

The effect of life-history and mode of inheritance on neutral genetic variability

BRIAN CHARLESWORTH*

Institute of Cell, Animal and Population Biology, Ashworth Laboratories, University of Edinburgh, King's Buildings, Edinburgh EH9 3JT, UK

(Received 5 June 2000 and in revised form 21 September 2000)

Summary

Formulae for the effective population sizes of autosomal, X-linked, Y-linked and maternally transmitted loci in age-structured populations are developed. The approximations used here predict both asymptotic rates of increase in probabilities of identity, and equilibrium levels of neutral nucleotide site diversity under the infinite-sites model. The applications of the results to the interpretation of data on DNA sequence variation in *Drosophila*, plant, and human populations are discussed. It is concluded that sex differences in demographic parameters such as adult mortality rates generally have small effects on the relative effective population sizes of loci with different modes of inheritance, whereas differences between the sexes in variance in reproductive success can have major effects, either increasing or reducing the effective population size for X-linked loci relative to autosomal or Y-linked loci. These effects need to be accounted for when trying to understand data on patterns of sequence variation for genes with different transmission modes.

1. Introduction

The advent of easily accessible methods for studying variation at the DNA level is producing a wealth of data on a variety of organisms. There is a growing interest in examining these results for interesting patterns, which may shed light on the evolutionary forces controlling levels of natural variability. In particular, comparisons of amounts of variability on autosomes and X chromosomes in *Drosophila* have been used to test hypotheses about factors influencing variation (Aquadro *et al.*, 1994; Begun, 1996; Moriyama & Powell, 1996; McVean & Charlesworth, 1999; Andolfatto, 2001; Begun & Whitley, 2000). There is also interest in using patterns of variation on evolving Y or neo-Y chromosomes to make inferences about the processes leading to the degeneration of Y chromosomes (Bachtrog & Charlesworth, 2000; Filatov *et al.*, 2000; Yi & Charlesworth, 2000). A clear understanding of the expectations for the levels of neutral variability on autosomes, X chromosomes and Y chromosomes, in the absence of any disturbing forces, is necessary for these purposes.

Under the standard infinite-sites model of molecular population genetics, the equilibrium level of neutral variability at a locus in a panmictic population is governed by the product of the mutation rate and effective population size (N_e) (Kimura, 1971). The relative levels of variability for genes with different modes of inheritance are thus dependent only on the relative values of N_e , if there are no sex differences in mutation rate. It is commonly assumed that the simplest discrete-generation formulae for N_e , which assume Poisson offspring number distributions (Wright, 1969), can be used for calculating expected levels of diversity. These formulae imply that an X chromosome has three-quarters the effective size of an autosome, and a Y chromosome gene only one-quarter (e.g. Aquadro *et al.*, 1994; Moriyama & Powell, 1996).

However, as has been pointed out many times (e.g. Nunney, 1993; Caballero, 1995; Charlesworth, 1996), there may well be sources of non-random variation in fertility that inflate the variance in reproductive success for one sex over the other. For example, an increased variance in male reproductive success due to sexual selection would inflate the ratio of the effective population size for X-linked genes to that for

* Tel: +44(0) 131 650 5750. Fax: +44(0) 131 650 6564. e-mail: brian.charlesworth@ed.ac.uk

autosomal genes, since X chromosome genes are transmitted twice as often through females as through males, the opposite to what is observed in some of the *Drosophila* examples mentioned above (Aquadro *et al.*, 1994; Begun, 1996; Moriyama & Powell, 1996; Begun & Whitley, 2000; Andolfatto, 2001). Additional factors, such as various forms of selection, have therefore been postulated to account for this pattern (Aquadro *et al.*, 1994; Begun, 1996; McVean & Charlesworth, 1999; Begun & Whitley, 2000).

In populations with age structure, such as humans and many species of *Drosophila*, there is an additional source of non-Poisson variance in net reproductive success, caused by the stochastic nature of survival during the reproductive phase and by differences in fertility among individuals of different ages (Felsenstein, 1971; Hill, 1972; Nunney, 1993). If there are sex differences in age-specific patterns of reproduction and survival, this could alter expectations for the effective sizes of different types of gene, e.g. lower male survival during adult life would reduce N_e for autosomal genes more than for X-linked genes. A well-developed body of theory has been worked out for calculating effective population size in age-structured populations (reviewed by Charlesworth, 1994, pp. 78–91).

This theory has, however, mostly been developed in terms of either the asymptotic rate of increase in probabilities of identity by descent of pairs of alleles, or the asymptotic rate of increase of variance in allele frequencies at a biallelic locus. No attention has been given to predictions of equilibrium levels of genetic diversity, and it is not self-evident that the above N_e values necessarily apply to this situation. One purpose of this paper is to develop such predictions, using the method originally used for the study of variability in a geographically structured population by Nagylaki (1980). As has been noted previously (Whitlock & Barton, 1997), there is a close analogy between geographically structured and age-structured populations. This can be usefully exploited for the present purpose; it also enables a unified approach to the derivation of expressions for N_e , as outlined in the Appendix. Related methods have also been used to predict rates of inbreeding in populations under selection (Wray & Thompson, 1990; Nomura, 1999; Woolliams *et al.*, 1999).

As will be seen, the standard formulae for N_e in fact provide excellent approximations for predicting expected equilibrium levels of genetic diversity under the infinite-sites model. Some simplifications to these formulae are proposed which facilitate comparisons of the effects of sex differences in demographic parameters on the N_e values for different modes of inheritance, and also incorporate the effects of other sources on non-random variation in reproductive success. The conclusions are applied to the interpret-

ation of levels of DNA variability in *Drosophila*, humans and plants.

2. Results

An age-structured population is a special class of structured population with gene flow among subclasses. While it is usually difficult to obtain explicit general results for the amount of genetic variability maintained in a structured population (Nagylaki, 1998; Wilkinson-Herbots, 1998), the case of a geographically structured population when gene flow is much stronger than mutation and genetic drift (the ‘strong-migration limit’) can be fully analysed to a satisfactory level of approximation (Nagylaki, 1980, 2000). It is straightforward to extend this approach to an age-structured population, provided that mating is random with respect to genotype, since the conditions for the validity of the strong migration limit are easily satisfied if the numbers of individuals in each age-class are large. A general derivation of the necessary results is presented in the Appendix. The population is assumed to be stationary in size and age-composition, and so large that stochastic fluctuations in numbers can be neglected.

These results show that an effective population size N_e can be defined for an age-structured population which determines both the asymptotic rate of increase in inbreeding coefficient per generation and the equilibrium level of nucleotide site diversity under the infinite-sites model. The use of this measure of N_e depends on appropriate definitions of generation time and mutation rate, which are determined by the genetic system in question, as described in Sections (i)–(iii) below. The genetic state of the population at time t is described by a vector $\mathbf{h}(t)$, made up of an ordered set of components, such that a given component $h_{ij}(t)$ is the probability of non-identity between two genes sampled from age–sex classes i and j at time t . Changes in \mathbf{h} are specified by the matrices \mathbf{A} and \mathbf{S} defined in the Appendix, which respectively describe the deterministic effects of the flow of genes among sex and age-classes, and the effect of stochastic sampling on variability (Nagylaki, 1980).

(i) Autosomal loci

Let n_{fx} and n_{mx} be the numbers of females and males in age-class x at time t , where x runs from 0 to d for females, and from 0 to d^* for males. Age-class 0 corresponds to newly formed zygotes, age-class 1 to individuals one time-unit older, etc. For internal consistency, the demographic state of the juvenile and reproductively active adult population at any one time is described by the column vector \mathbf{n} , whose components are $n_{f1}, n_{f2}, \dots, n_{fd}, n_{m1}, n_{m2}, \dots, n_{md^*}$. Let l_{sx} be the probability of survival from conception to age x for

individuals of sex s . Let m_{sx} be the expected number of offspring of sex s of a parent of sex s (for $x > d$, $m_{fx} = 0$; for $x > d^*$, $m_{mx} = 0$). The corresponding expectation of same-sex offspring produced at age x for a zygote of sex s is $k_{sx} = l_{sx}m_{sx}$.

Given these definitions, it is easy to obtain the elements of \mathbf{G} , a matrix that describes the deterministic flow of alleles among age- and sex-classes (Johnson, 1977), from which \mathbf{A} can be derived (equation (A 2)). The element g_{ij} is the probability that an allele now in class i came from class j in the previous time interval (where values of i between 1 and d refer to females of age $x = i$, and values between $d + 1$ and $d + d^*$ refer to males of age $x = i - d$). The following quantities are also needed to obtain an expression for N_e (Johnson, 1977).

Generation time, T , is defined as the mean age of the parents of newborn offspring. Assuming that the primary sex ratio is independent of parental age, we have

$$T = \frac{1}{2} \sum_{x=1}^d xk_{fx} + \frac{1}{2} \sum_{x=1}^{d^*} xk_{mx}. \tag{1}$$

The components of the left eigenvector, \mathbf{q} , of \mathbf{G} corresponding to its leading eigenvalue, normalized so that females and male aged 1 have values of one-half each, are given by

$$q_{fx} = \frac{1}{2} \sum_{y=x}^d k_{fy} \quad (1 \leq x \leq d), \tag{2a}$$

$$q_{mx} = \frac{1}{2} \sum_{y=x}^{d^*} k_{my} \quad (1 \leq x \leq d^*). \tag{2b}$$

Equation (1) can thus be rewritten as:

$$T = \sum_{x=1}^d q_{fx} + \sum_{x=1}^{d^*} q_{mx}. \tag{3}$$

If the components of \mathbf{q} are divided by T , we obtain a vector \mathbf{v} whose elements sum to 1, which is used in the calculations in the Appendix.

Using these expressions, a straightforward evaluation of the terms in \mathcal{S} , and their substitution into equation (A 7), yields the following well-known approximate expression for the reciprocal of the effective population of an autosomal locus (N_{eA}), neglecting terms of the order of the square of the reciprocal of the population size (Felsenstein, 1971; Johnson, 1977; Emigh & Pollak, 1979):

$$\frac{1}{N_{eA}} \approx \frac{1}{T} \left[\frac{1}{4n_{f1}} + \frac{1}{4n_{m1}} + \sum_{x=2}^d q_{fx}^2 \left(\frac{1}{n_{fx}} - \frac{1}{n_{fx-1}} \right) + \sum_{x=2}^{d^*} q_{mx}^2 \left(\frac{1}{n_{mx}} - \frac{1}{n_{mx-1}} \right) \right]. \tag{4}$$

This formula can be used as follows to determine the equilibrium level of diversity under the infinite-

sites model (Kimura, 1971). Let u be the probability that a gene of a new zygote contains a mutation (see equation (A 14)). Combining equations (A 12) and (A 13) with equation (A 14), we recover the standard discrete generation formula for equilibrium nucleotide site diversity under the infinite-sites model as

$$\pi = 4N_{eA}u + O(u). \tag{5}$$

(ii) *X-linked loci*

It is straightforward to evaluate the components of \mathbf{G} for the sex-linked case under the same assumptions as used in the autosomal case. If we normalize the components of its left leading eigenvector, \mathbf{q} , such that $q_{f1} = 2/3$ and $q_{m1} = 1/3$ (corresponding to the relative contributions of females and males to the X chromosome gene pool), we have

$$q_{fx} = \frac{2}{3} \sum_{y=x}^d k_{fy} \quad (1 \leq x \leq d), \tag{6a}$$

$$q_{mx} = \frac{1}{3} \sum_{y=x}^{d^*} k_{my} \quad (1 \leq x \leq d^*). \tag{6b}$$

The corresponding expression for generation time is

$$T^* = \sum_{x=1}^d q_{fx} + \sum_{x=1}^{d^*} q_{mx},$$

i.e.

$$T^* = \frac{2}{3} \sum_{x=1}^d xk_{fx} + \frac{1}{3} \sum_{x=1}^{d^*} xk_{mx}. \tag{7}$$

Applying these results to equation (A 7), we obtain an equation equivalent to that given by Pollak (1990) for the effective population size of an X-linked locus:

$$\frac{1}{N_{eX}} \approx \frac{1}{T^*} \left[\frac{4}{9n_{f1}} + \frac{2}{9n_{m1}} + \sum_{x=2}^d q_{fx}^2 \left(\frac{1}{n_{fx}} - \frac{1}{n_{fx-1}} \right) + 2 \sum_{x=2}^{d^*} q_{mx}^2 \left(\frac{1}{n_{mx}} - \frac{1}{n_{mx-1}} \right) \right]. \tag{8}$$

The mutation rate needed to obtain the expression for equilibrium diversity is obtained as follows. For a gene in a female zygote, the probability of a new mutation is the same as that given by equation (A 14). For a gene in a male zygote, it is

$$u^* = \sum_{x=1}^d u_{fx}k_{fx}. \tag{9}$$

Carrying out the same simplification as in the autosomal case, we find that

$$\pi = 4N_{eX} \left(\frac{2}{3}u + \frac{1}{3}u^* \right) + O(u). \tag{10}$$

(iii) *Y-linked and mitochondrial loci*

The cases of Y-linked and mitochondrial loci are equivalent to a haploid population, with the demo-

graphic parameters of males and females, respectively, so that the results of Felsenstein (1971) for the effective population size can be used. For Y-linked loci, we have

$$\frac{1}{N_{eY}} = \frac{2}{T^{**}} \left[\frac{1}{n_{m1}} + \sum_{x=2}^{a^*} q_{mx}^2 \left(\frac{1}{n_{mx}} - \frac{1}{n_{m(x-1)}} \right) \right], \quad (11)$$

where q_{mx} is now twice the quantity given by equation (2b), and T^{**} is the generation time for males:

$$T^{**} = \sum_{x=1}^{a^*} xk_{mx}.$$

The mutation rate for this case clearly involves only mutations derived from males, so that equation (5) applies with this modification.

For maternally transmitted organelle genomes, female demographic parameters replace the male ones in these formulae.

(iv) *Effects of non-random variation in fertility*

The formulae for N_e derived above assume that the only sources of non-random sampling of offspring from parents alive at a given time are differences in mean numbers of offspring among parents of different ages. This assumption can be relaxed, using the formulae given in the Appendix (equations (A 17) and (A 18)), which give additional terms that are to be added to the appropriate expressions for $1/N_e$. When there is no variation among parents in the primary sex ratio of their offspring, or their probabilities of survival to age 1, the additional terms in $1/N_e$ simplify as follows:

Autosomal loci:

$$\frac{1}{4T} \left(\frac{\Delta V_f}{n_{fT}} + \frac{\Delta V_m}{n_{mT}} \right) \quad (12a)$$

Sex-linked loci:

$$\frac{2}{9T^{**}} \left(\frac{2\Delta V_f}{n_{fT}} + \frac{\Delta V_m}{n_{mT}} \right), \quad (12b)$$

where ΔV_f is the difference between the total variance of offspring number per adult female and the component due to variance in mean number of offspring per adult female among females of different ages, scaled relative to the squared mean offspring number; ΔV_m is the corresponding expression for offspring per adult male; n_{sT} is the number of adult individuals of sex s present at a given time.

A similar calculation for the case of Y-linked genes yields the correction term

$$\frac{2\Delta V_m}{T^{**}n_{mT}}. \quad (13)$$

Organelle genomes can be treated similarly, substituting female for male parameters.

3. Applications

(i) *General considerations*

The main focus here will be the effect of demographic factors on the effective population size, extending the contributions of Nunney (1991, 1993). For purposes of numerical illustrations of different scenarios, it is convenient to approximate the summations in equations (4), (8) and (11) by integrals, but the frame of reference will be a discrete age-class model with a time-interval of one unit of time between successive age-classes. The reason for this is the technical problems that arise in dealing with the terms resulting from non-random variation in fertility when a continuous limit is taken, discussed in Section 3.iii below. In reality, of course, reproductive activity by females is usually a point process, corresponding to the production of litters or laying of clutches of eggs, so that a discrete-time model is the appropriate one to use.

If B is the number of births per unit time, the number of individuals of sex s of age x at time t is $n_{sx} = 0.5Bl_{sx}$, assuming a primary sex ratio of 0.5, which is reasonable for most cases of interest here. Substituting into the integral equivalent of equations (4), rearranging the terms of the resulting integral in the denominator, and using equation (12a), the effective population size for autosomal loci is given by

$$N_{eA} \approx \frac{BT}{\left(\frac{\Delta V_f}{2l_{fb}e_f} + \frac{\Delta V_m}{2l_{mb}e_m} \right) + 2 \int_0^\infty (q_{fx}m_{fx} + q_{mx}m_{mx})dx}, \quad (14)$$

where e_s is the life expectancy at maturity of an individual of sex s , l_{sb} is the probability that a zygote of sex s survives to maturity, and q_{sx} is defined by the integral equivalent of equations (2), respectively.

Similarly, the effective population size for X-linked loci is

$$N_{eX} \approx \frac{BT^{**}}{\frac{4}{9} \left\{ \frac{2\Delta V_f}{e_f l_{fb}} + \frac{\Delta V_m}{e_m l_{mb}} \right\} + \frac{8}{3} \int_0^\infty \{q_{fx}m_{fx} + q_{mx}m_{mx}\} dx}, \quad (15)$$

where q_{sx} is defined by the integral equivalent of equations (6).

The effective size for Y-linked loci is

$$N_{eY} \approx \frac{BT^{**}}{4 \left\{ \frac{\Delta V_m}{l_{mb}e_m} + 2 \int_0^\infty q_{mx}m_{mx} dx \right\}}, \quad (16)$$

where q_{mx} is twice the integral equivalent of q_{mx} in equation (2b).

(ii) *Effects of sex differences in demographic parameters*

This section will focus on the effects of purely demographic factors on N_e , ignoring the contributions from non-random sources of variation in offspring number. Analytical results on the effects on N_e of differences in male versus female life-history parameters can easily be obtained for the case of constant adult mortality and fertility. Age-independence of mortality rate under natural conditions is often observed, since external sources of mortality such as predation usually outweigh any intrinsic changes due to senescence (Charlesworth, 1994). Constant adult fertility is realistic for many natural populations of birds, small mammals and plants, but not for organisms like *Drosophila* and humans, where reproductive success reaches a peak early in adult life and then declines (Charlesworth, 1994). Nevertheless, it serves as a rough guide to more complex situations, and has previously been explored by Nunney (1991, 1993).

Let the ages at first reproduction for females and males be b_f and b_m . For adults, the death rate per unit time for an individual of sex s in a continuous time representation is denoted by μ_{sa} . Hence, for adults we have

$$l_{sx} = l_{sb} \exp -\mu_{sa}(x - b_s).$$

Equation (14) for autosomal loci then reduces to

$$N_{eA} \approx \frac{B(b_f + b_m + \mu_{fa}^{-1} + \mu_{ma}^{-1})}{(l_{fb}^{-1} + l_{mb}^{-1})}. \quad (17)$$

Similarly, equation (15) for X-linked loci becomes

$$N_{eX} \approx \frac{3B(2b_f + b_m + 2\mu_{fa}^{-1} + \mu_{ma}^{-1})}{4(2l_{fb}^{-1} + l_{mb}^{-1})}. \quad (18)$$

The ratio of effective sizes for X-linked and autosomal loci, in the absence of non-random variation in fertility, is thus

$$\frac{N_{eX}}{N_{eA}} \approx \frac{3(2b_f + b_m + 2\mu_{fa}^{-1} + \mu_{ma}^{-1})(l_{fb}^{-1} + l_{mb}^{-1})}{4(b_f + b_m + \mu_{fa}^{-1} + \mu_{ma}^{-1})(2l_{fb}^{-1} + l_{mb}^{-1})}. \quad (19)$$

As expected from the standard discrete-generation results, this ratio is equal to three-quarters if all the demographic parameters of the two sexes are equal (this is also true for cases where mortality and fertility rates vary with age). The effects of sex differences can be explored as follows. First, consider adult mortality. If adult mortality for males is lower than that for females, it is evident that the ratio $R_{XA} = (N_{eX}/N_{eA})$ is reduced, since less weight is given to male longevity in the expression for N_e for sex-linked than for autosomal loci. The converse is true if males have higher adult mortality. In *D. melanogaster*, there is evidence from laboratory studies that reproduction may impose a

higher survival cost on females than on males. For example, in the study of Roper *et al.* (1993), the adult life expectancy of females from the outbred flies in their table 1 was about 92% of that of males; it is possible that this effect could be even more marked in nature, although data are currently lacking. Another factor operating in the same direction is the longer development time commonly observed for males; the data of Roper *et al.* (1993) indicate that the development time of males was about 4% greater than that of females for F1 hybrids between lines selected for early reproduction, and about 3% greater for F1s between lines selected for late reproduction. But longer development time implies that survival to maturity may be lower for males than females; this will operate in the opposite direction, by reducing l_{mb} relative to l_{fb} .

l_{fb} and l_{mb} are related by the requirement for stationary population size, which implies that

$$\mu_{sa} = m_s l_{sb}. \quad (20)$$

If juvenile mortality rate per unit time is the same for males and females, l_{mb} can be determined by using l_{fb} from equation (20) and assuming a given value of b_f , thereby determining the juvenile mortality rate. Substitution of the resulting value for l_{mb} into the male version of equation (20) enables m_m to be determined, on the assumption that the fertility of matings is controlled by females. Taking m_f to be 10 female eggs per day as a rough average over a *Drosophila* female's reproductive span (cf. fig. 1 of Roper *et al.*, 1993), this implies that the probability of survival to maturity for a female is one-tenth the mortality rate per day.

Adult survival rates for *D. melanogaster* in nature are poorly known, but are likely to be low (Boesiger, 1968; Boulétreau, 1978), especially as population cage studies indicate short life expectations in a relatively benign environment (Frydenberg, 1962; Crow & Chung 1967). For illustrative purposes, assume a mortality rate per day for females of 0.09 and for males of 0.045, consistent with a large sex-difference of the type described above and with the survival data of Dobzhansky & Wright (1947) on a natural population of *D. pseudoobscura*, a species which is likely to be somewhat longer-lived than *D. melanogaster*. Taking a female development time of 9.00 days and a male value of 9.27, and assuming equal juvenile survival rates for males and females, equation (20) yields a value of $R_{XA} = 0.754$ for the ratio of X-linked to autosomal N_e values, i.e. a completely trivial difference from 0.75. If the adult mortality rate is half this value, but the ratio of male to female values is unchanged, $R_{XA} = 0.753$. If there were no differences between male and female probabilities of survival to maturity, R_{XA} values with the two different adult mortality rates become 0.742 and 0.740, still very close to 0.75, although now somewhat lower. It is, in fact,

Table 1. *Effects of variation in demographic parameters on relative N_e values of sex chromosomal and autosomal loci*

Row	μ_{ta}	μ_{ma}	b_f	b_m	R_{XA}	R_{YA}	R_{YX}
$\beta_f = 0.06, \beta_m = 0.06, \gamma_f = 0.905, \gamma_m = 0.905$							
1	0.05	0.05	9	9	0.750	0.250	0.333
2	0.05	0.025	9	9	0.728	0.273	0.375
3	0.10	0.05	9	9	0.719	0.283	0.394
4	0.20	0.10	9	9	0.710	0.294	0.414
5	0.20	0.10	9	9.27	0.726	0.274	0.378
6	0.20	0.05	9	9.27	0.696	0.308	0.442
7	0.025	0.05	9	9	0.773	0.228	0.295
8	0.05	0.10	9	9	0.782	0.220	0.281
9	0.10	0.20	9	9	0.792	0.212	0.268
10	0.10	0.20	9	9.27	0.811	0.200	0.246
$\beta_f = 0.06, \beta_m = 0.03, \gamma_f = 0.905, \gamma_m = 0.905$							
11	0.05	0.05	9	9	0.722	0.279	0.387
12	0.10	0.10	9	9	0.732	0.269	0.367
13	0.05	0.05	9	9.27	0.742	0.257	0.346
14	0.10	0.05	9	9	0.692	0.315	0.455
15	0.05	0.10	9	9.27	0.785	0.219	0.279

R_{XA} , R_{YA} and R_{YX} are the ratios of effective population sizes of X chromosome to autosome, Y chromosome to autosome, and Y chromosome to X chromosome, respectively. See text for explanation of other symbols.

very difficult to produce values of the ratio much less than 0.75 with this model, even with a much smaller male mortality rate and with no sex difference in probability of survival to maturity; e.g. with a male mortality rate that is one-half the female value, R_{XA} is 0.697 and 0.686 for the two female mortality rates considered above.

These results suggest that demographic factors of this kind are unlikely to have major effects on the relative effective population sizes of X-linked and autosomal loci in *Drosophila*. This model is somewhat unrealistic, in that it ignores the roughly triangular relation between reproductive output and age that is found in both humans and *Drosophila*. A simple model that describes this relation was proposed by McMillan *et al.* (1970) for egg production by *D. melanogaster* females, and adapted by Roff (1981) for the analysis of life-history evolution. Given that the age-specific profile of male reproductive success is roughly parallel to that of females in this species (Roper *et al.*, 1993), this model can also be used for male reproduction.

We can write

$$m_{sx} = \alpha_s [\exp(-\beta_s(x - b_s)) - \exp(-\gamma_s(x - b_s))], \quad (21)$$

where $\beta_s \ll \gamma_s$.

If the assumption of age-independent adult mortality is retained, integration of the product of this expression and the probability of survival to age yields the condition for stationarity:

$$l_{sb} = \frac{(\mu_{sa} + \beta_s)(\mu_{sa} + \gamma_s)}{\alpha_s(\gamma_s - \beta_s)}. \quad (22)$$

If the female parameters in equation (21) are treated as fixed, as well as the ages of maturity for males and females, this equation can be used to determine α_m by the same method as in the previous model, given values of b_s for the two sexes and values of β_m and γ_m .

Simple expressions for generation time and the integral involving $q_{sx}m_{sx}$ in the expressions for N_e are readily obtained. The contribution of sex s to generation time is

$$b_s + \frac{1}{(\gamma_s - \beta_s)} \left[\frac{(\mu_{sa} + \gamma_s)}{(\mu_{sa} + \beta_s)} - \frac{(\mu_{sa} + \beta_s)}{(\mu_{sa} + \gamma_s)} \right] \quad (23a)$$

and the integral of $q_{sx}m_{sx}$ is proportional to

$$\frac{\alpha_s}{(\gamma_s - \beta_s)} \left\{ \frac{(\mu_{sa} + \gamma_s)}{(\mu_{sa} + \beta_s)} - \frac{(\mu_{sa} + \beta_s)}{(\mu_{sa} + \gamma_s)} \right\}. \quad (23b)$$

These results can be substituted into equations (14)–(16) to obtain ratios of N_e values for X versus autosome, Y versus autosome, and Y versus X chromosome, under any given set of assumptions. Some numerical examples, based on the estimates of β and γ from the data of McMillan *et al.*, (1970) on female fecundity in an intercross between different strains of the Oregon-R stock of *D. melanogaster*, are shown in Table 1.

Rows 1–10 assume no sex difference in the shape of the relation between reproductive success and age, while rows 11–15 assume that males have a slower decline in reproductive success at advanced ages. The overall conclusion is similar to that above; even a large mortality differential in favour of males has only a small effect on R_{XA} (e.g. rows 2–4); the effect is

diminished by a difference in probability of survival of males and females to maturity, due to a difference in age at maturity (rows 5–6). The maximum effect is observed when there is an adult mortality differential in favour of females, as well as delayed maturity of males, both of which inflate R_{XA} , but even this is relatively small (row 10). The effects on the ratios of Y chromosome to autosome or X chromosome effective sizes are larger than those on R_{XA} if sex differences are sufficiently extreme, and go in the opposite direction to effects on R_{XA} . Rows 11–15 show that a late-life fertility advantage to males will, as expected, bring R_{XA} down and R_{YA} and R_{YX} up. This effect is counteracted by later reproductive maturity of males and/or higher adult mortality of males (rows 13 and 15).

A substantial late-life fertility advantage to males relative to females seems more likely for humans than for flies; demographic data indicate that human males continue to reproduce much later in life than females, as well as initiating reproduction later (e.g. Keyfitz, 1968, table 3.6). Numerical studies similar to those in Table 1 confirm the same general effect of a late-life fertility advantage to males, for parameter values that seem reasonable for human populations under natural conditions. The effect is, however, largely outweighed by the mortality differential against males and the later age at maturity of males, so that the net deviation from the predictions in line 1 of Table 1 is trivial. Overall, it is thus unlikely that demographic differences of these kinds will have major consequences for the relative effective population sizes under different modes of inheritance. The magnitudes of such effects are probably too small to be detectable from datasets on DNA sequence variation of the sizes that are currently available (see Section 4).

(iii) Effects of non-random variation in fertility

Differences in fertility variances between the two sexes, notably those resulting from sexual selection, are commonly invoked as potential causes of departures from Poisson expectation of the values of the ratios of effective population sizes (Crow & Morton, 1955; Nunney, 1993). In discrete-generation models, a large excess variance of male reproductive success over Poisson expectation can have large effects; in the limit, when this excess variance dominates other terms, we have $R_{XA} = 9/8$, $R_{YA} = 1/8$ and $R_{YX} = 1/9$ (cf. Caballero, 1995).

Predictions of the effects of sexual selection in the case of age-structured populations are complicated by the need to include the terms arising from purely demographic effects in the denominators of equations (14)–(16), which in general will tend to dilute the effects of sexual selection expressed in ΔV_m . Nunney (1991, 1993) has considered some aspects of these

effects in the context of sexual selection acting on a species with discrete breeding intervals, as in the case of annually reproducing temperate-zone iteroparous species. His results for the special cases he considers, which are expressed in terms of Hill's formula for N_e involving sex-specific variances of life-time reproductive success (Hill, 1972, 1979), can easily be derived from the general equations used here. They imply, that as the adult life-expectancy increases, the effect on N_e of a given intensity of sexual selection within a single breeding season diminishes.

This can be seen as follows, using the model used to obtain equations (22) and (23), and assuming $\gamma_s \gg \beta_s$, μ_a and $\Delta V_f = 0$. In the absence of sex differences in demographic parameters, the denominator of equation (14) can be approximated by

$$l_{ib}^{-1} \left(\frac{\Delta V_m \mu_{fa}}{2} + 2 \right) \quad (24)$$

so that the effect of a given level of sexual selection decreases with adult life expectancy (the reciprocal of μ_a).

The discrete-time model assumes, however, that offspring are produced and intermale competition for mates take place in successive and independent episodes. While this is reasonable for species with discrete breeding seasons, it is inappropriate for species such as *Drosophila* or humans, where there is no such sharp clustering of bouts of reproduction. Taking the continuous-time limit leads to the absurd result that the contributions from ΔV_m and ΔV_f disappear; this reflects the unrealistic nature of the assumption of independence between reproductive events that are close in time.

One extreme possibility is that some fraction of individuals fail to reproduce at any time in their lives; this could apply either to females who fail to enter breeding condition, or to males in a polygynous mating system who fail ever to obtain mates. Under these conditions, only individuals capable of breeding need be counted in the basic equations (14)–(16). This is tantamount to reducing l_{sb} for the sex in question by the appropriate factor, and re-expressing l_{sx} and m_{sx} in terms of values conditioned on individuals who are part of the breeding population. If this factor is small enough, there could be a substantial effect on the relative N_e values for different modes of inheritance, with N_{eX} being reduced relative to N_{eA} by a reduction in the proportion of breeding females, and vice versa by a reduction in the proportion of breeding males. The maximal effect of a reduction in the lifetime probability of a female breeding, p_f , to near-zero is to reduce R_{XA} to 9/16, and to increase R_{YA} and R_{YX} to $1/(8p_f)$ and $2/(9p_f)$, respectively. The effect of a lifetime reduction in the probability of a male breeding, p_m , to near-zero is to increase R_{XA} to 9/8, and to reduce R_{YA} and R_{YX} to 1/8 and 1/9,

respectively, the same effect as that of a large variance in male reproductive success.

In general, sexual competition between males will involve the passage over time of individuals in and out of groups of potential competitors for receptive females, which is difficult to model explicitly without recourse to simulations. In the case of *Drosophila*, a female continues to lay eggs fertilized by a given male for several days after mating, and under laboratory conditions is unreceptive to further courtship attempts (Ashburner, 1989, p. 142). The mean period, τ , over which a clutch derived from a given male is laid thus forms a natural time-interval for the model developed above. Genetic evidence for multiple matings in natural populations suggests, however, that this lack of receptivity is not absolute, and that clutches with mixed paternity are often produced (Harshman & Clark, 1998), which reduces the degree of non-Poisson variance in male reproductive success (Nunney, 1993). Ignoring this effect for the moment, under the lottery model of polygyny, which seems appropriate for *Drosophila* (Nunney, 1993), ΔV_m over the time-interval corresponding to the period over which eggs resulting from a given mating are laid would be expected to be approximately $(1 - \omega)/\omega$, where ω is the proportion of receptive females (Nunney, 1993). If most adult females are unreceptive, due to having recently mated, ω will be close to the fraction of the adult female population that has just reached sexual maturity over a given time-interval; this is equal to the reciprocal of the adult female life-expectancy in days if the population is stationary and receptive females mate within 24 h. On this model, e_m in equations (14)–(16) is measured in units of time-intervals of length τ , and so we obtain

$$\frac{\Delta V_m}{e_m} = (1 - e_r^{-1})\tau,$$

where e_s and τ are measured in days.

If n males contribute on average to the eggs laid by a single female over this period, this expression is reduced by a factor of approximately n (Nunney, 1993). Given the facts that eggs from a single mating of a *D. melanogaster* female continue to be produced over about 7 days (Ashburner 1989, p. 143) and that broods of wild-caught females seem to show evidence of insemination by roughly two males on average (Harshman & Clark, 1998), the net estimate of $\Delta V_m/e_m$ is therefore around 3.5 for this species. This may well be a considerable overestimate, since the production by a female of eggs fathered by a single male tends to decrease over a period of a few days (Ashburner, 1989, p. 142).

The expression in equation (24) then reduces approximately to $3.75/l_{fb}$, if the term in e_r is ignored. Assuming no other life-history differences between the sexes, similar calculations for the other terms in

equations (14)–(16), using equation (24), give $R_{XA} = 0.89$, $R_{YA} = 0.17$ and $R_{YX} = 0.19$, i.e. there is a substantial departure from the Poisson expectations in the direction of more X-linked variability than expected and less Y-linked variability.

4. Discussion

The results presented above bring out the fact that the relative values of the effective population sizes experienced by genes with different modes of inheritance are influenced by many different demographic variables, which in practice are very difficult to estimate even for laboratory populations. The analyses in Section 3.ii suggest, however, that rather extreme differences in survival and age-specific fertility patterns between males and females would be required to produce major departures of R_{XA} , R_{YA} and R_{YX} from their Poisson expectations of 0.75, 0.25 and 0.33.

(i) *Drosophila* populations

In the case of *D. melanogaster*, there is little reason to expect large effects of sex differences in demography. The major disturbing factor for this species would seem to be non-random variation in fertility. Considering male effects only, more or less plausible guesses about the effects of lottery model competition for mates (Nunney, 1993) suggest that the equilibrium level of variability on the X chromosome could be nearly 90% of that for the autosomes, and variability on the Y only 20% of the autosomal value (see Section 3.iii). A factor operating in the opposite direction is non-random variation in female breeding success; the most extreme effect would occur if some fraction of females simply failed to attain breeding condition. There is abundant evidence that female flies can control the development of their oocytes in response to their nutritional status (Soller *et al.*, 1999), so that it is quite likely that inseminated females will refrain from breeding in nature if they are in poor condition.

Given the 1:1 primary sex ratio in this species, this effect could greatly reduce the effective size for the X relative to the autosomes and Y chromosome. Observations on females caught in nature in southern France indicated that, despite the fact that they are mostly inseminated if they are over 24 h old, between 50% and 70% of wild-caught females lacked vitellogenic activity, although they can breed successfully when established in the laboratory (Boulétreau, 1978). If these figures correspond to life-time probabilities of not breeding, $(1 - p_f)$, the argument presented in Section 3.iii indicates that the terms in q_{fx} in the denominators of equations (14)–(16) should be increased by $1/p_f$. Taking the lower value of $p_f = 0.3$, in the absence of male fertility variation and other life-history difference, we would then have $R_{XA} = 0.64$

and $R_{YA} = 0.54$. With sexual selection of the intensity just discussed, we have $R_{XA} = 0.73$ and $R_{YA} = 0.28$.

An analysis of published data on DNA sequence diversity in *D. melanogaster* and *D. simulans* shows that African populations have levels of X chromosomal variability that are similar to or even larger than those for the autosomes, whereas the non-African populations have a mean X chromosomal nucleotide site diversity level that is approximately 87% of the mean autosomal level in *D. melanogaster* and 76% of that in *D. simulans* (Andolfatto, 2001). A similar pattern has been observed with microsatellite data (C. Schlötterer, personal communication). The standard errors of these estimates are, however, very high, but there does seem to be a significant reduction in variability on the X chromosome relative to the autosomes in non-African populations of *D. melanogaster* (Andolfatto, 2001). Begun & Whitley (2000) also found a significant reduction in X-linked relative to autosomal variability in *D. simulans*, using data from various, predominantly non-African, populations.

These results are consistent with the possibility that adult females are generally in good breeding condition in African populations in these species, and that there is strong sexual selection which reduces the effective size for the autosomes much more than for the X, whereas the poor breeding success of females in non-African populations reduces the effective size of the X more than that of the autosomes. Further ecological studies, especially of African populations, are needed to test this possibility. These calculations assume no differences in mutation rates between the sexes, which obviously can also influence relative nucleotide diversities. There is, however, no indication of any such difference in *D. melanogaster* and its relatives (Bauer & Aquadro, 1997).

Before appealing to selectionist interpretations of these patterns, it is thus important to exclude these purely demographic effects, which must, however, take into account the fact that the time required to approach a new equilibrium level of variability after a reduction in effective population size is of the order of N_e . If the effective sizes of *D. melanogaster* and *D. simulans* are around 1 million, as suggested by the nucleotide site diversity data, equilibration of the diversity values of X chromosomal loci after movement of these species out of Africa would require something of the order of 50000 years, assuming 20 generations per year. This is substantially longer than current estimates of the timing of their spread into Europe and Asia (around 10000 years ago: David & Capy, 1988), but the evidence for this date is weak. It is conceivable that fly populations accompanied modern humans as they moved out of Africa. Alternatively, a period of greatly reduced population size following movement out of Africa, followed by

subsequent expansion, could have speeded up equilibration of relative levels of diversity on X and autosome.

(ii) Human populations

The pattern of relative diversity values for X, Y and autosomes in humans is confusing at present, with very different results emerging from different studies whose interpretation is obscured by the fact that different studies use different sampling strategies, and different amounts of coding and non-coding sequences (Charlesworth & Charlesworth, 2000). In addition, if there is a difference in mutation rates between males and females (Miyata *et al.*, 1987; Li, 1997), or between the X chromosome and autosomes or the Y chromosome (McVean & Hurst, 1997), the ratio of diversities will differ from the R values. For example, if the male mutation rate is α times the female mutation rate, the expected ratio of X-linked to autosomal diversities is $2(2+\alpha)R_{XA}/3(1+\alpha)$. The value of α is still controversial, with estimates ranging between 1.7 and 5 (Anagnostopoulos *et al.*, 1999; Bohossian *et al.*, 2000; Nachman & Crowell, 2000). Taking $\alpha = 3$, which is in the mid-range of estimated values, the expected ratio of X-linked to autosomal diversities is $0.833R_{XA}$.

The mean diversity for 11 X-linked loci (including both coding and non-coding regions, in variable proportions) is 6.5×10^{-4} (from table 1 of Przeworski *et al.*, 2000), whereas the mean for 16 autosomal loci with nearly equal amounts of coding and non-coding sequences is 3.9×10^{-4} (Shen *et al.*, 2000). Taking these at face value, the estimate of R_{XA} is 2.0, much larger than expected even with intense sexual selection on males. In contrast, a study of several regions of the human Xq22 region from a world-wide sample of 24 men, based on a chemical mutation-detection method, gave a diversity value of 1.7×10^{-4} (Anagnostopoulos *et al.*, 1999). Using the same autosomal value as before, this gives an estimate of R_{XA} of 0.49, suggesting an excess of female variance in reproductive success. Studies that use comparable sampling strategies for X chromosome and autosomes are needed to resolve this major discrepancy.

For the Y chromosome, the corresponding expression for the expected ratio of Y-linked to autosomal diversities is $R_{YA}(1+\alpha)/2\alpha$. An extensive survey of world-wide Y chromosome variation in humans suggests that there is approximately 5 times less variation at non-coding sites at four Y-chromosomal loci than for comparable autosomal samples (Shen *et al.*, 2000)—slightly but probably not significantly greater than the 4-fold difference expected with a Poisson variance of reproductive success and no difference in mutation rate, but considerably larger than the value of 1.5 expected with $\alpha = 3$ and

implying an R_{YA} value of 0.133. This is consistent with an appreciable effect of sexual selection on the effective size of human populations.

(iii) Plant populations

A recent study of variability at a locus on the sex chromosomes of the dioecious perennial flowering plant *Silene latifolia* showed that there is 20 times as much variation among X-linked copies of the gene as among Y-linked copies (Filatov *et al.*, 2000). This is much greater than the maximum effect of 9 fold which occurs when a large fraction of males fail to breed throughout the whole of their life. Studies of the flowering phenology of this species show no evidence for such an extreme effect (Delph, 1999), so it seems unlikely that sexual selection could have such a large effect. Since adults are known to live for several years, and to flower each season, the effect of male competition occurring within flowering seasons on the relative effective sizes of different inheritance modes is likely to be counteracted by the purely demographic terms in the denominators of equations (14)–(16) (see Section 3.iii above), so that it seems extremely unlikely that the reduction in Y chromosomal variability could be caused by this factor.

There is evidence, however, that males suffer higher adult mortality, an adult sex ratios are often strongly male biased (Mulcahy, 1967; Lovett-Doust *et al.*, 1987). A simple calculation shows that, in a discrete age-class model, the relative numbers of males and females in a stationary population is equal to the inverse of the ratio of their probabilities of death per time-interval, if the primary-sex ratio is 1:1. An adult sex ratio of 1:2, which is close to the most extreme values reported, therefore indicates that males have a death rate that is approximately twice that of females. From the calculations presented in Table 1, this seems unlikely to reduce R_{YX} to as small a value as the ratio of observed diversities.

(iv) General remarks

The method of derivation used here employs gene identities and diversities. Orive (1993) applied a coalescent approach to age- and stage-structured populations, but did not obtain simple expressions even for the mean coalescent time for a pair of sampled alleles. Given the proportionality between the expected pairwise diversity measure under the infinite-sites model and the mean coalescent time for a pair of alleles (Hudson, 1990), the results derived in Section 2 must apply to the mean coalescent time, i.e. $2N_e$ is the mean time to coalescence of a pair of alleles in the age-structured case. Use of the properties of strong-migration models of structured populations, in which the time-scale over which migration between

classes occurs is shorter than for all other evolutionary processes (Nagylaki, 1980, 2000; Notohara, 1993; Nordborg, 1997; Wilkinson-Herbots, 1998), implies that the time to coalescence of a pair of alleles is independent of the age-class from which they originate, to the level of approximation assumed. In turn, this means that all the useful properties of the exponential distribution of coalescence times (Hudson, 1990) can be applied.

Appendix

(i) Properties of the general recursion relation

Following Felsenstein (1971) and Johnson (1977), we can treat the changes in probabilities of identity between genes sampled from different sex and age classes as follows. The state of the population at time t is described by the vector of probabilities of non-identity, $\mathbf{h}(t)$, defined in the text. In the absence of finite population size effects, we have the general recursion relation

$$\mathbf{h}(t) = \mathbf{A}\mathbf{h}(t-1). \quad (\text{A } 1)$$

The element $a_{ij,kl}$ of \mathbf{A} that corresponds to the subscript pairs ij and kl is the probability that a gene sampled from class i is derived from class k in the previous time-interval, and a gene from class j is derived from class l .

In an infinitely large population, two randomly sampled genes from a given age-sex class come from two separate individuals with probability 1, and hence represent two independent draws from this class. If g_{ik} is the probability that a gene sampled from class i is derived from class k in the previous time-interval, we thus have

$$a_{ij,kl} = g_{ik}g_{jl}. \quad (\text{A } 2)$$

In this case, \mathbf{A} is identical in structure to the matrix \mathbf{K} used by Nagylaki (1980) in dealing with migration between demes, and $\mathbf{G} = \{g_{ik}\}$ corresponds to his migration matrix \mathbf{M} . Both \mathbf{G} and \mathbf{A} are stochastic matrices, with leading eigenvalues of 1. Of interest in what follows are the components of the left eigenvector \mathbf{v} of \mathbf{G} that corresponds to the leading eigenvalue (Emigh & Pollak, 1979; Nagylaki, 1980). If these are normalized to sum to 1, the i th component, v_i , corresponds to the proportional contribution of the i th class to the ancestry of future generations, i.e. it is a measure of reproductive value (Whitlock & Barton, 1997).

In addition to \mathbf{A} , we need to consider the effects of random sampling on \mathbf{h} . This can be represented by a matrix of stochastic terms, \mathbf{S} , such that

$$\mathbf{h}(t) = (\mathbf{A} + \mathbf{S})\mathbf{h}(t-1). \quad (\text{A } 3)$$

If the population size of each class is of order n , the elements of \mathbf{S} are of order $1/n$ (for details of the

relevant stochastic terms see Felsenstein, 1971, and Johnson, 1977). If these are sufficiently small in relation to the non-zero components of \mathbf{G} (the ‘strong-migration limit’ assumption: Nagylaki, 1980), a perturbation argument implies that the right eigenvector ($\boldsymbol{\mu}_0$) of $\mathbf{A} + \mathbf{S}$ that corresponds to its leading eigenvalue has elements which differ from each other by terms of order $1/n$, since the corresponding elements of the eigenvector of \mathbf{A} are all equal (Emigh & Pollak, 1979; Nagylaki, 1980, eq. 10). Without loss of generality, we can thus write the components of $\boldsymbol{\mu}_0$ as $1 + O(1/n)$.

(ii) *The inbreeding effective population size*

Following Felsenstein (1971) and Nagylaki (1980), we can define a weighted mean of the $h_{ij}(t)$ as

$$\bar{h}(t) = \sum_{ij} v_i v_j h_{ij}(t). \tag{A 4}$$

This weighted mean is invariant under the transformation \mathbf{A} (Nagylaki, 1980).

We can write the leading eigenvalue of $\mathbf{A} + \mathbf{S}$ as

$$\lambda_0 = 1 - \frac{1}{2TN_e}, \tag{A 5}$$

where T is the generation time. We have

$$\lambda_0 \boldsymbol{\mu}_0 = (\mathbf{A} + \mathbf{S})\boldsymbol{\mu}_0. \tag{A 6}$$

Substituting from (A 4) and (A 5) into (A 6), and using the above result on the components of $\boldsymbol{\mu}_0$, we obtain

$$\frac{1}{2TN_e} = - \sum_{ij,kl} v_i v_j s_{ij,kl} + O\left(\frac{1}{n^2}\right). \tag{A 7}$$

This corresponds to equations (12) of Nagylaki (1980).

The effective population size in this sense gives the asymptotic rate at which the probabilities of non-identity change, and thus corresponds to the ‘inbreeding effective size’ (Whitlock & Barton, 1997).

(iii) *The mutation effective population size*

It is not obvious that the same quantity can be used to determine the equilibrium level of variability in the population (the ‘mutation effective size’ of Whitlock & Barton, 1997), although this is known to be the case for the infinite-alleles and infinite-sites models in the strong migration limit with population subdivision (Nagylaki, 1980).

In the present case, Nagylaki’s argument can be used to obtain the equilibrium nucleotide site diversity under the infinite-sites model, by adding a mutation vector \mathbf{u} to equation (A 3). Mutations are only of evolutionary relevance when transmitted to gametes, so that u_{ij} is defined as zero for all classes for which neither i nor j equals 1 or $d + 1$. If these probabilities are of order u , the infinite-sites assumption implies that u multiplied by the mutation effective population

size is sufficiently low that the equilibrium probability of non-identity of a pair of alleles is $\ll 1$; the mutational contribution to the change in \mathbf{h} can then be treated as independent of the state of the population, to order u .

We can define the vector \mathbf{h}^0 as the limit of the equilibrium value of \mathbf{h} as u tends to zero, keeping nu fixed. Following Nagylaki (1980), we can write

$$\mathbf{h} = \mathbf{h}^0 + u\mathbf{h}^1 + O(u^2) \tag{A 8}$$

and we find that \mathbf{h}^0 is an eigenvector of \mathbf{A} with eigenvalue 1. From the above results, this implies that the elements of \mathbf{h}^0 are all equal to a constant, h^0 say. Substituting into equation (A 8), we obtain

$$uh_{ij}^1 = h^0 \sum_{kl} s_{ij,kl} + u \sum_{kl} a_{ij,kl} h_{kl}^1 + u_{ij}. \tag{A 9}$$

Weighting the right-hand side by $v_i v_j$ and summing over all ij , we obtain

$$-h^0 \sum_{ij,kl} v_i v_j s_{ij,kl} = \sum_{ij} v_i v_j u_{ij}. \tag{A 10}$$

We can define a weighted measure of the rate per time-interval with which mutations arise that distinguish a pair of alleles (which is twice the corresponding mutation rate per haploid genome) as

$$2\tilde{u} = \sum_{ij} v_i v_j u_{ij}. \tag{A 11}$$

It follows from (A 7) and (A 10) that the equilibrium nucleotide site diversity is given by

$$\pi = 4N_e T\tilde{u} + O(u). \tag{A 12}$$

For an autosomal locus, it follows from the remarks before equation (A 8) that we have $u_{1j} = u_{d+1j} = u$ ($j \neq 1, d + 1$); $u_{11} = u_{1,d+1} = u_{d+1,1} = u_{d+1,d+1} = 2u$, where u is the probability that a new zygote contains a mutation. All other u_{ij} are zero. Hence

$$\sum_{ij} v_i v_j u_{ij} = 2u(v_1 + v_{d+1}) \sum_{j=1}^{d+d^*} v_j = \frac{2u}{T}. \tag{A 13}$$

From equations (2.29) of Charlesworth (1994), for an autosomal locus, we also have

$$u = \frac{1}{2} \left[\sum_{x=1}^d u_{fx} k_{fx} + \sum_{x=1}^{d^*} u_{mx} k_{mx} \right], \tag{A 14}$$

where u_{sx} is the probability that a gamete originating from an individual of age x and sex s contains a new mutation.

(iv) *Non-random variation in fertility*

The expression for \mathbf{S} used to derive N_e assumed random sampling of genes from parents of a given age–sex class (Johnson, 1977; Emigh & Pollak, 1979), implying that the probability of non-identity of a pair of genes in age-class 1 that originated from parents belonging to the same age–sex class k at time t is equal

to $h_{kk}(t)$. With variation in fertility additional to that corresponding to differences in mean fertility among age-classes, this is no longer true, since there is no longer a random chance that the two genes are replicates of the same gene from class k . Methods for accounting for such variation in the context of the inbreeding effective population size with discrete generations have been developed (Caballero, 1995; Nagylaki, 1995), and can be adapted as follows for the age-structured case.

Consider first the autosomal case. Only pairs of genes sampled from age-class 1 need be considered, since reproduction will not affect identity probabilities for other pairs. For two genes sampled from the same female from age-class 1, there is no change to the previous results, since they are either replicates of the same gene, or are derived from a male and a female. If they come from different females, and sampling of offspring is random except for differences in fertility across age-classes, the probability that two genes are derived from a replicate of the same gene (neglecting higher-order terms in $1/n$) is approximated by

$$\sum_{x=b}^d g_{1x}^2 \frac{1}{2n_{1x}} = \frac{1}{8} \sum_{x=b}^d \frac{n_{1x}^2 f_{1x}^2}{n_{1x}(n_{1T} \bar{f}_{1x})^2},$$

where f_{1x} is the expected number of daughters contributed to age-class 1 by a female aged x , \bar{f}_{1x} is the mean of this across all age-groups, n_{1T} is the total number of females of reproductive age, and b is the age of reproductive maturity for females.

This simplifies to

$$\sum_{x=b}^d g_{1x}^2 \frac{1}{2n_{1x}} = \frac{(1 + V_{ffA}/\bar{f}_{1x}^2)}{8n_{1T}} \tag{A 15a}$$

where V_{ffA} is the variance in mean numbers of daughters per female across reproductive ages.

This term is implicitly included in the calculations leading to equation (4), since it represents the chance that two random genes from this age–sex class are identical as a result of replication of the parental gene (this can be verified by direct calculation). More generally, however, expected numbers of daughters can be assigned individually to each female present at a given time, such that the probability that two offspring genes are replicates of the same gene of female i is proportional to the square of this female’s fertility. This yields the following approximate expression for the probability of identity of two genes from females of age-class 1:

$$\frac{(1 + V_{ffT}/\bar{f}_{1x}^2)}{8n_{1T}}, \tag{A 15b}$$

where V_{ffT} is the total variance in numbers of daughters per female among all females of reproductive age.

The difference between (A 15b) and (A 15a) represents the relevant contribution to the complement

of the sum of the $s_{11,kt}$ terms corresponding to female parents, and can conveniently be written as:

$$\frac{\Delta V_{ff}}{8n_{1T}}, \tag{A 16}$$

where ΔV_{ff} is the difference between the total variance of daughters per female and the variance between age-classes in mean numbers of daughters per female, scaled relative to the overall mean number of daughters per female.

A similar calculation for contributions from male parents to females aged 1 can be made, substituting the number of males of reproductive age, n_{mT} , for n_{1T} and writing ΔV_{mf} for the corresponding scaled difference in variance in numbers of daughters per male. We also have equivalent terms arising from males, of age-class 1, where now fertility is measured as sons per female and sons per male, inserting ΔV_{fm} and ΔV_{mm} into (A 16) and its equivalent for male progeny, respectively.

Finally, terms from genes sampled from a male and a female of age-class 1 must be considered. These involve the covariance between the number of male and female offspring contributed to age-class 1 by parents of a given sex. If these covariances are scaled by the product of the mean numbers of male and female offspring per parent, we obtain the equivalents of equations (A 15a) and (A 15b) as

$$\frac{\Delta C_{ffm}}{4n_{1T}} \tag{A 15c}$$

and

$$\frac{\Delta C_{mfm}}{4n_{mT}} \tag{A 15d}$$

where ΔC_{ffm} is the difference between the total scaled covariance in female and male offspring number for female parents, and the scaled covariance for mean numbers of male and female offspring per female of a given age class; ΔC_{mfm} is the corresponding expression for male parents.

Weighting these terms by the appropriate components of v , and substituting into equation (A 7), we obtain the following contribution to the expression for the reciprocal of N_e :

$$\frac{1}{16T} \left[\frac{(\Delta V_{ff} + \Delta V_{fm} + 2\Delta C_{ffm})}{n_{1T}} + \frac{(\Delta V_{mf} + \Delta V_{mm} + 2\Delta C_{mfm})}{n_{mT}} \right]. \tag{A 17}$$

The same general procedure can be followed for the other genetic systems studied here. For the case of sex-linkage, we get

$$\frac{1}{9T^*} \left[\frac{(\Delta V_{ff} + \Delta V_{fm} + 2C_{ffm})}{n_{1T}} + \frac{2\Delta V_{mf}}{n_{mT}} \right]. \tag{A 18}$$

These are equivalent to the expressions in Pollak (1980, 1990) and Caballero (1995).

I thank P. Andolfatto, D. J. Begun, D. Charlesworth, W. G. Hill, T. Nagylaki, C. Schlötterer and two reviewers for their comments on this paper. This work was supported by grants from the Royal Society and the EPSRC/BBSRC.

References

- Anagnostopoulos, T., Green, P. M., Rowley, G., Lewis, C. M. & Gianelli, F. (1999). DNA variation in a 5-Mb region of the X chromosome and estimates of sex-specific/type-specific mutation rates. *American Journal of Human Genetics* **64**, 508–517.
- Andolfatto, P. (2001). Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *D. simulans*. *Molecular Biology and Evolution* **18**, 279–290.
- Aquadro, C. F., Begun, D. J. & Kindahl, E. C. (1994). Selection, recombination, and DNA polymorphism in *Drosophila*. In *Non-Neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 46–56. London: Chapman and Hall.
- Ashburner, M. (1989). *Drosophila: A Laboratory Handbook*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Bachtrog, D. & Charlesworth, B. (2000). Reduced variability on the neo-Y chromosome of *Drosophila miranda*. *Current Biology* **10**, 1025–1031.
- Bauer, V. L. & Aquadro, C. F. (1997). Rates of DNA sequence evolution are not sex biased in *Drosophila melanogaster* and *D. simulans*. *Molecular Biology and Evolution* **14**, 1252–1257.
- Begun, D. J. (1996). Population genetics of silent and replacement variation in *Drosophila simulans* and *D. melanogaster*. X/autosome differences? *Molecular Biology and Evolution* **13**, 1405–1405.
- Begun, D. J. & Whitley, P. (2000). Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. *Proceedings of the National Academy of Sciences of the USA* **97**, 5960–5965.
- Boesiger, E. (1968). Estimation global de l'âge de femelles de *Drosophila melanogaster* capturées dans des populations naturelles. *Comptes Rendues de la Société de Biologie* **162**, 358–361.
- Bohossian, H. B., Skaletsky, H. & Page, D. C. (2000). Unexpectedly similar rates of nucleotide substitution found in male and female hominids. *Nature* **406**, 622–625.
- Boulétreau, J. (1978). Ovarian activity and reproductive potential in a natural population of *Drosophila melanogaster*. *Oecologia* **35**, 319–342.
- Caballero, A. (1995). On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* **139**, 1007–1011.
- Charlesworth, B. (1994). *Evolution in Age-structured Populations*, 2nd edn. Cambridge: Cambridge University Press.
- Charlesworth, B. (1996). Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genetical Research* **68**, 131–150.
- Charlesworth, B. & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London, Series B* **355**, 1563–1572.
- Charlesworth, D. & Charlesworth, B. (1998). Sequence variation: looking for effects of genetic linkage. *Current Biology* **8**, R658–661.
- Crow, J. F. & Chung, Y. J. (1967). Measurement of effective generation length in *Drosophila* population cages. *Genetics* **57**, 951–955.
- Crow, J. F. & Morton, N. E. (1955). Measurement of gene-frequency drift in small populations. *Evolution* **9**, 202–214.
- David, J. R. & Capy, P. (1988). Genetic variation of *Drosophila melanogaster* natural populations. *Trends in Genetics* **4**, 106–111.
- Delph, L. F. (1999). Sexual dimorphism and life history. In *Gender and Sexual Dimorphism in Flowering Plants* (ed. M. A. Geber, T. E. Dawson & L. F. Delph), pp. 149–174. Berlin: Springer.
- Dobzhansky, T. & Wright, S. (1947). Genetics of natural populations. XV. Rate of diffusion of a mutant gene through a population of *Drosophila pseudoobscura*. *Genetics* **32**, 303–339.
- Emigh, T. H. & Pollak, E. (1979). Fixation probabilities and effective population numbers in diploid populations with overlapping generations. *Theoretical Population Biology* **15**, 86–107.
- Felsenstein, J. (1971). Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* **68**, 581–597.
- Filatov, D. A., Monéger, F., Negrutiu, I. & Charlesworth, D. (2000). Evolution of a plant Y-Chromosome: variability in a Y-linked gene of *Silene latifolia*. *Nature* **404**, 388–390.
- Frydenberg, O. (1962). Estimation of some genetical and vital statistics parameters of Bennett populations of *Drosophila melanogaster*. *Hereditas* **48**, 83–103.
- Harshman, L. G. & Clark, A. G. (1998). Inference of sperm competition from broods of field-caught *Drosophila*. *Evolution* **52**, 1334–1341.
- Hill, W. G. (1972). Effective size of populations with overlapping generations. *Theoretical Population Biology* **3**, 278–289.
- Hill, W. G. (1979). A note on effective population size with overlapping generations. *Genetics* **92**, 317–322.
- Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford Surveys of Evolutionary Biology* **7**, 1–45.
- Johnson, D. L. (1977). Inbreeding in populations with overlapping generations. *Genetics* **87**, 581–591.
- Keyfitz, N. (1968). *Introduction to the Mathematics of Population*. Reading, MA: Addison-Wesley.
- Kimura, M. (1971). Theoretical foundations of population genetics at the molecular level. *Theoretical Population Biology* **2**, 174–208.
- Li, W.-H. (1997). *Molecular Evolution*. Sunderland, MA: Sinauer.
- Lovett-Doust, J., O'Brien, G. & Lovett-Doust, L. (1987). Effect of density on secondary sex characteristics and sex ratio in *Silene alba* (Caryophyllaceae). *American Journal of Botany* **74**, 40–46.
- McMillan, I., Fitz-Earle, M. & Robson, D. S. (1970). Quantitative genetics of fertility. I. Lifetime egg production of *Drosophila melanogaster*: theoretical. *Genetics* **65**, 344–353.
- McVean, G. A. T. & Charlesworth, B. (1999). A population genetic model for the evolution of synonymous codon usage: patterns and predictions. *Genetical Research* **74**, 145–158.
- McVean, G. A. T. & Hurst, L. D. (1997). Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. *Nature* **386**, 388–392.
- Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. (1987). Male-driven molecular evolution: a

- model and nucleotide sequence analysis. *Cold Spring Harbor Symposium on Quantitative Biology* **52**, 863–867.
- Moriyama, E. N. & Powell, J. R. (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution* **13**, 261–277.
- Mulcahy, D. L. (1967). Optimal sex ratios in *Silene alba*. *Heredity* **22**, 411–423.
- Nachman, M. W. & Crowell, S. L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304.
- Nagylaki, T. (1980). The strong-migration limit in geographically structured populations. *Journal of Mathematical Biology* **9**, 101–114.
- Nagylaki, T. (1995). The inbreeding effective population number in dioecious populations. *Genetics* **139**, 473–485.
- Nagylaki, T. (1998). The expected number of heterozygous sites in a subdivided population. *Genetics* **149**, 1599–1604.
- Nagylaki, T. (2000). Geographical invariance and the strong-migration limit in subdivided populations. *Journal of Mathematical Biology* **41**, 123–142.
- Nomura, T. (1999). On the methods for predicting the effective size of populations under selection. *Heredity* **83**, 485–489.
- Nordborg, M. (1997). Structured coalescent processes on different time scales. *Genetics* **146**, 1501–1514.
- Notohara, M. (1993). The strong-migration limit for the genealogical process in geographically structured populations. *Journal of Mathematical Biology* **31**, 115–122.
- Nunney, L. (1991). The influence of age structure and fecundity on effective population size. *Proceedings of the Royal Society of London, Series B* **246**, 71–76.
- Nunney, L. (1993). The influence of mating system and overlapping generations on effective population size. *Evolution* **47**, 1329–1341.
- Orive, M. (1993). Effective population size in organisms with complex life histories. *Theoretical Population Biology* **44**, 316–340.
- Pollak, E. (1980). Effective population numbers and mean times to extinction in dioecious populations with overlapping generations. *Mathematical Biosciences* **52**, 1–25.
- Pollak, E. (1990). The effective population size of an age-structured population with a sex-linked locus. *Mathematical Biosciences* **101**, 121–130.
- Przeworski, M. A., Hudson, R. R. & Di Rienzo, A. (2000). Adjusting the focus on human variation. *Trends in Genetics* **16**, 296–302.
- Roff, D. A. (1981). On being the right size. *American Naturalist* **118**, 405–422.
- Roper, C., Pignatelli, P. & Partridge, L. (1993). Evolutionary effect of selection on age at reproduction in larval and adult *Drosophila*. *Evolution* **47**, 445–455.
- Shen, P., Wang, F., Underhill, P. A., Franco, C., Yang, W.-H., Roxas, A., Sung, R., Lin, A. A., Hyman, R. W., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (2000). Population genetic implications from sequence variation in four Y chromosome genes. *Proceedings of the National Academy of Sciences of the USA* **97**, 7354–7359.
- Soller, M., Bownes, M. & Kubli, E. (1999). Control of oocyte maturation in sexually mature *Drosophila* females. *Developmental Biology* **208**, 337–351.
- Whitlock, M. C. & Barton, N. H. (1997). The effective size of a subdivided population. *Genetics* **146**, 427–441.
- Wilkinson-Herbots, H. M. (1998). Genealogy and subpopulation differentiation under various models of population structure. *Journal of Mathematical Biology* **37**, 535–585.
- Woolliams, J. A., Bijma, P. & Villaneuva, B. (1999). Expected genetic contributions and their impact on gene flow and genetic gain. *Genetics* **153**, 1009–1020.
- Wray, N. R. & Thompson, R. (1990). Predictions of rates of inbreeding in selected populations. *Genetical Research* **55**, 41–54.
- Wright, S. (1969). *Evolution and the Genetics of Populations*, vol. 2, *The Theory of Gene Frequencies*. Chicago: University of Chicago Press.
- Yi, S. & Charlesworth, B. (2000). Contrasting patterns of molecular evolution of the genes on the new and old sex chromosomes of *Drosophila miranda*. *Molecular Biology and Evolution* **17**, 703–717.