Life history effects on neutral polymorphism levels of autosomes and sex chromosomes

Supplementary Information

[a] Department of Biological Sciences, Columbia University, New York, NY 10027

[1] To whom correspondence should be addressed: ga2373@columbia.edu or gs2742@columbia.edu

Table of Contents

1. Haploid Model

Here we rigorously solve the haploid model with age-structure and endogenous reproductive variance, and relate our results with those of previous work that considered special cases of our model. In Section 1.2 we consider the implications of our assumptions about endogenous reproductive variance. In Section 1.3 we solve for the joint stationary distribution of the age and relative reproductive success associated with an allele, going backwards in time. Based on this distribution, we derive the stationary per-generation coalescence rate for a sample of two alleles, to obtain Eq. 7 in the main text:

$$N_e = \frac{M \cdot G}{W}. \tag{S1}$$

In Section 1.4, we recast these results in terms of total reproductive variance, showing that the relationship derived by Hill for the case with age-structure alone (1):

$$N_e = G \cdot M_1/V, \tag{S2}$$

applies to the extended model with endogenous reproductive variance, and derive Eq. 11 in for the total reproductive variance in this case, i.e.,

$$V = W \cdot (M_1/M). \tag{S3}$$

In section 1.5 we consider the case without endogenous reproductive variance, to show that the our results for the effective population size (Eq. S1) reduce to the formula obtained by Felsenstein (2), and to consider a simple example of how age-structure affects the effective population size.

1.1 Requirements on f_a

When we introduced the haploid model with endogenous reproductive variance, we assumed that each newborn is assigned a relative reproductive success vector $\vec{r}$, and denoted the (constant) proportion of individuals with a given vector $\vec{r}$ in age class a by $f_a(\vec{r})$ (see Table S1 for summary of notation). Here we describe the requirements on the probability mass function f_a that these assumptions entail. First, given that the probability of being born to a parent of age a is p_a, and to a specific parent of age a and with reproductive success $\vec{r}$ is $p_a \cdot \frac{r_a}{M_a}$, we require that $E_{f_a}(r_a) = \sum_{\vec{r}} f_a(\vec{r}) \cdot r_a = 1$ for any age a. Second, given that the number of individuals with a given $\vec{r}$ can only decrease with age (due to mortality), we further require that $M_a \cdot f_a(\vec{r}) \geq M_{a+1} \cdot f_{a+1}(\vec{r})$.

Third, requiring that the number of individuals of a given age *a* and with a given $\vec{r}$ is constant and equal to $M_a \cdot f_a(\vec{r})$, implies that this number needs to be an integer. Notably, if we would like to model the distribution of relative reproductive success using a given (continuous or discrete) distributions $\widetilde{f_a}(\vec{r})$, which satisfies the first two requirements, we would need to discretize $\widetilde{f_a}$ to obtain a probability mass function $\tilde{f}_a'$ such that $M_a \cdot \tilde{f}_a'(\vec{r})$ is always an integer. However, if we assume that the relative sizes of the age-class, i.e., the ratios M_i/M_j, are constant, and increase the total population sizes, the discretized functions $\tilde{f}_a'$ will approach $\widetilde{f_a}$, and the value of the $W_{i,j} = E_{\tilde{f}_j'}(r_i \cdot r_j) = E(r_i \cdot r_j | \text{survival to age } j)$ terms, which summarize the effect of endogenous reproductive variance on the effective population size, will approach $E_{\widetilde{f_j}}(r_i \cdot r_j)$. We implicitly assumed this limit when we considered the special case in which relative reproductive success is independent of age and of mortality rates. More generally, while the assumption that for any age *a*, $M_a \cdot f_a(\vec{r})$ is an integer, might appear highly restrictive, these restrictions are relaxed under the standard coalescent assumption that the population size is sufficiently large.

Notation	Definition	Remarks
p_a	Probability that a newborn descends from a parent of age a	$\sum_a p_a = 1$
q_a	Probability that a newborn descends from a parent of age $\geq a$	$q_a = \sum_{i \geq a} p_i$
G	Average generation time	$G = \sum_a a \cdot p_a$
M_a	Number of individuals of age a (M_1 is the number of newborns per-year)	$M_{a+1} \leq M_a$
$\vec{r}$	Relative reproductive success, where component r_a is the relative reproductive success at age a	
$f_a(\vec{r})$	The frequency of individuals with relative reproductive success $\vec{r}$ among individuals of age a	
$g_a(\vec{r})$	Given an individual I of age a and a newborn n, $g_a(\vec{r})$ is the probability that I has relative reproductive success $\vec{r}$, conditioned on n being descended from I	$g_a(\vec{r}) = r_a f_a(\vec{r})$
$\epsilon(a, \vec{r})$	Joint stationary probability of age a and relative reproductive success $\vec{r}$ along a lineage, going backwards in time	$\epsilon(a, \vec{r}) = \frac{1}{G} \sum_{j \geq a} p_j g_j(\vec{r})$
ϵ_a	Marginal stationary distribution of age a	$\epsilon_a = \frac{q_a}{G}$
M	Effective age-class size	See Eq. S16
$W_{i,j}$	Average value of $r_i \cdot r_j$ among individuals of sex s and age a	Defined for $i \leq j$
W	Weighted average of the $W_{i,j}$	See Eq. S15
X, X_a	An individual's number of offspring, throughout its life or at age a, respectively	
V	Reproductive variance (i.e., $V = Var(X)$)	See Eq. S28
S_a	The event of surviving to age $\geq a$	

Table S1: Notation for the haploid model, with parameters of the model in red.

1.2 Solution backwards in time

Here, we extend the derivations of Sagitov and Jagers (3) to account for endogenous reproductive variance. Tracing an allele backward in time, the age a_t and relative reproductive success $\vec{r}_t$ of the individual I_t who carries the allele t years in the past defines a Markov chain $(a_t, \vec{r}_t)$. To define the transition probabilities of the chain, we distinguish between two cases. First, if the individual carrying the

allele is not a newborn, i.e., $a_t > 1$, then at time t+1 that individual will be one year younger and its relative reproductive success $\vec{r}$ will remain unchanged, i.e., $(a_{t+1}, \vec{r}_{t+1}) = (a_t - 1, \vec{r}_t)$ with probability one. Second, if individual carrying the allele is a newborn, i.e., $a_t = 1$, then a_{t+1} equals a with probability p_a. The probability mass function of $\vec{r}_{t+1}$ conditional on a_{t+1}, follows from Bayes' theorem, further conditioning on the fact that the parent, $I_{t+1} = I$, necessarily reproduced successfully

$$P(\vec{r}_I = \vec{r} | I_{t+1} = I, a_{t+1} = a)$$
$$= \frac{P(I_{t+1}=I|\vec{r}_{t+1}=\vec{r}, a_{t+1}=a) \cdot P(\vec{r}_I=\vec{r}|\, a_{t+1}=a)}{P(I_{t+1}=I)} = \frac{(r_a/M_a) \cdot f_a(\vec{r})}{\sum_{\vec{k}} (r_a/M_a) \cdot f_a(\vec{k})} = r_a \cdot f_a(\vec{r}). \quad \text{(S4)}$$

We denote this probability by $g_a(\vec{r}) \equiv r_a \cdot f_a(\vec{r})$, and conclude that

$$P((a_{t+1}, \vec{r}_{t+1}) = (a, \vec{r}) | a_t = 1) = p_a \cdot g_a(\vec{r}). \quad \text{(S5)}$$

g_a is a proper probability mass function since $\sum_{\vec{r}} g_a(\vec{r}) = \sum_{\vec{r}} r_a \cdot f_a(\vec{r}) = 1$. Moreover, the parent's expected value of r_a is $E_{\vec{r}\sim g_a}(r_a) = E_{\vec{r}\sim f_a}(r_a^2) = 1 + V_{\vec{r}\sim f_a}(r_a) \geq 1$. The latter inequality makes intuitive sense, as it implies that the allele is more likely to be descended from an individual that has higher than average relative reproductive success in its age class.

We rely on the transition probabilities to derive and solve a recursion for the stationary probability $\epsilon(a, \vec{r})$ of age, a and relative reproductive successes, $\vec{r}$, of the individuals carrying the allele. Namely,

$$\epsilon(a, \vec{r}) = \epsilon(a+1, \vec{r}) + \left(\sum_{\vec{k}} \epsilon(1, \vec{k})\right) \cdot p_a \cdot g_a(\vec{r}), \quad \text{(S6)}$$

where the first term corresponds to aging within the same individual and the second corresponds to parenting a newborn. In order to solve these recursions we first consider the marginal stationary distribution of age, $\epsilon_a = \sum_{\vec{r}} \epsilon(a, \vec{r})$. To this end, we sum the recursions over $\vec{r}$ to obtain recursions on the marginal distribution,

$$\epsilon_a = \epsilon_{a+1} + \epsilon_1 \cdot p_a, \quad \text{(S7)}$$

where we also require that $\sum_a \epsilon_a = 1$. This recursion was solved by Sagitov and Jagers (3) for the case without endogenous reproductive variance, yielding

$$\epsilon_a = q_a / G, \quad \text{(S8)}$$

where $q_a \equiv \sum_{j \geq a} p_j$. Substituting this expression into Eq. S6, the recursions simplify to

$$\epsilon(a, \vec{r}) = \epsilon(a+1, \vec{r}) + \frac{1}{G} \cdot p_a \cdot g_a(\vec{r}), \quad \text{(S9)}$$

where we further require that $\sum_{a,\vec{r}} \epsilon(a, \vec{r}) = 1$. The solution of these recursions is

$$\epsilon(a,\vec{r}) = \frac{1}{G}\sum_{j\geq a} p_j g_j(\vec{r}). \quad \text{(S10)}$$

The marginal stationary probability mass function of $\vec{r}$ is $\sum_a \epsilon(a,\vec{r}) = \frac{1}{G}\sum_j \left(j \cdot p_j\right) \cdot g_j(\vec{r})$, which is a proper probability mass function because $\frac{1}{G}\sum_j j \cdot p_j = 1$, and $\sum_{\vec{r}} g_a(\vec{r}) = 1$ for any age *a*.

We rely on the stationary distribution to derive the probability of coalescence of two alleles, along the same lines as detailed in the main text for the case without endogenous reproductive variance. For the coalescence to occur at time *t* in the past, one of the alleles (A) would descent from the other (B) or both would descent from the same parental allele at that time (this is contrary to the case of non-overlapping generations, in which coalescence necessarily occurs when both alleles descent from the same parental allele in the previous generation). Specifically, if allele B is in an individual of age *a* and relative reproductive success $\vec{r}$ at time *t* (with probability $\epsilon(a,\vec{r})$), then allele A must be in a newborn at time *t*-1 (with probability ϵ_1) having descended from the same individual carrying allele B (with probability $p_a \cdot \frac{r_a}{M_a}$). Summing over the individual's possible ages and reproductive success vectors, we obtain the probability

$$\sum_{a,\vec{r}} \epsilon(a,\vec{r}) \cdot \epsilon_1 \cdot p_a \cdot \frac{r_a}{M_a} = \frac{1}{G^2}\sum_a \frac{\sum_{j\geq a} p_a p_j \sum_{\vec{r}} r_a g_j(\vec{r})}{M_a} = \frac{1}{G^2}\sum_a \frac{\sum_{j\geq a} p_a p_j W_{a,j}}{M_a}, \quad \text{(S11)}$$

where for $j \geq i$,

$$W_{i,j} = \sum_{\vec{r}} r_i g_j(\vec{r}) = \sum_{\vec{r}} r_i r_j f_j(\vec{r}) = E_{\vec{r}\sim f_j}(r_i \cdot r_j) \quad \text{(S12)}$$

is the expectation of $(r_i \cdot r_j)$ conditional on surviving to age $\geq j$. Further allowing for either allele or both to be the newborn (using the inclusion-exclusion principal to subtract the probability $\epsilon_1^2 \sum_a \frac{p_a^2 W_{a,a}}{M_a}$ that both alleles were in a newborn prior to coalescence), and measuring the coalescence rate in generations (rather than years), we obtain the per-generation coalescence rate and corresponding effective population size:

$$\frac{1}{N_e} = \frac{1}{G}\sum_a \frac{p_a^2 W_{a,a} + 2\sum_{j>a} p_a p_j W_{a,j}}{M_a}. \quad \text{(S13)}$$

Eq. S13 can be rearranged to obtain Eq. 7 of the main text. To this end, we define

$$w_i = (p_i^2 W_{i,i} + 2\sum_{j>i} p_i p_j W_{i,j})/W, \quad \text{(S14)}$$

where

$$W = \sum_i p_i^2 W_{i,i} + 2\sum_{i<j} p_i p_j W_{i,j} \quad \text{(S15)}$$

is a weighted average of the $W_{i,j}$. Noting that $\sum_a w_a = 1$, we then define the effective age class size as a weighted harmonic average of the age class sizes,

$$\frac{1}{M} = \sum_a \frac{w_a}{M_a}. \quad \text{(S16)}$$

Substituting this expression into Eq. S13, we obtain Eq. 7 of the main text:

$$\frac{1}{N_e} = W/(M \cdot G). \quad \text{(S17)}$$

1.3 Reproductive variance

To recast our results for N_e in terms of the total reproductive variance V, we first consider the case with non-overlapping generations in a haploid population of constant size, i.e., with Wright-Fisher sampling. We denote the number of offspring of the i^{th} individual by k_i and the census size by N. The expected number of offspring is 1, i.e., $\frac{1}{N}\sum_i k_i = 1$, and we denote the variance in number of offspring, which we also refer to as the reproductive variance, by $V = \frac{1}{N}\sum_i (k_i - 1)^2$. In the standard neutral model, without endogenous reproductive variance, $V = 1$. Since the probability that two distinct gametes descend from the same ancestor in the previous generation is $\sum_i \frac{k_i}{N} \cdot \frac{k_i - 1}{N-1} = \frac{V}{N-1}$, we find that the effective population size is

$$N_e = \frac{N-1}{V} \cong \frac{N}{V}, \quad \text{(S18)}$$

which is the expression derived by Wright (4) and presented in Eq. 9 of the main text.

To extend Eq. S18 to the case with overlapping generations, we consider the first two moments of an individual's number of offspring, X, throughout its lifetime. First, we note that an individual's number of offspring can be expressed as a sum over the number at each age, i.e., $X = \sum_a X_a$, where X_a is the number of offspring at age a; $X_a = 0$ if the individual does not survive to that age. In these terms, the first two moments are

$$E(X) = \sum_a E(X_a) \text{ and } E(X^2) = \sum_a E(X_a^2) + 2\sum_{i<j} E(X_i \cdot X_j). \quad \text{(S19)}$$

Denoting the event of surviving to age $\geq a$ by S_a, we note that

$$E\left(X_a^i\right) = Pr(S_a) \cdot E\left(X_a^i \middle| S_a\right) = \frac{M_a}{M_1} \cdot E\left(X_a^i \middle| S_a\right), \quad \text{(S20)}$$

The latter term, $E\left(X_a^i \middle| S_a\right)$, can be simplified further by conditioning on $\vec{r}$. Since the probability mass function of $\vec{r}$ conditional on S_a is f_a,

$$E\left(X_a^i \middle| S_a\right) = E_{\vec{r}\sim f_a} E\left(X_a^i \middle| S_a, \vec{r}\right). \quad \text{(S21)}$$

Moreover, the distribution of X_a conditional on S_a and $\vec{r}$ is simply $(X_a|\vec{r}, S_a) \sim Bin(M_1, p_a \cdot r_a/M_a)$, and therefore

$$E(X_a|S_a) = E_{\vec{r}\sim f_a}\left(\frac{M_1 r_a p_a}{M_a}\right) = \frac{M_1 p_a}{M_a}$$

and

$$E(X_a^2|S_a) = E_{\vec{r}\sim f_a}\left(M_1 \frac{r_a p_a}{M_a} + 2\binom{M_1}{2}\left(\frac{r_a p_a}{M_a}\right)^2\right) = \frac{M_1 p_a}{M_a} + 2\binom{M_1}{2}\left(\frac{p_a}{M_a}\right)^2 W_{a,a}. \quad \text{(S22)}$$

Substituting these expression into Eq. S20, we find that

$$E(X_a) = p_a \text{ and } E(X_a^2) = p_a + \frac{M_1 - 1}{M_a} p_a^2 \cdot W_{a,a}. \quad \text{(S23)}$$

To calculate the remaining terms in Eq. S19, $E(X_i \cdot X_j)$ for $j > i$, we note that conditioning on S_j, and on $\vec{r}|S_j$,

$$E\left(X_i \cdot X_j\right) = Pr\left(S_j\right) \cdot E\left(X_i \cdot X_j \middle| S_j\right) = \frac{M_j}{M_1} \cdot E_{\vec{r}\sim f_j} E\left(X_i \cdot X_j \middle| S_j, \vec{r}\right). \quad \text{(S24)}$$

The latter term is easily calculated, since conditional on S_j and $\vec{r}$, X_i and X_j are independent binomial variables, with $\left(X_i \middle| \vec{r}, S_j\right) \sim Bin(M_1, p_i \cdot r_i/M_i)$ and $\left(X_j \middle| \vec{r}, S_j\right) \sim Bin\left(M_1, p_j \cdot r_j/M_j\right)$, yielding

$$E\left(X_i \cdot X_j\right) = \frac{M_j}{M_1} \cdot E_{\vec{r}\sim f_j}\left(\frac{M_1^2 p_i p_j r_i r_j}{M_i M_j}\right) = \frac{M_1 p_i p_j W_{i,j}}{M_i}. \quad \text{(S25)}$$

Substituting the expressions from Eqs. S23 and S25 into Eq. S19 we find that

$$E(X) = 1 \text{ and } E(X^2) = 1 + M_1 \sum_i \frac{p_i^2 \cdot W_{i,i} + 2\sum_{j>i} p_i p_j W_{i,j}}{M_i} - \sum_i \frac{p_i^2 \cdot W_{i,i}}{M_i}. \quad \text{(S26)}$$

Assuming that the total population size is sufficiently large for the ratios M_i/M_j and terms $W_{i,j}$ to be approximated as fixed, and for the higher order term $\sum_i \frac{p_i^2 \cdot W_{i,i}}{M_i}$ to be negligible, we find that

$$E(X) = 1 \text{ and } E(X^2) \cong 1 + \frac{M_1}{M} W, \quad \text{(S27)}$$

and therefore the total reproductive variance is

$$V = E(X^2) - E^2(X) = \frac{M_1}{M} W, \quad \text{(S28)}$$

which is Eq. 11 of the main text. These assumptions correspond to the standard practice of neglecting higher order terms in $1/N$ in models with non-overlapping generations. From Eqs. S17 and S28 we find that the effective population size is

$$N_e = (G \cdot M_1)/V, \quad \text{(S29)}$$

which is the same form as in the case without age-structure (Eq. S18), and the general form presented in Eq. 10 of the main text.

1.4 Age-structure alone

Felsenstein (1971) used a different approach to solve the haploid model without endogenous reproductive variance, relying on the definition of the effective population size as the inbreeding effective number (2). To see that his results agree with ours (as well as with those of Sagitov and Jagers (3)), consider the case without endogenous reproductive variance, where Eq. S17 reduces to

$$N_e = MG = \frac{G}{\sum_i \frac{p_i^2 + 2\sum_{j>i} p_i p_j}{M_i}} = \frac{G}{\sum_i \frac{p_i}{M_i}(q_i + q_{i+1})}, \quad \text{(S30)}$$

where $q_i = \sum_{j \geq i} p_i$. Noting that $p_i(q_i + q_{i+1}) = (q_i - q_{i+1})(q_i + q_{i+1}) = q_i^2 - q_{i+1}^2$, we find that

$$N_e = \frac{G}{\sum_i \frac{p_i}{M_i}(q_i + q_{i+1})} = \frac{GM_1}{\sum_i \frac{M_1}{M_i}(q_i^2 - q_{i+1}^2)} = \frac{GM_1}{1 + \sum_i q_{i+1}^2 (\frac{M_1}{M_{i+1}} - \frac{M_1}{M_i})}, \quad \text{(S31)}$$

where Felsenstein's functional form (p. 585 in (2)) is on the rightmost side.

To better understand the effect of age-structure on the effective population size, consider a simple example in which there is no endogenous reproductive variance, and no age-dependence in reproductive success. In other words, the only difference among individuals' numbers of offspring arise from the stochasticity of mortality and reproduction. In this case, the probability of having a parent of age a is proportional to the size of the age class, i.e., $p_a = M_a/N$ where $N = \sum_a M_a$ is the census size. Following our derivations, the effective population size (Eq. 3 in the main text) then reduces to $N_e = \frac{G}{(2G-1)}N$, and if the generation time $G \gg 1$ then $N_e \approx \frac{1}{2}N$. In other words, the age structure reduces the effective population size to half of the census size

2. Diploid Model

2.1 Overview

Here we rigorously define and solve the diploid model with two sexes and endogenous reproductive variance, and derive formulas for the effective population sizes of X and autosomes. While the diploid model is more elaborate, the model and results follow along the same lines as we described for the haploid model. In Section 2.2 we detail the assumptions of the diploid model and introduce the notation required for the derivations that follow. In Section 2.3 we solve for the joint stationary distribution of the age and relative reproductive success of autosomal and X-linked alleles. We build on the joint stationary distribution to solving for the stationary per-generation coalescence rates and corresponding effective population sizes on X and autosomes. Since some of the explicit equations we derive are not presented in the main text, we briefly review them here.

Notably, to extend the haploid formula for the effective population size, $N_e = MG/W$ (Eqs. 7 and S17), to the diploid case, we require explicit expressions for the effective age-class size M, generation time G, and W, corresponding to the X and autosomes. First, we define these measures for each sex in the same way that we did in the haploid model (i.e., as in Eqs. S15 and S16). We then define G and W for X and autosomes, as simple weighted averages over their values in males and females:

$$G_X = \frac{2}{3}G_F + \frac{1}{3}G_M \text{ and } G_A = \frac{1}{2}(G_M + G_F) \tag{S32}$$

(which is Eq. 12 in the main text), and

$$W_X = \frac{2}{3}W_F + \frac{1}{3}W_M \text{ and } W_A = \frac{1}{2}(W_M + W_F), \tag{S33}$$

where the weights reflect the relative number of generations that X and autosomal linked loci spend in males and females (see Table S2 for notation). The effective age class sizes on X and autosomes are defined as weighted harmonic averages. In the case without endogenous reproductive variance, they are defined as

$$\frac{1}{M_X} = \frac{1/3}{M_M} + \frac{2/3}{M_F} \text{ and } \frac{1}{M_A} = \frac{1/2}{M_M} + \frac{1/2}{M_F}. \tag{S34}$$

To account for sex-specific endogenous reproductive variances, the weights further account for the endogenous reproductive variances effect on the relative probability of coalescence in males and females,

$$\frac{1}{M_X} = \frac{1/3(W_M/W_X)}{M_M} + \frac{2/3(W_F/W_X)}{M_F} \text{ and } \frac{1}{M_A} = \frac{1/2(W_M/W_X)}{M_M} + \frac{1/2(W_F/W_X)}{M_F}. \tag{S35}$$

Using these definitions, the effective population size for the X and autosomes take the form

$$N_e^A = \frac{2G_A M_A}{W_A} \text{ and } N_e^A = \frac{2G_X M_X}{W_X}, \tag{S36}$$

where the factor 2, which is absent in the haploid case (Eqs. 7 and S17), accounts for the effective number of age classes in the population (i.e., G classes in the haploid population, but $2G$ classes in the case with two sexes). To translate these effective sizes into coalescence rates, we also account for ploidy, yielding per generation rates of $1/2N_e^A = W_A/4G_A M_A$ on autosomes and $1/(3/2)N_e^X = W_X/3G_X M_X$ on the X. Based on Eq. S36, the mutation rate on X and autosomes (Eq. X), and the standard forms for polymorphism levels (Eq. X), we obtain the following expression for the ratio of X to autosome polymorphism levels:

$$\frac{E(\pi_X)}{E(\pi_A)} = \frac{3}{4} \cdot \frac{f\left(\frac{\mu_M}{\mu_F}\right) \cdot f\left(\frac{G_M}{G_F}\right)}{f\left(\frac{W_M/W_F}{M_M/M_F}\right)} \tag{S37}$$

(note that this differs from Eq. 18 in the main text).

In Section 2.4, we recast the results for the effective population size (Eq. S36) and polymorphism ratio (Eq. S37) in terms of male and female reproductive variances. First, we show that the reproductive variances in males and females, V_M and V_F, are given by

$$V_s = \frac{M_1}{\gamma_s}\frac{W_s}{M_s} - \frac{1-\gamma_s}{\gamma_s^2}, \tag{S38}$$

where the index s corresponds to M or F, and γ_M and γ_F are the proportions of males and females among newborns, respectively. Thus, this equation does not assume a sex ratio of 1. Rewriting Eq. S36 in terms of male and female reproductive variances we find that

$$N_e^X = \frac{4G_X M_1}{\frac{2}{3}\gamma_M V_M + \frac{4}{3}\gamma_F V_F + \frac{2}{3}\frac{\gamma_F}{\gamma_M} + \frac{4}{3}\frac{\gamma_M}{\gamma_F}} \text{ and } N_e^A = \frac{4G_A M_1}{\gamma_M V_M + \gamma_F V_F + \frac{\gamma_F}{\gamma_M} + \frac{\gamma_M}{\gamma_F}}, \tag{S39}$$

where M_1 is the number of newborns of both sexes per year, and that

$$\frac{E(\pi_X)}{E(\pi_A)} = \frac{3}{4} \cdot \frac{f(\mu_M/\mu_F) \cdot f(G_M/G_F)}{f\left(\frac{\gamma_F/\gamma_M + \gamma_M V_M}{\gamma_M/\gamma_F + \gamma_F V_F}\right)}. \tag{S40}$$

These equations reduce to Eqs. 15 and 18 in the main text when the sex ratio at birth equals 1.

In Section 2.5, we derive the expressions for the effective population sizes in terms of the allelic reproductive variances for X and autosomes, V_X^* and V_A^* (Eq. 13 in the main text):

$$N_e^A = \frac{G_A \cdot M_1}{V_A^*} \text{ and } N_e^X = \frac{G_X \cdot M_1}{V_X^*}, \quad \text{(S41)}$$

where the expression for the X only hold when the sex ratio at birth equals 1 (see below).

Notation	Definition	Remarks
$p_{s,a}$	Probability that a parent of sex s is of age a	$\Sigma_a p_{F,a} = \Sigma_a p_{M,a} = 1$
$q_{s,a}$	Probability that a parent of sex s is of age $\geq a$	$q_{s,a} = \Sigma_{i \geq a} p_{s,i}$
G_M, G_F	Male and female generation times	$G_s = \Sigma_a a \cdot p_{s,a}$
G_X, G_A	Generation times for X and autosomes	See Eq. S32
$M_{s,a}$	Number of individuals of sex s and age a	$M_{s,a+1} \leq M_{s,a}$
M_1	Number of newborns of both sexes per-year	
γ_M, γ_F	Proportions of males and females among newborns	$\gamma_s = M_{s,1}/M_1$
$\vec{r}$	Relative reproductive success	
$f_{s,a}(\vec{r})$	The proportion of individuals with relative reproductive success $\vec{r}$ among individuals of sex s and age a	
$g_{s,a}(\vec{r})$	Given a newborn that descended from a parent of sex s and age a, $g_{s,a}(\vec{r})$ is the probability that the parent has relative reproductive success $\vec{r}$	$g_{s,a}(\vec{r}) = r_a f_{s,a}(\vec{r})$
$\epsilon^X(s,a,\vec{r})$, $\epsilon^A(s,a,\vec{r})$	Joint stationary probability of sex s, age a and relative reproductive success $\vec{r}$ for the X and autosomes	See Eqs. S49 and S50
$\epsilon^X_{s,a}, \epsilon^A_{s,a}$	Marginal stationary distribution of age a for the X and autosomes	
M_M, M_F	Effective male and female age-class sizes	See Eq. S61
M_X, M_A	Effective X and autosome linked age-class sizes	See Eq. S65
$W_{s,i,j}$	Expectation of $r_i \cdot r_j$ among individuals of sex conditional of surviving to age $a \geq j$	Defined for $j \geq i$
W_M, W_F	Weighted averages of the $W_{M,i,j}$ and the $W_{F,i,j}$, respectively	See Eq. S60
W_X, W_A	Weighted averages of W_M and W_F for X and autosome linked loci	See Eq. S33
$X_s, X_{s,a}$	$X_{s,a}$ and X_s are Random variables describing the numbers of offspring an individual has at age a or throughout life, respectively	
V_M, V_F	Male and female reproductive variances (i.e., $V_s = V(X_s)$)	See Eq. S84
$S_{s,a}$	The event of a newborn of sex s surviving to age $\geq a$	
$f(x)$	$f(x) \equiv (2x+4)/(3x+3)$	
μ_M, μ_F	Male and female expected mutation rates per generation	See Section 3
μ_X, μ_A	Expected mutation rates per generation on X and autosomes; $\mu_X = \frac{1}{3}\mu_M + \frac{2}{3}\mu_F$ and $\mu_A = \frac{1}{2}\mu_M + \frac{1}{2}\mu_F$	
X^X_m, X^A_m	The number of newborns carrying a random X or autosome linked allele m	
V^*_X, V^*_A	Reproductive variances of X and autosome linked alleles, respectively (i.e., $V^*_X \equiv V(X^X_m)$ and $V^*_A \equiv V(X^A_m)$)	See Eqs. S90 and S96

Table S2: Notation for the diploid model with two sexes, with parameters of the model in red.

2.2 Assumptions and notation

We consider a panmictic, diploid population of constant size, with two sexes, and sex- and age-dependent mortalities, fecundities and reproductive variances. We measure age in years, and assume that the number of individuals of sex s and age a, $M_{s,a}$, is constant. Specifically, the sizes of the newborn age classes, $M_{M,a}$ and $M_{F,a}$, may take any integer values, meaning that at this point we make do not assume that the sex-ratios at birth equals 1. More generally, the size of classes can vary between sexes, but for each sex they decrease with age, i. e. , $M_{s,a+1} \leq M_{s,a}$, reflecting sex- and age-specific mortalities. We further assume that age classes are partitioned according to individuals' age-dependent reproductive success. Namely, individuals are randomly assigned a vector $\vec{r}$ at birth, reflecting their expected relative reproductive success at each age (see below). We then assume that the number of individuals in the population of sex s, age a and relative reproductive success $\vec{r}$, is constant and equal to $M_{s,a} \cdot f_{s,a}(\vec{r})$, where $f_{s,a}$ is the probability mass function of $\vec{r}$ among individuals of sex s and age a. Individuals with the same value of $\vec{r}$ are chosen to survive to the next age class at random, i.e., there are no differences in viability, but $M_{s,a} \cdot f_{s,a}(\vec{r}) \geq M_{s,a+1} \cdot f_{s,a+1}(\vec{r})$ due to mortality, where rates of mortality can depend on the value of $\vec{r}$.

Sex and age dependent fertility and reproductive success are described backwards in time, in terms of the probability of an individual being chosen as a parent. Every newborn has a mother and a father, which are chosen independently. The probability that the parent of sex s is of a given age is described by a discrete distribution $A_s = (p_{s,a})_{a=1}^{\infty}$, where the expectations $G_M = E(A_M)$ and $G_F = E(A_F)$ are the generation times for males and females, respectively. The average probability per individual of age a is therefore $p_{s,a}/M_{s,a}$, which can be viewed as the fertility associated with that age and sex. The probability of being born to a specific parent of age a and relative reproductive success $\vec{r}$ is $p_{s,a} \cdot \frac{r_a}{M_{s,a}}$, where r_a is the a-th component of $\vec{r}$. The value of r_a thus reflects an individual's expected (rather than actual) relative reproductive success.

Similar to the haploid case (cf. Section 1.1), our assumptions imply several requirements on the form of the probability mass functions $f_{s,a}$. First, requiring that the probability of a parent of sex s being of age a is $p_{s,a}$, implies that for any sex s and age a, $E_{f_{s,a}}(r_a) = 1$. Second, requiring that $M_{s,a} \cdot f_{s,a}(\vec{r}) \geq M_{s,a+1} \cdot f_{s,a+1}(\vec{r})$ implies that $f_{s,a}(\vec{r})/f_{s,a+1}(\vec{r}) \geq M_{s,a+1}/M_{s,a}$. Third, requiring that for any sex s and age a $M_{s,a} \cdot f_{s,a}(\vec{r})$ is an integer, implies that the probability mass functions $f_{s,a}$ are discrete and can only take values

$i/M_{s,a}$ for $i = 0, 1, \ldots, M_{s,a}$. While the latter requirement may appear to be highly restrictive, if we fix the ratios $M_{s,a}/M_{s',a'}$ and assume that the total population size is sufficiently large, we can relax this requirement and assume any continuous or discrete distributions $f_{s,a}$ that satisfy the first two requirements (by the same reasoning we applied to the haploid case in Section 1.2).

2.3 Solution backwards in time

Here, we extend the derivations of Pollak (5) to account for endogenous reproductive variance. Tracing an allele backward in time, the sex s_t, age a_t and relative reproductive success $\vec{r}_t$ of the individual I_t carrying the allele t years in the past defines a Markov chain, $(s_t, a_t, \vec{r}_t)$. To define the transition probabilities of the chain, we distinguish between two cases. First, if the allele is not carried by a newborn, i.e., if $a_t > 1$, then at time $t+1$ the individual carrying it was one year younger, and its sex s and relative reproductive success $\vec{r}$ remain unchanged, i.e., $(s_{t+1}, a_{t+1}, \vec{r}_{t+1}) = (s_t, a_t - 1, \vec{r}_t)$ with probability one. Second, if the allele is carried by a newborn, i.e., if $a_t = 1$, then the sex of the parent, s_{t+1}, is equally likely to be male or female if the allele is autosomal or if it is X-linked and the newborn was a female, but if it is X-linked and the newborn was a male then the sex of the parent will be female with probability 1. Conditional on the parent's sex, s_{t+1}, its age $a_{t+1} = a$ with probability $p_{s_{t+1},a}$. The probability mass function of $\vec{r}_{t+1}$ conditional on (s_{t+1}, a_{t+1}), follows from Bayes' theorem, further conditioning on the fact that the parent, $I_{t+1} = I$, necessarily reproduced successfully

$$P(\vec{r}_I = \vec{r} | I_{t+1} = I, a_{t+1} = a)$$
$$= \frac{P(I_{t+1}=I|\vec{r}_{t+1}=\vec{r}, a_{t+1}=a) \cdot P(\vec{r}_I=\vec{r}|\, a_{t+1}=a)}{P(I_{t+1}=I)} = \frac{(r_a/M_{s,a}) \cdot f_{s,a}(\vec{r})}{\sum_{\vec{k}} (r_a/M_{s,a}) \cdot f_{s,a}(\vec{k})} = r_a \cdot f_{s,a}(\vec{r}) \qquad \text{(S42)}$$

We denote this probability by $g_{s,a}(\vec{r}) \equiv r_a \cdot f_{s,a}(\vec{r})$, and conclude that when $a_t = 1$, s_{t+1} is distributed as we described above and $P\big((a_{t+1}, \vec{r}_{t+1}) = (a, \vec{r}) \big| s_{t+1}\big) = p_{s,a} \cdot g_{s,a}(\vec{r})$.

$g_{s,a}$ is a proper probability mass function since $\sum_{\vec{r}} g_{s,a}(\vec{r}) = \sum_{\vec{r}} r_a \cdot f_{s,a}(\vec{r}) = 1$. Moreover, the parent's expected value of r_a is $E_{\vec{r} \sim g_{s,a}}(r_a) = E_{\vec{r} \sim f_{s,a}}(r_a^2) = 1 + V_{\vec{r} \sim f_{s,a}}(r_a) \geq 1$. The latter inequality makes intuitive sense, as it implies that the allele is more likely to be descended from an individual that has higher than average relative reproductive success in its age class.

We rely on the transition probabilities to derive and solve recursions for the stationary probabilities, $\epsilon_A(s,a,\vec{r})$ and $\epsilon_X(s,a,\vec{r})$, of sex, s, age, a, and relative reproductive successes, $\vec{r}$, of autosome and X linked alleles, respectively. For autosomal alleles

$$\epsilon^A(s,a,\vec{r}) = \epsilon^A(s,a+1,\vec{r}) + \left(\sum_{t,\vec{k}} \epsilon^A\left(t,1,\vec{k}\right)\right) \cdot \frac{1}{2} p_{s,a} \cdot g_{s,a}(\vec{r}), \tag{S43}$$

where the first term corresponds to aging by one year and the second term corresponds to parenting a newborn. For X linked alleles

$$\epsilon^X(s,a,\vec{r}) = \epsilon^X(s,a+1,\vec{r}) + \left(\frac{1}{2}\sum_{\vec{k}} \epsilon^X(F,1,\vec{k}) + \mathbb{I}_{s=F} \sum_{\vec{k}} \epsilon^X(M,1,\vec{k})\right) \cdot p_{s,a} \cdot g_{s,a}(\vec{r}), \tag{S44}$$

where $\mathbb{I}$ denotes an indicator function, and, similar to the autosomal case, the first term corresponds to aging by one year and the second term corresponds to parenting a newborn.

In order to solve these recursions, we first consider the marginal stationary distribution of age and sex, $\epsilon^A_{s,a} = \sum_{\vec{r}} \epsilon^A(s,a,\vec{r})$ for autosomes and $\epsilon^X_{s,a} = \sum_{\vec{r}} \epsilon^X(s,a,\vec{r})$. To this end, we sum the recursions over $\vec{r}$ to obtain recursions on the marginal distributions,

$$\epsilon^A_{s,a} = \epsilon^A_{s,a+1} + \left(\epsilon^A_{M,1} + \epsilon^A_{F,1}\right) \cdot \frac{1}{2} p_{s,a} \text{ and } \epsilon^X_{s,a} = \epsilon^X_{s,a+1} + \left(\frac{1}{2}\epsilon^X_{F,1} + \mathbb{I}_{s=F}\epsilon^X_{M,1}\right) \cdot p_{s,a}, \tag{S45}$$

where we also require that $\sum_{s,a} \epsilon^A_{s,a} = \sum_{s,a} \epsilon^X_{s,a} = 1$. These recursions were solved by Pollack (5) for the case without endogenous reproductive variance, yielding

$$\epsilon^A_{s,a} = q_{s,a}/2G_A,\ \epsilon^X_{M,a} = q_{M,a}/3G_X \text{ and } \epsilon^X_{F,a} = 2q_{F,a}/3G_X, \tag{S46}$$

where $q_{s,j} \equiv \sum_{j \geq a} p_{s,j}$ is the probability that a parent of sex s is at least j years old. Substituting these expressions into Eqs. S43 and S44, the recursions simplify to

$$\epsilon^A(s,a,\vec{r}) = \epsilon^A(s,a+1,\vec{r}) + \frac{1}{2G_a} p_{s,a} \cdot g_{s,a}(\vec{r}) \tag{S47}$$

for autosomes and

$$\epsilon^X(s,a,\vec{r}) = \epsilon^X(s,a+1,\vec{r}) + \frac{1+\mathbb{I}_{s=F}}{3G_X} \cdot p_{s,a} \cdot g_{s,a}(\vec{r}) \tag{S48}$$

for the X, where we further require that $\sum_{s,a,\vec{r}} \epsilon^A(s,a,\vec{r}) = \sum_{s,a,\vec{r}} \epsilon^X(s,a,\vec{r}) = 1$. The solution to these recursions is

$$\epsilon^A(s,a,\vec{r}) = \frac{1}{2G_A} \epsilon(s,a,\vec{r}) \tag{S49}$$

for autosomes and

$$\epsilon^X(s,a,\vec{r}) = \frac{1+\mathbb{I}_{s=F}}{3G_X}\epsilon(s,a,\vec{r}) \quad \text{(S50)}$$

for the X, where $\epsilon(s,a,\vec{r}) \equiv \sum_{j\geq a} p_{s,j} \cdot g_{s,j}(\vec{r})$.

The marginal stationary probability mass function of $\vec{r}$ follows,

$$\epsilon^A_{\vec{r}} = \sum_{s,a} \epsilon^A(s,a,\vec{r}) = \sum_{s,j} \frac{j \cdot p_{s,j}}{2G_A} \cdot g_{s,j}(\vec{r}) \quad \text{(S51)}$$

for autosomes, and

$$\epsilon^X_{\vec{r}} = \sum_{s,a} \epsilon^X(s,a,\vec{r}) = \sum_{s,j} \frac{(1+\mathbb{I}_{s=F}) \cdot j \cdot p_{s,j}}{3G_X} \cdot g_{s,j}(\vec{r}) \quad \text{(S52)}$$

for the X. These are proper probability mass functions since they are weighted averages of the probability mass functions $g_{s,j}$, since $\sum_{s,j} \frac{j \cdot p_{s,j}}{2G_A} = \sum_{s,j} \frac{(1+\mathbb{I}_{s=F}) \cdot j \cdot p_{s,j}}{3G_X} = 1$.

Similar to the haploid case, we rely on the stationary distribution to derive the probability of coalescence of two alleles. Consider the autosomal case first. For coalescence to occur at time *t* in the past, one of the alleles (A) would descent from the other (B) or both would descent from the same parental allele at that time. Specifically, if allele B is in an individual of sex *s*, age *a* and relative reproductive success $\vec{r}$ at time *t* (with probability $\epsilon^A(s,a,\vec{r})$), then allele A must be in a newborn at time *t*-1 (with probability $\epsilon_{M,1} + \epsilon_{F,1}$), having descended from the same individual carrying allele B (with probability $\frac{1}{2} p_{s,a} \cdot \frac{r_a}{M_a}$) and from allele B specifically (with probability $1/2$). Summing over the individual's possible sexes, ages and reproductive success vectors, we obtain the probability

$$\sum_{s,a,\vec{r}} \epsilon^A(s,a,\vec{r}) \cdot \left(\epsilon^A_{M,1} + \epsilon^A_{F,1}\right) \cdot \frac{1}{2} p_{s,a} \cdot \frac{r_a}{2M_{s,a}} = \frac{1}{8(G_A)^2} \sum_{s,a} \frac{\sum_{j\geq a} p_{s,a} p_{s,j} \sum_{\vec{r}} r_a \cdot g_{s,j}(\vec{r})}{M_{s,a}}$$

$$= \frac{1}{8(G_A)^2} \sum_{s,a} \frac{\sum_{j\geq a} p_{s,a} p_{s,j} W_{s,a,j}}{M_{s,a}}, \quad \text{(S53)}$$

where, for $j \geq i$,

$$W_{s,i,j} \equiv E_{\vec{r}\sim f_{s,j}}\left(r_i \cdot r_j\right) = E_{\vec{r}\sim g_{s,j}}(r_i) = \sum_{\vec{r}} r_i \cdot g_{s,j}(\vec{r}) \quad \text{(S54)}$$

is the expectation of $\left(r_i \cdot r_j\right)$ over individuals of sex *s* and age *j*. Further allowing for either allele or both to be the newborn (using the inclusion-exclusion principal to subtract the probability

$$\left(\epsilon^A_{M,1} + \epsilon^A_{F,1}\right)^2 \sum_{s,a,\vec{r}} \left(\frac{1}{2} p_{s,a}\right)^2 \cdot \frac{r_a g_{s,a}(\vec{r})}{2M_{s,a}} = \frac{1}{8(G_A)^2} \sum_{s,a} \frac{p_{s,a} p_{s,a} W_{s,a,a}}{M_{s,a}} \quad \text{(S55)}$$

that both alleles were in a newborn prior to coalescence), the autosomal stationary coalescence rate per year is

$$\frac{1}{8(G_A)^2}\sum_{s,a}\frac{p_{s,a}^2 W_{s,a,a}+2\sum_{j>a}p_{s,a}p_{s,j}W_{s,a,j}}{M_{s,a}}. \quad \text{(S56)}$$

The per generation coalescence rate (in terms of the autosomal generation time G_A) and corresponding effective population size are therefore

$$\frac{1}{2N_e^A}=\frac{1}{8\cdot G_A}\sum_{s,a}\frac{p_{s,a}^2 W_{s,a,a}+2\sum_{j>a}p_{s,a}p_{s,j}W_{s,a,j}}{M_{s,a}}. \quad \text{(S57)}$$

For the X, the stationary coalescence rate per year is

$$\begin{aligned}&2\left(\epsilon_{M,1}^X+\frac{1}{2}\epsilon_{F,a}^X\right)\sum_{\mathrm{a},\vec{r}}\epsilon^X(F,a,\vec{r})\cdot p_{F,a}\cdot\frac{r_a}{2M_{F,a}}+2\cdot\frac{1}{2}\epsilon_{F,a}^X\sum_{\mathrm{a},\vec{r}}\epsilon^X(M,a,\vec{r})\cdot p_{M,a}\cdot\frac{r_a}{M_{M,a}}\\&-\left(\epsilon_{M,1}^X+\frac{1}{2}\epsilon_{F,1}^X\right)^2\sum_{a,\vec{r}}p_{F,a}^2\cdot\frac{r_a g_{F,a}(\vec{r})}{2M_{F,a}}-\left(\frac{1}{2}\epsilon_{F,1}^X\right)^2\sum_{a,\vec{r}}p_{M,a}^2\cdot\frac{r_a g_{M,a}(\vec{r})}{M_{M,a}}\\&=\frac{1}{9(G_X)^2}\sum_s(1+\mathbb{I}_{s=F})\sum_a\frac{p_{s,a}^2 W_{s,a,a}+2\sum_{j>a}p_{s,a}p_{s,j}W_{s,a,j}}{M_{s,a}},\end{aligned} \quad \text{(S58)}$$

and the corresponding per generation coalescence rate, which defines the effective population size for the X, N_e^X, is

$$\frac{1}{(3/2)N_e^X}=\frac{1}{9G_X}\sum_s(1+\mathbb{I}_{s=F})\sum_a\frac{p_{s,a}^2 W_{s,a,a}+2\sum_{j>a}p_{s,a}p_{s,j}W_{s,a,j}}{M_{s,a}} \quad \text{(S59)}$$

(defined in terms of the X-linked generation time G_X).

As outlined in Section 2.1, the effective population sizes, N_e^X and N_e^A, can be rewritten in terms of the effective age class sizes, to obtain expressions that are analogous to Eqs. 7 and S17 in the haploid case. To this end, the terms G, W and M in Eq. S17 need to be defined for the X and autosomes. First, we define these terms separately for males and females, by applying the haploid definitions. Specifically, we define

$$W_s=\sum_i p_{s,i}^2 W_{s,i,i}+2\sum_{i<j}p_{s,i}p_{s,j}W_{s,i,j} \quad \text{(S60)}$$

as a weighted average of the $W_{s,i,j}$, and define

$$\frac{1}{M_s}=\sum_a\frac{w_{s,a}}{M_{s,a}} \quad \text{(S61)}$$

as a weighted harmonic average of the age classes sizes of sex *s*, with weights

$$w_{s,i}=(p_{s,i}^2 W_{s,i,i}+2\sum_{j>i}p_{s,i}p_{s,j}W_{s,i,j})/W_s \quad \text{(S62)}$$

(note that $\sum_a w_{s,a}=1$).

To extend the definitions of G, W and M to the X and autosomes, we define them as weighted averages over males and females. Specifically, *G* and *W* are defined as simple weighted averages,

$$G_A = \frac{1}{2}(G_M + G_F) \text{ and } G_X = \frac{2}{3}G_F + \frac{1}{3}G_M \tag{S63}$$

(this is Eq. 12 in the main text) and

$$W_A = \frac{1}{2}(W_M + W_F) \text{ and } W_X = \frac{2}{3}W_F + \frac{1}{3}W_M. \tag{S64}$$

The effective age class size M for X and autosomes is defined as a weighted harmonic average,

$$\frac{1}{\mathrm{M_A}} = \frac{1/2(\mathrm{W_M}/\mathrm{W_A})}{\mathrm{M_M}} + \frac{1/2(\mathrm{W_F}/\mathrm{W_A})}{\mathrm{M_F}} \text{ and } \frac{1}{\mathrm{M_X}} = \frac{1/3(\mathrm{W_M}/\mathrm{W_X})}{\mathrm{M_M}} + \frac{2/3(\mathrm{W_F}/\mathrm{W_X})}{\mathrm{M_F}}. \tag{S65}$$

Expressing Eqs. S57 and S59 in these terms, we find that

$$N_e^A = \frac{2M_A G_A}{W_A} \text{ and } N_e^X = \frac{2M_X G_X}{W_X}. \tag{S66}$$

The factor 2, which is absent in the haploid case (Eq. S17), reflects the effective number of age classes (i.e., G classes of size M in the haploid model, but $2G$ classes in the diploid model with two sexes).

Assuming the standard expressions for neutral heterozygosity, $E(\pi_A) = 4N_e^A\mu_A$ and $E(\pi_X) = 3N_e^X\mu_X$ (see Section 3), and rearranging the expressions in Eq. S66, we find that

$$\frac{E(\pi_X)}{E(\pi_A)} = \frac{3}{4} \cdot \frac{f(\mu_M/\mu_F)\cdot f(G_M/G_F)}{f\left(\frac{W_M/W_F}{M_M/M_F}\right)}. \tag{S67}$$

When the mutation rate, age structure, and endogenous reproductive variance are identical in both sexes Eq. S67 reduces to the naïve neutral expectation of ¾. When these factors differ among sexes, Eq. S67 provides a simple expression for the effect of each factor.

2.4 Reproductive variance

To recast our results for the effective population sizes in terms of total reproductive variances in males and females, V_M and V_F, we follow the same steps as described for the haploid case (Section 1.3). First, we consider the case with non-overlapping generations in a diploid population of constant size, with N_M males and N_F females. We denote the total population size by $N \equiv N_M + N_F$, the proportions of males and females by $\gamma_s \equiv \frac{N_s}{N}$, and the number of offspring of the i^{th} individual of sex s by k_i^s. To maintain a constant population size, we require that the number of offspring arising from parents of each sex equals N, and therefore the sex-specific expectations are $E(k_i^s) = \frac{1}{N_s}\sum_i k_i^s = \frac{N}{N_s}$. We denote the sex-specific variances by $V_s \equiv V(k_i^s)$.

We are interested in the probability that two distinct alleles descend from the same allele in the previous generation, as this probability equals $1/2N_e^A$ for autosomes and $1/(3/2)N_e^X$ for the X. For autosomes, the probability that the two alleles descend from individuals of sex *s* is ¼, the probability that they descend from the same individual of that sex is $\sum_{i=1}^{N_s} \frac{k_i^s}{N} \cdot \frac{k_i^s-1}{N-1}$, and the probability that they descend from the same allele is 1/2 , therefore

$$\frac{1}{2N_e^A} = \frac{1}{8}\sum_s \sum_{i=1}^{N_s} \frac{k_i^s}{N} \cdot \frac{k_i^s-1}{N-1}. \quad \text{(S68)}$$

Substituting $\sum_{i=1}^{N_s} \frac{k_i^s}{N} \cdot \frac{k_i^s-1}{N-1} = \frac{\gamma_s}{N-1}\left(E\left(k_i^{s\,2}\right) - E(k_i^s)\right) = \frac{1}{N-1}(\gamma_s V_s + \frac{1}{\gamma_s} - 1)$ into Eq. S68, we find that

$$N_e^A = \frac{4(N-1)}{\gamma_M V_M + \gamma_F V_F + \frac{\gamma_F}{\gamma_M} + \frac{\gamma_M}{\gamma_F}} \cong \frac{4N}{\gamma_M V_M + \gamma_F V_F + \frac{\gamma_F}{\gamma_M} + \frac{\gamma_M}{\gamma_F}} \quad \text{(S69)}$$

(cf. (4)).

For the X chromosome, the probability that two alleles descend from individuals of sex *s* depends on γ_M and γ_F. However, as we go further backwards in time, this probability approaches 1/9 for both being male and 4/9 for both being female, regardless of γ_M and γ_F. The probability that both alleles descend from the same individual of that sex is $\sum_{i=1}^{N_s} \frac{k_i^s}{N} \cdot \frac{k_i^s-1}{N-1}$, and the probability that they descend from the same allele is ½ for females and 1 for males, and therefore

$$\frac{1}{(3/2)N_e^X} = \frac{1}{9}\sum_{i=1}^{N_M} \frac{k_i^M}{N} \cdot \frac{k_i^M-1}{N-1} + \frac{1}{2} \cdot \frac{4}{9}\sum_{i=1}^{N_F} \frac{k_i^F}{N} \cdot \frac{k_i^F-1}{N-1}, \quad \text{(S70)}$$

and thus

$$N_e^X = \frac{4(N-1)}{\frac{2}{3}\gamma_M V_M + \frac{4}{3}\gamma_F V_F + \frac{2}{3}\frac{\gamma_F}{\gamma_M} + \frac{4}{3}\frac{\gamma_M}{\gamma_F}} \cong \frac{4N}{\frac{2}{3}\gamma_M V_M + \frac{4}{3}\gamma_F V_F + \frac{2}{3}\frac{\gamma_F}{\gamma_M} + \frac{4}{3}\frac{\gamma_M}{\gamma_F}}. \quad \text{(S71)}$$

Assuming a sex ratio of 1 (i.e., $\gamma_M = \gamma_F = 1/2$), Eqs. S69 and S71 reduce to

$$N_e^A = \frac{4N}{2 + \frac{1}{2}V_M + \frac{1}{2}V_F} \text{ and } N_e^X = \frac{4N}{2 + \frac{1}{3}V_M + \frac{2}{3}V_F}. \quad \text{(S72)}$$

To extend Eq. S73 to the case with overlapping generations, we consider the first two moments of an individual's number of offspring, X_s, throughout its lifetime. First, we note that an individual's number of offspring can be expressed as a sum over the number at each age, i.e., $X_s = \sum_a X_{s,a}$, where $X_{s,a}$ denotes the number of offspring at age *a*; and $X_{s,a} = 0$ if the individual does not survive to age *a*. In these terms, the first two moments are

$$E(X_s) = \sum_a E(X_{s,a}) \text{ and } E(X_s^2) = \sum_a E(X_{s,a}^2) + 2\sum_{j>i} E(X_{s,i} \cdot X_{s,j}). \quad \text{(S74)}$$

Denoting the event of surviving to age $\geq a$ by $S_{s,a}$, we note that

$$E\left(X_{s,a}^i\right) = Pr\left(S_{s,a}\right) \cdot E\left(X_{s,a}^i \middle| S_{s,a}\right) = \frac{M_{s,a}}{M_{s,1}} \cdot E\left(X_{s,a}^i \middle| S_{s,a}\right). \quad \text{(S75)}$$

The latter term, $E\left(X_{s,a}^i \middle| S_{s,a}\right)$, can be simplified further by conditioning on $\vec{r}$. Since the probability mass function of $\vec{r}$ conditional on $S_{s,a}$ is $f_{s,a}$,

$$E\left(X_{s,a}^i \middle| S_{s,a}\right) = E_{\vec{r}\sim f_{s,a}} E\left(X_{s,a}^i \middle| S_{s,a}, \vec{r}\right). \quad \text{(S76)}$$

Moreover, the distribution of $X_{s,a}$ conditional on $S_{s,a}$ and $\vec{r}$ is

$$\left(X_{s,a} \middle| \vec{r}, S_{s,a}\right) \sim Bin\left(M_1, p_{s,a} \cdot r_a / M_{s,a}\right), \quad \text{(S77)}$$

where $M_1 = M_{M,1} + M_{F,1}$ is the number of newborns of both sexes per-year, and therefore

$$E\left(X_{s,a} \middle| S_{s,a}\right) = E_{\vec{r}\sim f_{s,a}}\left(\frac{M_1 r_a p_{s,a}}{M_{s,a}}\right) = \frac{M_1 p_{s,a}}{M_{s,a}} \quad \text{(S78)}$$

and

$$E\left(X_{s,a}^2 \middle| S_{s,a}\right) = E_{\vec{r}\sim f_{s,a}}\left(M_1 \frac{r_a p_{s,a}}{M_{s,a}} + 2\binom{M_1}{2}\left(\frac{r_a p_{s,a}}{M_{s,a}}\right)^2\right) = \frac{M_1 p_{s,a}}{M_{s,a}} + 2\binom{M_1}{2}\left(\frac{p_{s,a}}{M_{s,a}}\right)^2 W_{s,a,a}.$$

Substituting these expression into Eq. S75, we find that

$$E(X_{s,a}) = \frac{p_{s,a}}{\gamma_s} \text{ and } E\left(X_{s,a}^2\right) = \frac{p_{s,a}}{\gamma_s} + \frac{M_1 - 1}{M_{s,a}} \frac{p_{s,a}^2}{\gamma_s} \cdot W_{s,a,a}, \quad \text{(S79)}$$

where γ_M and γ_F are the proportions of males and females at birth (i.e., $\gamma_s = M_{s,1}/M_1$). To calculate the remaining terms in Eq. S74, $E(X_{s,i} \cdot X_{s,j})$ for $j > i$, we note that conditioning on $S_{s,j}$, and on $\vec{r}|S_{s,j}$,

$$E\left(X_{s,i} \cdot X_{s,j}\right) = P\left(S_{s,j}\right) \cdot E\left(X_{s,i} \cdot X_{s,j} \middle| S_{s,j}\right) = \frac{M_{s,j}}{M_{s,1}} \cdot E_{\vec{r}\sim f_{s,j}} E\left(X_{s,i} \cdot X_{s,j} \middle| S_{s,j}, \vec{r}\right). \quad \text{(S80)}$$

The latter term is easily calculated, since conditional on $S_{s,j}$ and $\vec{r}$, $X_{s,i}$ and $X_{s,j}$ are independent binomial variables: $\left(X_{s,i} \middle| \vec{r}, S_{s,j}\right) \sim Bin\left(M_1, p_{s,i} \cdot r_i / M_{s,i}\right)$ and $\left(X_{s,j} \middle| \vec{r}, S_{s,j}\right) \sim Bin\left(M_1, p_{s,j} \cdot r_j / M_{s,j}\right)$, and therefore

$$E\left(X_{s,i} \cdot X_{s,j}\right) = \frac{M_{s,j}}{M_{s,1}} \cdot E_{\vec{r}\sim f_{s,j}}\left(\frac{M_1^2 p_{s,i} p_{s,j} r_i r_j}{M_{s,i} M_{s,j}}\right) = \frac{M_1 p_{s,i} p_{s,j} W_{s,i,j}}{\gamma_s M_{s,i}}. \quad \text{(S81)}$$

Substituting the expressions from Eqs. S79 and S81 into Eq. S74 we obtain

$$E(X_s) = \frac{1}{\gamma_s} \text{ and } E(X_s^2) = \frac{1}{\gamma_s} + \frac{M_1}{\gamma_s}\sum_i \frac{p_{s,i}^2 \cdot W_{s,i,i} + 2\sum_{j>i} p_{s,i} p_{s,j} W_{s,i,j}}{M_{s,i}} - \sum_a \frac{p_{s,a}^2 \cdot W_{s,a,a}}{\gamma_s M_{s,a}}. \quad \text{(S82)}$$

Assuming that the total population size is sufficiently large for the ratios $M_{s,i}/M_{t,j}$ and terms $W_{s,i,j}$ to be approximated as fixed, and for the higher order terms $\sum_a \frac{p_{s,a}^2 \cdot W_{s,a,a}}{\gamma_s M_{s,a}}$ to be negligible, we find that

$$E(X_s) = \frac{1}{\gamma_s} \text{ and } E(X_s^2) \cong \frac{1}{\gamma_s} + \frac{M_1}{\gamma_s}\frac{W_s}{M_s}. \quad \text{(S83)}$$

The total reproductive variances of sex s, are therefore

$$V_s = E(X_s^2) - E^2(X_s) \cong \frac{M_1}{\gamma_s}\frac{W_s}{M_s} - \frac{1-\gamma_s}{\gamma_s^2}. \quad \text{(S84)}$$

From Eqs. S66 and S84, we obtain that

$$N_e^A = \frac{4G_A M_1}{\gamma_M V_M + \gamma_F V_F + \frac{\gamma_F}{\gamma_M} + \frac{\gamma_M}{\gamma_F}} \text{ and } N_e^X = \frac{4G_X M_1}{\frac{2}{3}\gamma_M V_M + \frac{4}{3}\gamma_F V_F + \frac{2}{3}\frac{\gamma_F}{\gamma_M} + \frac{4}{3}\frac{\gamma_M}{\gamma_F}}, \quad \text{(S85)}$$

where $G_A M_1$ and $G_X M_1$ are the total numbers of newborns per-generation, for autosomes and the X, respectively. Eq. S85 thus generalizes Eqs. S69 and S71 to the case with age-structure.

Assuming that $E(\pi_A) = 4N_e^A \mu_A$ and $E(\pi_X) = 3N_e^X \mu_X$ (see Section 3), we find that

$$\frac{E(\pi_X)}{E(\pi_A)} = \frac{3}{4} \cdot \frac{f(\mu_M/\mu_F) \cdot f(G_M/G_F)}{f\left(\frac{\gamma_F/\gamma_M + \gamma_M V_M}{\gamma_M/\gamma_F + \gamma_F V_F}\right)}. \quad \text{(S86)}$$

Further assuming that the sex ratio at birth is 1 (i.e. that $\gamma_M = \gamma_F = {}^1/_2$), Eqs. S85 and S86 reduce to

$$N_e^A = \frac{4G_A M_1}{2 + \frac{1}{2}V_M + \frac{1}{2}V_F}, \; N_e^X = \frac{4G_X M_1}{2 + \frac{1}{3}V_M + \frac{2}{3}V_F} \text{ and } \frac{E(\pi_X)}{E(\pi_A)} = \frac{3}{4} \cdot \frac{f(\mu_M/\mu_F) \cdot f(G_M/G_F)}{f\left(\frac{2+V_M}{2+V_F}\right)}, \quad \text{(S87)}$$

which are the expressions presented in Eqs. 15 and 18 of the main text.

2.5 Allelic reproductive variance

In the main text, we derived the effective population size for X and autosomes in terms of alleles rather than individuals, on the premise that we can then use the expressions obtained in the haploid model (Eqs. 10 and S29). Here we establish this premise, showing it to always be correct for autosomal alleles, whereas for the X it applies so long as the sex ratio at birth equals 1 (i.e., $\gamma_M = \gamma_F = 1/2$).

Consider an allele m carried by an individual I_m of sex s_m. We define the allele's (realized) reproductive success as the number of I_m's offspring who endogenous a copy of m, and denote it by X_m^A when m is autosomal and by X_m^X when it is X-linked. To obtain expressions for the effective population size in allelic terms, we calculate the first two moments of X_m^A and X_m^X, where m is chosen at random among alleles in newborns (in particular., m is carried by a male with probability $2M_{M,1}/(2M_{M,1} + 2M_{F,1})$ for autosomes and $M_{M,1}/(2M_{M,1} + 2M_{F,1})$ for the X).

First consider the case of an autosomal allele. To that end, we denote the total number of offspring of individual I_m by X_I. Since each offspring of I_m carries a copy of m with probability ½, the conditional distribution $X_m^A|X_I \sim Bin(X_I, \frac{1}{2})$. Based on the law of total variance, therefore

$$E(X_m^A) = \frac{1}{2}E(X_I) \text{ and } V(X_m^A) = \frac{1}{4}[E(X_I) + V(X_I)]. \quad \text{(S88)}$$

Further conditioning on the sex of the individual carrying the allele, s_I, we note that $E(X_I|s_I) = 1/\gamma_{s_I}$ (Eq. S83) and $V(X_I|s_I) = V_{s_I}$, where the individual I_m is male with probability γ_M and female with probability γ_F. Applying the law of total variance again, we obtain

$$E(X_I) = 2 \text{ and } V(X_I) = \gamma_M V_M + \gamma_F V_F + \frac{(\gamma_M - \gamma_F)^2}{\gamma_M \gamma_F}. \quad \text{(S89)}$$

Substituting these expressions into Eq. S89, we find that $E(X_m^A) = 1$ and

$$V_A^* \equiv V(X_m^A) = \frac{1}{4}[\gamma_M V_M + \gamma_F V_F + \frac{\gamma_M}{\gamma_F} + \frac{\gamma_F}{\gamma_M}]. \quad \text{(S90)}$$

When the sex-ratio at birth is 1, and thus $\gamma_M = \gamma_F = 1/2$, Eq. S91 reduces to Eq. 14 for V_A^* in the main text. From Eqs. S85 and S90, we obtain

$$N_e^A = \frac{G_A \cdot M_1}{V_A^*} \quad \text{(S92)}$$

for any sex-ratio, which is Eq. 13 in the main text. While direct analogy with Eqs. 10 and S29 for the haploid case would result in an effective population size of $N_e = 2G_A \cdot M_1/V_A^*$, given that the effective population sizes are defined by requiring coalescence rates of $1/N_e$ in haploids and $1/(2 \cdot N_e^A)$ in diploids, Eq. S93 is, in fact, analogous to Eqs. 10 and S29.

Next, consider the case of an X-linked allele. If the individual carrying the allele, I_m, is a male, then only his female offspring will inherit the allele, and thus, $X_m^X|(s_I = M, X_I) \sim Bin(X_I, \gamma_F)$. Since $E(X_I|s_I = M) = 1/\gamma_M$ (Eq. S83) and $V(X_I|s_I) = V_{s_I}$, the law of total variance implies that

$$E(X_m^X|s_I = M) = \gamma_F/\gamma_M \text{ and } V(X_m^X|s_I = M) = \gamma_F^2 V_M + \gamma_F. \quad \text{(S94)}$$

The case in which I_m is a female is similar to the autosomal case, and thus, $X_m^X|(s_I = F, X_I) \sim Bin(X_I, 1/2)$,

$$E(X_m^X|s_I = F) = \frac{1}{2\gamma_F} \text{ and } V(X_m^X|s_I = F) = \frac{1}{4}V_F + \frac{1}{4\gamma_F}. \quad \text{(S95)}$$

Given that there are $M_{M,1}$ X-linked alleles in newborn males and $2M_{F,1}$ in newborn females, the probability that an X-linked allele in a newborn is in a male is $\gamma_M/(1 + \gamma_F)$ and the probability it is in a female is $2\gamma_F/(1 + \gamma_F)$. Applying the law of total variance therefore implies that

$$E(X_m^X) = 1 \text{ and } V_X^* = Var(X_m^X) = \frac{\gamma_M \gamma_F^2}{1+\gamma_F} V_M + \frac{\gamma_F}{2(1+\gamma_F)} V_F + \frac{1+2\gamma_M\gamma_F}{2(1+\gamma_F)} + \frac{(1-2\gamma_F)^2}{2\gamma_M\gamma_F}. \quad \text{(S96)}$$

When the sex-ratio at birth is 1, and thus $\gamma_M = \gamma_F = 1/2$, Eq. S96 reduces to Eq. 14 for V_X^* in the main text. From Eqs. S85 and S96, we find that

$$N_e^X = \frac{G_X \cdot M_1}{V_X^*}, \quad \text{(S97)}$$

which is Eq. 13 in the main text, only holds when $\gamma_M = \gamma_F = 1/2$. Thus, the haploid result (Eqs. 10 and S29) applies to X-linked alleles only when the sex ratio at birth equals 1.

To gain some intuition as to why this result fails in the general case, consider the reproductive success of an X-linked allele in consecutive generations. As we have shown above, an allele's expected reproductive success is γ_F/γ_M in males and $1/(2\gamma_F)$ in females (averaged over the sexes the expectation is 1). Now consider the expected reproductive success in the next generation: if the allele was in a male in the previous generation it will necessarily be in a female, and the expected reproductive success of the offspring allele would be $1/(2\gamma_F)$; if the allele was in a female in the previous generation, the expected reproductive success is obtained by averaging over the sex of the offspring, and is $\frac{1}{2} + \gamma_F$. Thus, unless $\gamma_M = \gamma_F = 1/2$, the reproductive success of an X-linked allele will be negatively correlated between parents and offspring. Thus, the assumption of the haploid model that the reproductive success of individuals and their offspring are independent variables is clearly violated in this case.

3. Mutational process

Here we describe the assumptions on the mutational model and derive formulas for the expected levels of heterozygosity. To incorporate what has recently been revealed about the dependencies of mutation rates on sex and age (e.g., (6-8)), we allow for mutation rate in the diploid model to depend on sex and age. Namely, we assume that the number of de novo mutations that a parent of sex *s* and age *a* bequeaths to its newborn is a random variable with expectation $\mu_{s,a}$ per base pair. While in the main text, we consider a specific model for $\mu_{s,a}$ motivated by pedigree studies in humans, our derivation here treat $(\mu_{s,a})_{a=1}^{\infty}$ as parameters and assumes no specific form. Since mutation rates vary with sex and age, the mutation rates per generation in males and females depend on the distributions of their breeding ages (i.e. A_M and A_F, which were defined in Section 2). We denote the expected mutation rate per generation in males by $\mu_M = E_{A_M}(\mu_{M,a}) = \sum_a p_{M,a} \cdot \mu_{M,a}$ and the expected rate in females by $\mu_F = E_{A_F}(\mu_{F,a})$. The average rates on the autosomes and the X are given by $\mu_A = \frac{1}{2}(\mu_M + \mu_F)$ and $\mu_X = \frac{2}{3}\mu_F + \frac{1}{3}\mu_M$. For the haploid model, we assume the expected number of mutations μ_a to be dependent of age and define the per generation rate as $\mu = E_A(\mu_a)$. In the special case in which the parameters $\mu_{s,a}$ (or the μ_a in the haploid case) depend linearly on age, these expectations will depend only on the expected generation times G_M and G_F, i.e., they are insensitive to higher moments of the distributions of breeding ages in males and females. As we show below, higher moments of the distributions of mutation rates per generation do not affect our results, which is how we avoid any further assumptions about these distributions.

The standard expressions for heterozygosity (e.g., $E(\pi_A) = 4N_e^A\mu_A$) are usually derived assuming that the genealogical and mutational processes are independent (9). This assumption is violated in our case, because both the time to the most recent common ancestor and the number of accumulated mutations depend on the ages of the individuals along the lineage. To derive the expected autosomal heterozygosity $E(\pi_A)$ under these conditions, we track alleles *A* and *B* backwards in time. Let X_i denote the number of mutations occurring on the lineage leading from allele *A* in the i^{th} generation and *T* denote the number of generations until the alleles coalesce. The number of mutations on the lineage leading to allele *A* is then $\sum_{i=1}^{T} X_i$. Although X_i and *T* are dependent variables, Wald's equation (10) implies that $E(\sum_{i=1}^{T} X_i) = E(T) \cdot E(X_i)$ (to see that Wald's equation holds, note that the indicator function $\mathbb{I}_{T \geq n}$ is independent of X_n, since the first depends on the sexes and ages in the first $n-1$ generations, and the second on the n^{th} generation). We have shown previously that $E(T) = 2N_e^A$. Since $E(X_i|s_i, a_i) = \mu_{s,a}$ (where s_i and a_i are

the sex and age in the i^{th} generation), it follows that $E(X_i) = E(\mu_{s,a}) = \mu_A$. We conclude that the lineage leading to allele A has on average $E(\sum_{i=1}^{T} X_i) = 2N_e^A \cdot \mu_A$ mutations and therefore $E(\pi_A) = 4N_e^A \mu_A$. A similar argument shows that for the haploid model $E(\pi) = 2N_e\mu$.

The same argument cannot be readily applied to the X-chromosome, as the sexes s_i and s_{i+1} in consecutive generations along the lineage are dependent variables, leading to a dependence between X_{i+1} and s_i, in violation of the conditions for Wald's equation to hold. Instead, we define T as the number of females on the lineage until the coalescence occurs, and define X_i as the number of mutations between the i^{th} and $i+1$ females on the lineage. Under this definition, Wald's equation holds and $E(\pi_X) = 2E(\sum_{i=1}^{T} X_i) = 2E(X_i)E(T)$. It is easily shown that $E(X_i) = \frac{3}{2}\mu_X$ and $E(T) = N_e^X$, so that $E(\pi_X) = 3N_e^X \mu_X$.

Bibliography

1. Hill WG (1972) Effective size of populations with overlapping generations. *Theor. Popul. Biol.* 3(3):278-289.
2. Felsenstein J (1971) Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* 68(4):581-597.
3. Sagitov S & Jagers P (2005) The coalescent effective size of age-structured populations. *Annals of Applied Probability* 15(3):1778-1797.
4. Wright S (1939) Statistical genetics in relation to evolution, Vol. 802. *Hermann et Cie.: Paris*.
5. Pollak E (2011) Coalescent theory for age-structured random mating populations with two sexes. *Math. Biosci.* 233(2):126-134.
6. Kong A*, et al.* (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488(7412):471-475.
7. Segurel L, Wyman MJ, & Przeworski M (2014) Determinants of mutation rate variation in the human germline. *Annu Rev Genomics Hum Genet* 15:47-70.
8. Wong WS*, et al.* (2016) New observations on maternal age effect on germline de novo mutations. *Nat Commun* 7:10486.
9. Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology* 7(1):44.
10. Blackwell D (1946) On an Equation of Wald. *Annals of Mathematical Statistics* 17(1):84-87.