loci. Thus, even ignoring the effect of selective forces, you would expect to get a lower proportion of heterozygotes at X-linked loci than at autosomal loci due to having effectively fewer copies of each gene.

Stewart (1976), using the infinite allele model, showed that the between population variance in heterozygosity at a single neutral autosomal locus was given by

$$\text{Var } H \simeq \frac{8N_e u}{(1+4N_e u)^2(2+4N_e u)(3+4N_e u)}.$$

Following Nagylaki (1981), an approximation for the variance at a single sex-linked locus should be given, for reasonably large N_{eX} and small u, by

$$\text{Var } H \simeq \frac{8N_{eX} u}{(1+4N_{eX} u)^2(2+4N_{eX} u)(3+4N_{eX} u)}.$$

To compare theoretical values with those given in table 1, we need to calculate the variation in mean heterozygosity over loci. If we assume the species in the various groups have approximately similar effective population sizes and the loci chosen are unlinked (a reasonable assumption generally as the number of loci studied is not large relative to the number of chromosomes) then the between species variation in heterozygosity should be approximately given by

$$\text{Var } (\bar{H}) \simeq \frac{1}{\bar{n}} \text{Var } H.$$

where $\bar{n}$ is the harmonic mean of the number of loci studied for each species in the species group

$$\left(\text{i.e. } \frac{1}{\bar{n}} = \left(\frac{1}{n_1}+\frac{1}{n_2}+\ldots+\frac{1}{n_k}\right)/k\right).$$

Equating the observed H to $E(H)$ as done by Nei, Fuerst & Chakraborty (1976) for autosomal data, we can estimate $4N_{eX}u$ and hence predict Var $(\bar{H})$. The observed and predicted values are given in Table 3. Given the assumptions made, the agreement is surprisingly good and suggests that a lot of the variability might be due to mutation and random genetic drift. Though some selection models may lead to similar predictions. A similar approach for the diploid data does not

Table 3. *Observed and predicted values for the between species standard deviation* (S.D.) *in mean heterozygosity for three haplodiploid groups*

Category	Standard deviation		Overall mean heterozygosity	$\bar{n}$*
	Observed	Predicted		
Solitary	0·026	0·030	0·039	13·1
Primitively eusocial	0·018	0·017	0·011	12·8
Advanced eusocial	0·025	0·033	0·045	12·4

* $\bar{n}$ = harmonic mean of number of loci studied.

produce a very good agreement between observed and predicted values. This is probably due to the great diversity of effective population sizes within the diploid groups.

(ii) *Models including selection*

Exact analyses for X-linked genes are complex. However, as we discussed earlier when looking at selection effects in large populations or at neutral models in finite populations, the gene frequencies in the two sexes generally converge quickly and do not tend to differ greatly due to the finiteness of the population. If we assume that $p_m \simeq p_f$ then we can use a univariate diffusion approximation.

If we consider directional selection with dosage compensation and no dominance, i.e.

Females			Males	
$\frac{A_1A_1}{1+s}$	$\frac{A_1A_2}{1+s/2}$	$\frac{A_2A_2}{1}$	$\frac{A_1}{1+s}$	$\frac{A_2}{1}$

then using equation (3.3),

$$E(\bar{p}_{t+1}-\bar{p}_t) = E(\Delta\bar{p}_t) \simeq \frac{2s\bar{p}_t(1-\bar{p}_t)}{3} \quad (\text{if } p_{mt} \simeq p_{ft} \simeq \bar{p}_t) \tag{4.11}$$

for small s and from equation (4.9),

$$\text{Var}\,(\Delta\bar{p}_t) \simeq \frac{\bar{p}_t(1-\bar{p}_t)}{2N_{eX}}, \tag{4.12}$$

where N_{eX} is as given in equation (4.6), if we assume that $1/N_{eX}$ is of the same order as s. From Kimura (1964) the probability of ultimate fixation of A_1 is then given by

$$u(\bar{p}_0) = \int_0^{\bar{p}_0} G(\bar{p}_t)\,d\bar{p}_t \Big/ \int_0^1 G(\bar{p}_t)\,d\bar{p}_t,$$

(where $G(\bar{p}_t) = \exp\{-\int[2E(\Delta\bar{p}_t)/\text{Var}\,(\Delta\bar{p}_t)]\,d\bar{p}_t\},$)

$$= \int_0^{\bar{p}_0} \exp(-8N_{eX}s\bar{p}_t/3)\,d\bar{p}_t \Big/ \int_0^1 \exp(-8N_{eX}s\bar{p}_t/3)\,d\bar{p}_t,$$

$$= (1-\exp(-8N_{eX}s\bar{p}_0/3))/(1-\exp(-8N_{eX}s/3)), \tag{4.13}$$

as compared to $u(p_0) = (1-e^{-2N_e s\bar{p}_0})/(1-e^{-2N_e s})$

https://doi.org/10.1017/S0016672300026550 Published online by Cambridge University Press

for autosomal genes with selection without dominance. A similar approach for a rather more specific problem was taken by Hedrick (1980). When the numbers of each sex are equal, $N_{eX} = 3N/4$, and $N_e = N$ for autosomal genes. On substituting in (4.13), we find that $u(\bar{p}_0) = (1-e^{-2Ns\bar{p}_0})/(1-e^{-2Ns})$, i.e. is identical to the result for autosomal genes given the same initial average gene frequency. If however you consider the probability of fixation of a single advantageous mutant gene then as $\bar{p}_0 = 2/(3N)$ for X-linked genes and $\bar{p}_0 = 1/(2N)$ for autosomal genes, the probability of ultimate fixation for large N and small s is approximately $4s/3$ for X-linked genes and s for autosomal genes i.e. an X-linked advantageous mutant is more likely to be fixed due to its higher initial gene frequency. Some similar general statements were made without proof by Lester & Selander (1979). Equivalent expressions to (4.13) can be derived for other selection models.

To check some of the assumptions in the above analysis, I used the Markov chain approach mentioned previously for very small populations. For M males and F females, there are $(M+1)\times(2F+1)$ possible values for the gene frequencies of a particular allele, A_1. If $\mathbf{q}_t = (q^{0,0}, q_t^{1,0}, q_t^{2,0}, \ldots, q_t^{M,2F})$ where $q_t^{i,j}$ is the probability of having i copies of A_1 in the males and j copies in the females at time t then using sampling theory we can derive the elements of the matrix $\mathbf{P}$ such that $\mathbf{q}_t = \mathbf{q}_{t-1}\mathbf{P}$. If $\mathbf{u} = (u^{1,0}, u^{2,0}, \ldots, u^{M,2F-1})$ is a column vector of the probabilities of ultimate fixation of A_1 from all the states excluding (0, 0) and $(M, 2F)$ then

$$\mathbf{u} = \mathbf{P}^*\mathbf{u}+\mathbf{K}$$

where $\mathbf{P}^*$ is the matrix $\mathbf{P}$ with the first and last rows and columns removed and $\mathbf{K}$ is a column vector derived from the last column of $\mathbf{P}$ with the first and last element removed. Then

$$\mathbf{u} = (\mathbf{I}-\mathbf{P}^*)^{-1}\mathbf{K} \tag{4.14}$$

where $\mathbf{I}$ is the identity matrix.

Table 4 gives approximate expressions for $u(\bar{p}_0)$ from equation (4.13) and exact expressions from (4.14) for various values of M, F and s. Three starting conditions were considered: a single mutant gene in the males, a single mutant gene in the females and an initial average gene frequency of 0·5. The approximation is surprisingly good even for small values of N_{eX}. From table 4 or by using Kimura's (1964) approximate expression for autosomal genes, we can see that

$$u(\bar{p}_0) \simeq \bar{p}_0+\frac{8N_{eX}\,s}{3}\frac{\bar{p}_0(1-\bar{p}_0)}{2} \tag{4.15}$$

for small s.

Using the approximate expressions for the mean change and variance in change of the gene frequency, (4.11) and (4.12), under directional selection at X-linked loci, we can also compare the mean times to fixation given fixation occurs, $\bar{t}_f$, and the mean time to extinction given extinction occurs, $\bar{t}_e$, between X-linked and autosomal loci. These can be calculated using the equations of Kimura & Ohta (1969*a*, *b*), i.e.

$$\bar{t}_f = \int_{\bar{p}_0}^{1} \psi(\bar{p}_t)\,u(\bar{p}_t)\,(1-u(\bar{p}_t))\,d\bar{p}_t+\frac{1-u(\bar{p}_0)}{u(\bar{p}_0)}\int_0^{\bar{p}_0} \psi(\bar{p}_t)\,[u(\bar{p}_t)]^2\,d\bar{p}_t$$

https://doi.org/10.1017/S0016672300026550 Published online by Cambridge University Press

Table 4. *Probability of ultimate fixation of* A_1 *with directional selection without dominance. 'Exact' values are calculated from the Markov Chain approach and 'Approx.' values from the Diffusion Approximation*

			$p_{m0} = 1/M, p_{f0} = 0$				$p_{m0} = 0, p_{f0} = 1/(2F)$				$\bar{p}_0^* = \frac{1}{2}$			
			$s = 0{\cdot}01$		$s = 0{\cdot}1$		$s = 0{\cdot}01$		$s = 0{\cdot}1$		$s = 0{\cdot}01$		$s = 0{\cdot}1$	
M	F	N_{eX}	Exact	Approx	Exact	Approx	Exact	Approx	Exact	Approx	Exact	Approx	Exact	Approx
1	1	1·5	0·3383	0·3378	0·3821	0·3786	0·3383	0·3378	0·3821	0·3786	—	—	—	—
2	1	1·8	0·1700	0·1700	0·2000	0·2017	0·3392	0·3387	0·3904	0·3879	0·5062	0·5060	0·5593	0·5597
1	2	2·25	0·3400	0·3400	0·3987	0·4018	0·1708	0·1709	0·2090	0·2109	0·5075	0·5075	0·5710	0·5744
2	2	3	0·1724	0·1723	0·2252	0·2267	0·1725	0·1723	0·2272	0·2267	0·5104	0·5100	0·5983	0·5987
4	4	6	0·0896	0·0896	0·1533	0·1564	0·0897	0·0896	0·1552	0·1564	0·5203	0·5200	0·6852	0·6900
12	3	6	0·0299	0·0300	0·0521	0·0545	0·1193	0·1192	0·2021	0·2041	0·5200	0·5200	0·6830	0·6900
2	8	6	0·1776	0·1780	0·2841	0·2933	0·0448	0·0449	0·0771	0·0808	0·5190	0·5200	0·6750	0·6900
6	6	9	0·0621	0·0621	0·1335	0·1373	0·0621	0·0621	0·1353	0·1373	0·5302	0·5300	0·7613	0·7685

* Average over various possible starting values.

Table 5. *Approximate mean times to fixation given that fixation occurs for autosomal and* X*-linked loci,* $(p_0 = \bar{p}_0 = 0{\cdot}25)$

	$M = 10, F = 10$			$M = 8, F = 20$			$M = 100, F = 100$		
Selection coefficient s	Autosomal $\bar{t}_f$ $(N_e = 20)$	X-linked $\bar{t}_{fX}$ $(N_{eX} = 15)$	$\bar{t}_{fX}/\bar{t}_f$	Autosomal $\bar{t}_f$ $(N_e = 22{\cdot}86)$	X-linked $\bar{t}_{fX}$ $(N_{eX} = 20)$	$\bar{t}_{fX}/\bar{t}_f$	Autosomal $\bar{t}_f$ $(N_e = 200)$	X-linked $\bar{t}_{fX}$ $(N_{eX} = 150)$	$\bar{t}_{fX}/\bar{t}_f$
0	69·04	51·78	0·750	78·91	69·04	0·875	690·44	517·83	0·750
0·01	68·87	51·65	0·750	78·65	68·73	0·874	561·70	421·27	0·750
0·05	65·01	48·76	0·750	73·05	62·32	0·853	196·05	147·04	0·750
0·1	56·17	42·13	0·750	61·12	50·02	0·817	37·15	27·86	0·750
0·25	34·60	25·62	0·750	35·35	27·44	0·776	—	—	—

Table 6. *Approximate mean times to extinction given extinction occurs for autosomal and* X*-linked loci* $(\bar{p}_0 = p_0 = 0{\cdot}25)$

	$M = 10, F = 10$			$M = 8, F = 20$			$M = 100, F = 100$		
Selection coefficient s	Autosomal $\bar{t}_e$ $(N_e = 20)$	X-linked $\bar{t}_{eX}$ $(N_{eX} = 15)$	$\bar{t}_{eX}/\bar{t}_e$	Autosomal $\bar{t}_e$ $(N_e = 22{\cdot}86)$	X-linked $\bar{t}_{eX}$ $(N_{eX} = 20)$	$\bar{t}_{eX}/\bar{t}_e$	Autosomal $\bar{t}_e$ $(N_e = 200)$	X-linked $\bar{t}_{eX}$ $(N_{eX} = 150)$	$\bar{t}_{eX}/\bar{t}_e$
0·0	36·97	27·73	0·750	42·25	36·97	0·875	369·68	277·26	0·750
0·01	36·85	27·64	0·750	42·07	36·75	0·874	286·23	214·67	0·750
0·05	34·24	25·68	0·750	38·31	32·48	0·848	106·55	79·91	0·750
0·1	28·62	21·47	0·750	30·88	25·04	0·811	65·36	49·02	0·750
0·25	17·02	12·77	0·750	17·80	14·04	0·789	33·12	24·84	0·750

and

$$\bar{t}_e = \int_0^{\bar{p}_0} \psi(\bar{p}_t)\, u(\bar{p}_t)\,(1-u(\bar{p}_t))\, d\bar{p}_t + \frac{u(\bar{p}_0)}{1-u(\bar{p}_0)} \int_{\bar{p}_0}^{1} \psi(\bar{p}_t)\,(1-u(\bar{p}_t))^2\, d\bar{p}_t$$

where

$$\psi(\bar{p}_t) = 2\int_0^1 G(x)\, dx/(\mathrm{Var}\,(\Delta\bar{p}_t)\, G(\bar{p}_t)),$$

$$G(\bar{p}_t) = \exp\{-\int 2E(\Delta\bar{p}_t)/\mathrm{Var}\,(\Delta\bar{p}_t)\, d\bar{p}_t\}$$

and

$$u(\bar{p}_t) = \int_0^{\bar{p}_t} G(x)\, dx \Big/ \int_0^1 G(x)\, dx$$

is the probability of ultimate fixation given initial frequency $\bar{p}_t$, $\bar{t}_e$ and $\bar{t}_f$ can only be calculated using numerical integration except for the case of $s = 0$. Results for some particular parameter values are given in Tables 5 and 6. For equal numbers of each sex, both mean times are reduced by a quarter for X-linked genes for any value of N, s or $\bar{p}_0$ due to the fact that $N_{eX}/N_e = 0{\cdot}75$ when $M = F$. For unequal numbers of each sex, the mean times are reduced in proportion to N_{eX}/N_e for small s and by more than that for larger s. Comparing the mean times when $N_{eX} = N_e$ (e.g. $\bar{t}_{eX}$ and $\bar{t}_{fX}$ for $M = 8$, $F = 20$ with $\bar{t}_e$ and $\bar{t}_f$ for $M = F = 10$) shows that the values for X-linked genes are slightly lower than for the autosomal genes due to selection acting more directly. Thus as we found when looking at deterministic rates of change of gene frequency, X-linked genes are lost or fixed more quickly than autosomal genes starting from the same initial frequency. This confirms the suggestion of Lester & Selander (1979).

Other selection models, such as those which maintain polymorphisms in large populations, can similarly be studied for small populations. An interesting problem to look at is to find those models which considerably reduce the rate of loss of genetic variability. Robertson (1962) considered the rate of fixation and loss of genes for the case of overdominance at autosomal loci (i.e. relative fitnesses of $1-s_1'$, 1, $1-s_2'$ for A_1A_1, A_1A_2 and A_2A_2 respectively where $s_1', s_2' > 0$) and showed that, if $s_1'/(s_1'+s_2') > 0{\cdot}8$ or $< 0{\cdot}2$, overdominance could actually increase the rate of fixation compared with a neutral model depending on the actual value of N_e and that only when s_1' and s_2' were reasonably similar was fixation and loss considerably slowed. The models considered earlier for maintaining variability at X-linked genes can be approximated such that Robertson's results can be used for X-linked genes.

For overdominance with dosage compensation (model A),

$$E(\Delta\bar{p}_t) \simeq \tfrac{1}{3}\bar{p}_t\bar{q}_t(3s_2 - s_1 - 2(s_1+s_2)\bar{p}_t),$$

and for selection in opposite directions in the two sexes, completely recessive in the females (model C),

$$E(\Delta\bar{p}_t) \simeq \tfrac{1}{3}\,\bar{p}_t\,\bar{q}_t(2s_1 - s_2 - 2s_1\bar{p}_t),$$

assuming in both equations that s_1 and s_2 are reasonably small and $p_{mt} \simeq p_{ft} \simeq \bar{p}_t$. For overdominance at an autosomal locus,

$$E(\Delta p_t) \simeq p_t\, q_t(s_2' - (s_1'+s_2')\, p_t)$$

for reasonably small s_1' and s_2'. Thus model A is equivalent to autosomal overdominance with $s_1' = 3s_1 - s_2$ and $s_2' = 3s_2 - s_1$. For a balanced polymorphism, we need $s_1', s_2' > 0$, i.e. $s_1 < 3s_2$ and $s_2 < 3s_1$, which is what we obtained earlier, ignoring quadratic terms. If we further restrict values such that $0{\cdot}2 < s_1'/(s_1' + s_2') < 0{\cdot}8$ following Robertson (1962), we need $s_1 < \frac{13}{7} s_2$ and $s_2 < \frac{13}{7} s_1$ in order to slow the rate of fixation of genes relative to the neutral model. Similarly for model C, $s_1' = \frac{1}{3} s_2$ and $s_2' = \frac{2}{3} s_1 - \frac{1}{3} s_2$ and the restrictions on the parameters reduce from $0 < s_2 < 2s_1$ for a balanced polymorphism to $s_2 < 8s_1/5$ and $s_1 < 5s_2/2$ to reduce the rate of fixation. In both cases the reduction in the allowable parameter space is similar to that for autosomal genes but slightly larger. Thus the range of possible parameter values which can lead to maintenance of genetic variability has been further reduced.

5. DISCUSSION

In this paper we have found that, if directional selection acts on an X-linked gene or a gene in haplodiploids, disadvantageous alleles are lost more quickly and advantageous alleles are fixed more quickly than at autosomal loci because selection acts more directly in the males and the effective number of copies of each gene is less for X-linked genes. This may partially explain why differences between some closely-related species, which can still interbreed, have been found to be mainly on the X-chromosome (e.g. Grula & Taylor (1980)) even when the X-chromosome forms a relatively small part of the genome.

We have also shown that rare deleterious genes maintained by mutation will have a higher incidence of the deleterious form than at equivalent autosomal loci making detection of such mutants somewhat easier. Further in finite populations, e.g. laboratory stocks, James (1979) has shown that any recessive mutation that occurs at X-linked loci will either be detected very quickly (i.e. within 3 generations) or not be detected at all due to the hemizygous state in the males. In contrast, at autosomal loci, many generations may pass before a recessive mutant is detected (Robertson, 1978). These points plus the fact that the initial frequency of a mutant will be higher at X-linked loci due to the reduced number of copies leads us to expect proportionally more loci to be detected on the X-chromosome than on the autosomes as McKusick (1975) found in humans.

Electrophoretic data obtained from haplodiploids suggests quite strongly that haplodiploids have a lower level of heterozygosity than similar diploid insects. There is, however, little data yet comparing X-linked and autosomal loci in the same species. What is available is somewhat inconclusive but does not show any marked differences. When looking at theoretical selection models which can maintain genetic variability, it is clear that there are many more restrictions on the allowable parameter values for X-linked loci than autosomal loci and so, if similar selective forces occur at X-linked and autosomal loci, we would expect fewer loci to remain polymorphic due to such forces and therefore a lower level of heterozygosity is to be expected. Neutral models where variability is maintained by mutation also lead us to expect reduced heterozygosity due to the reduced effective population size. This may not however be relevant when comparing across

species whose effective population size may be different anyway. Thus theory leads us to expect the lower level of genetic variability reported but, as with autosomal loci, it is still somewhat debatable whether selection or mutation and random drift are the important forces in maintaining the variability which does occur. However, in marsupials, where there is paternal *X*-inactivation and the range of selection models that can maintain a polymorphism is very limited, the fact that a similar level of variability has been observed at both autosomal and *X*-linked loci suggests that a lot of this variability (at both autosomal and *X*-linked loci) is probably being maintained by the joint action of mutation and random genetic drift. Further, as the between species variation in heterozygosity seems to fit with what we would expect from the neutral theory, this also suggests that mutation and drift may be more important than selective forces in maintaining variability at *X*-linked genes and genes in haplodiploids.

In this paper, only single locus results have been considered. Very little data on linkage disequilibrium in haplodiploids or between *X*-linked genes is available and very little theoretical work has been done. The main result (Bennett, 1963) so far obtained is that the effective recombination fraction is $\frac{2}{3}$ of the recombination fraction in females as no crossing over occurs in males.

In conclusion, it seems sensible to consider the question as to why haplodiploidy and *X*-linkage occur as they allow the effect of selection to act more directly reducing individuals' (usually males') fitnesses. The *X*-chromosome is, for higher organisms at least, only a small part of the genome and so such deleterious effects will generally be small and may be outweighed by the advantages of such a chromosomal sex-determining mechanism. Haplodiploids however clearly have a larger selective load on them. Bull (1981) has considered the evolution of haplodiploidy and has shown that haplodiploidy can evolve as long as the fitness of the haploid males is greater than half the fitness of diploid males. This is likely to be true in relatively inbred populations where the occurrence of deleterious recessives will be low. However, a crucial condition for the evolution of haplodiploidy is that individuals must be able to develop from unfertilized eggs which seems to be very rare in higher organisms.

I would like to thank Professor R. N. Curnow for encouragement during work on this paper and for comments on early drafts, and O. Mayo and the referee for their helpful comments on the first version of the paper.

REFERENCES

Bennett, J. H. (1958). The existence and stability of selectively balanced polymorphisms at a sex-linked locus. *Australian Journal of Biological Sciences* **11**, 598–602.

Bennett, J. H. (1963). Random mating and sex linkage. *Journal of Theoretical Biology* **4**, 28–36.

Berkelhammer, R. C. (1983). Intraspecific genetic variation and haplodiploidy, eusociality and polygyny in the Hymenoptera. *Evolution* **37**, 540–545.

Bull, J. J. (1981). Co-evolution of haplodiploidy and sex determination in the Hymenoptera. *Evolution* **35**, 568–580.

Cannings, C. (1967). Equilibrium, convergence and stability at a sex-linked locus under Natural Selection. *Genetics* **56**, 613–618.

Carson, H. L. (1955). Variation in genetic recombination in natural populations. *Journal of Cellular and Comparative Physiology* **45**, (Suppl. 2), 221–235.

COOPER, D. W. (1976). Studies on Metatherian Sex Chromosomes: II. The improbability of a stable balanced polymorphism at an *X*-linked locus with the paternal *X*-inactivation system of Kangaroos. *Australian Journal of Biological Sciences* **29**, 245–250.

COOPER, D. W., JOHNSTON, P. G., SHARMAN, G. B. & VANDEBERG, J. L. (1977). The control of gene activity on eutherian and metatherian *X*-chromosomes: a comparison. In *Reproduction and Evolution* (ed. J. H. Calaby and C. H. Tyndale-Biscoe). Proceedings of the Fourth International Symposium on Comparative Biology of Reproduction, Australian Academy of Sciences, Canberra, 81–87.

COOPER, D. W., JOHNSTON, P. G., VANDEBERG, J. L., MAYNES, G. M. & CHEW, G. K. (1979). A comparison of genetic variability at *X*-linked and autosomal loci in Kangaroos, Man and Drosophila. *Genetical Research* **33**, 243–252.

CROZIER, R. H. & PAMILO, P. (1979). Frequency dependent models for *X*-linked loci in Kangaroos. *Australian Journal of Biological Sciences* **32**, 469–474.

CURTSINGER, J. W. (1980). On the opportunity for polymorphism with sex-linkage or haplo-diploidy. *Genetics* **96**, 995–1006.

GRULA, J. W. & TAYLOR, O. R. (1980). The effect of *X*-chromosome inheritance on mate selection behaviour in the sulfur butterflies, *Colias Eurytheme* and *C. Philodice*. *Evolution* **34**, 688–695.

HALDANE, J. B. S. (1926). A mathematical theory of natural and artificial selection. Part III. *Proceedings of the Cambridge Philosophical Society* **23**, 363–372.

HALDANE, J. B. S. (1935). The rate of spontaneous mutation of a human gene. *Journal of Genetics* **31**, 317–326.

HALDANE, J. B. S. & JAYAKAR, S. D. (1964). Equilibrium under Natural Selection at a sex-linked locus. *Journal of Genetics* **59**, 29–36.

HEDRICK, P. W. (1980). Selection in finite populations: III. An experimental examination. *Genetics* **96**, 297–313.

JAMES, J. W. (1979). The time of detection of sex-linked recessives in small populations. *Genetical Research* **34**, 11–17.

JENNINGS, H. S. (1916). The numerical results of diverse systems of breeding. *Genetics* **1**, 53–89.

KERR, W. E. & KERR, L. S. (1952). Concealed variability in the *X*-chromosome of Drosphila Melanogaster. *The American Naturalist* **86**, 405–408.

KIMURA, M. (1964). Diffusion models in Population Genetics. *Journal of Applied Probability* **1**, 177–232.

KIMURA, M. & CROW, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

KIMURA, M. & OHTA, T. (1969*a*). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771.

KIMURA, M. & OHTA, T. (1969*b*). The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* **63**, 701–709.

LESTER, L. J. & SELANDER, R. K. (1979). Population Genetics of Haplodiploid insects. *Genetics* **92**, 1329–1345.

LI, C. C. (1955). *Population Genetics*. Chicago and London: University of Chicago Press.

MAYO, O. (1976). Neutral alleles at *X*-linked loci. *Human Heredity* **26**, 263–266.

MCKUSICK, V. A. (1975). *Mendelian Inheritance in Man, 4th ed.* Baltimore and London: Johns Hopkins, University Press.

NAGYLAKI, T. (1977). Selection in one- and two-locus systems. *Lecture Notes in Biomathematics 15* (ed. S. Levin) Berlin, Heidelberg and New York: Springer-Verlag.

NAGYLAKI, T. (1979). Selection in dioecious populations. *Annals of Human Genetics, London* **43**, 143–150.

NAGYLAKI, T. (1981). The inbreeding effective population number and the expected homozygosity for an *X*-linked locus. *Genetics* **47**, 731–737.

NEI, M., FUERST, P. A. & CHAKRABORTY, R. (1976). Testing the neutral mutation hypothesis by distribution of single locus heterozygosity. *Nature* **262**, 491–493.

OHNO, S. (1967). *Sex chromosomes and Sex-linked genes*. Berlin, Heidelberg and New York: Springer Verlag.

PAMILO, P. (1979). Genic variation at sex-linked loci: Quantification of regular selection models. *Hereditas* **91**, 129–133.

POWELL, J. R. (1975). Protein variation in natural populations of animals. *Evolutionary Biology* **8**, 79–119.

PRAKASH, S. (1973). Patterns of gene variation in central and marginal populations of Drosophila Robusta. *Genetics* **75**, 347–369.

PRAKASH, S. (1977*a*). Gene polymorphism in natural populations of Drosophila Persimilis. *Genetics* **85**, 513–530.

PRAKASH, S. (1977*b*). Further studies on gene polymorphism in the main-body and geographically isolated populations of Drosophila Pseudoobscura. *Genetics* **85**, 713–719.

PRAKASH, S., LEWONTIN, R. C. & HUBBY, J. L. (1969). A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of Drosophila Pseudoobscura. *Genetics* **61**, 841–858.

ROBERTSON, A. (1962). Selection for heterozygotes in small populations. *Genetics* **47**, 1291–1300.

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

STEWART F. M. (1976). Variability in the amount of heterozygosity maintained by neutral mutations. *Theoretical Population Biology* **9**, 188–201.

WARD, P. S. (1980). Genetic variation and population differentiation in the *Rhytidoponera Impressa* group, a species complex of Ponerine ants (Hymenoptera: Formicidae). *Evolution* **34**, 1060–1076.

WILTON, A. (1979). Studies of the nature of chromosomal heterosis in Drosophila Melanogaster. Doctorial Thesis, University of Sydney.

WRIGHT, S. (1933). Inbreeding and Homozygosis. *Proceedings of the National Academy of Sciences, U.S.A.* **19**, 411–420.

YANG, C-K., LEE, C-T., CHEE, C-L. (1966). A stochastic process for sex-linked gene frequency in random-mated small populations (Chinese with English abstract). *Acta Biochimica et Biophysica Sinica* **6**, 118–129.