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

Guy Amster [a,1] & Guy Sella [a,1]

[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

Abstract

In human and other hominid (great apes) populations, estimates of the relative levels of neutral polymorphism on the X and autosomes differ from each other and from the naive theoretical expectation of ¾. These differences have garnered considerable attention over the past decade, with studies highlighting the potential importance of several factors, including historical changes in population size and linked selection near genes. Here, we examine a more realistic neutral model than has been considered to date, which incorporates sex- and age-dependent mortalities, fecundities, reproductive variances and mutation rates, and ask whether such a model can account for diversity levels observed far from genes. To this end, we derive analytical expressions for the X to autosome ratio of polymorphism levels, which incorporate all of these factors and clarify their effects. In particular, our model shows that the genealogical effects of life history can be reduced to ratios of sex-specific generation times and reproductive variances. Applying our results to hominids by relying on estimated life-history parameters and approximate relationships of mutation rates to age and sex, we find that life history effects, and the effects of male and female generation times in particular, may account for much of the observed variation in X to autosome ratios of polymorphism levels across populations and species.

Introduction

Neutral polymorphism patterns on X (or Z) and autosomes reflect a combination of evolutionary forces. Everything else being equal, the ratio of X to autosome polymorphism levels should be ¾, because the number of X-chromosomes in a population is ¾ that of autosomes. However, autosomes spend an equal number of generations in diploid form in both sexes, whereas the X spends two thirds of the number of generations in diploid form in females and a third in haploid form in males. As a result, the ratio of X to autosome polymorphism levels can also be shaped by differences in male and female life history and mutation processes as well as by differences in the effects of linked selection (1).

The differential effects of these factors on X and autosome has been discussed for many species (2-4), and in particular, they have been the focus of considerable recent interest in humans and other hominids (5-13). In particular, it has been noted that linked selection could affect X and autosomes differently because of differences in recombination rates, the density of selected regions, and the efficacy and mode of selection. Notably, the hemizygosity of the X in males leads to a more rapid fixation of recessive beneficial alleles and to a more rapid purging of recessive deleterious one (14, 15). Accounting for these effects and for differences in recombination rates suggests that in humans—in mammals more generally—the effects of linked selection should be stronger on the X ((16), but see (2)). To evaluate these effects empirically, several studies have examined how polymorphism levels on the X and autosomes vary with genetic distance from putatively selected regions, e.g., from coding and conserved non-coding regions (17, 7, 9, 18, 10, 11). In most hominids, including humans, such comparisons confirm that linked selection reduces X to autosome ratios. They further suggest that the effects are minimal sufficiently far from genes (17, 7), potentially allowing the effects of other factors shaping X to autosome ratios to be examined in isolation.

Even far from genes, however, the X to autosome ratios in humans and other hominids appear to differ from the naive expectation (7, 9, 10). To control for the effects of higher mutation rates in males and variation in mutation rates along the genome, polymorphism levels on X and autosomes are typically divided by divergence to an outgroup. In regions far from genes, the normalized X to autosome ratios range between ¾ and 1 among human populations, generally decreasing with the distance from Africa (5, 7, 9, 11). Ratios exceeding ¾ have also been observed in most other hominids ((10), but see (13)).

Proposed explanations for these departures from ¾ and the variation among populations and species fall into two main categories: those based on population history, such as historical changes in population size, and those based on life history, such as differences between male and female reproductive variances (i.e., the variance in the number of offspring that individuals have throughout their life). If we assume that the effective population size on the X is generally smaller than on autosomes, then changes in population size will have a different impact on polymorphism levels on X and autosomes (19-22). Notably, population bottlenecks that occurred sufficiently recently, such as the Out of Africa bottleneck in human evolution, will have decreased the X to autosome ratio, because a greater proportion of X-linked loci will have coalesced during the bottleneck (22). While what is known of human population history (including the Out of Africa bottleneck) is indeed expected to lead to lower X to autosome ratios in non-Africans, estimates of the expected reduction in diversity levels fall short of explaining its full extent (11). Differences in population history between males and females may have also contributed to the differences among human populations. For example, Keinan and Reich (8) suggested that an increase in the ratio of males to female effective population sizes (or generation times) during an Out-of-Africa bottleneck contributed to the lower X to autosome ratios in non-Africans. While the possibility of inferring such historical differences from extant polymorphism patterns is intriguing, and there are comparable, well established sex-differences in the more recent past (e.g., (23)), differences in life history may offer a more straight-forward explanation of the observations.

Notably, sex differences in life history traits that are readily observed today could have substantially affected X to autosome ratios. Among traits, perhaps the most straightforward effect arises from the higher reproductive variance in males than in females (e.g., due to sexual selection), which causes higher coalescence rates on autosomes and an increased X to autosome ratio (24, 5). Theoretical studies suggest that the resulting increased ratio is bound by $^9/_8$ (24). In reality the effect is likely much smaller, yet differences in male and female variance among extant human populations and hominids (25) suggest that it could have contributed to observed differences in X to autosome ratios, as well as account for why ratios often exceed $^3/_4$.

Differences between male and female generation times could have affected X to autosome ratios in more intricate but substantial ways. Mutation rates in mammals increase with paternal age (and to a lesser extent with maternal age), notably in humans (26-29). Ignoring genealogical effects (see below), we would therefore expect increased paternal generation times to decrease the X to autosome ratios of polymorphism and divergence. In that regard, given that male and female generation times vary considerably among populations and species (30, 31), the common practice of dividing polymorphism estimates by divergence levels might not fully control for male mutational biases and could result in X to autosomes ratios that deviate from ¾. Sex differences in generation times also affect the genealogical process, with longer generation times in males than in females resulting in slower coalescence rates on the X compared to autosomes, and therefore in increased polymorphism ratios.

The overall effect of these factors has not been considered jointly. We do so here, by studying how these life history traits and mutational differences between sexes affect neutral polymorphism levels on X and autosomes. Earlier work by Felsenstein (32) modeled the effects of age structure on the effective population size of haploid populations, and that work was later extended by Hill (33), Johnson (34), Charlesworth (35, 24) and Orive (36), who incorporated the effects of age structure in diploid populations, as well as the effects of sex- (but not age-) dependent rates of mutation. Where our work departs from those is in building on coalescent models of age-structured populations developed by Sagitov and Jagers (37) and Pollack (38), which we describe below.

Results

The haploid model. To illustrate how we treat the coalescence process in an age-structured population, we first consider the haploid model proposed by Felsenstein (32). We assume a haploid panmictic population that is divided into age classes of one year (for convenience), and denote the (constant) number of individuals of age $a \geq 1$ by M_a, where $M_{a+1} \leq M_a$ due to mortality (see Table 1 for a summary of notations). We further assume that each of the M_1 newborns is independently descended from a random parent, with the parent's age chosen from a distribution $A = (p_a)_{a=1}^{\infty}$ with expectation G (the generation time). The parent is then chosen with uniform probability within the age class. We begin with the case without endogenous reproductive variance, but incorporate this effect below.

Notation	Definition
p_a	Probability that a newborn descends from a parent of age a
M_a	Number of individuals of age a
G	Average generation time
M	Effective age-class size
$\vec{r}$	Relative reproductive success, where component r_a is the relative reproductive success at age a
$W_{i,j}$	Expected value of $r_i \cdot r_j$ conditioned on survival to age j $(j \geq i)$
W	Weighted average of $W_{i,j}$

Table 1: Notation for the haploid model.

This model was solved by Sagitov and Jagers (37) in a coalescent framework, and here we provide an intuitive account of the solution. First, we consider the rate of coalescence for a sample of two alleles, where to this end, we trace their lineages backward in time. For the alleles to coalesce in a given age class a, one of them would have to be a newborn in the previous generation; this occurs with probability p_a/G per year. The other allele would have to be present in the same age class, which occurs with probability $\sum_{j\geq a} p_a/G$, because it could have been born to a parent of age $j \geq a$, $j - a + 1$ generations ago (the rigorous derivation establishes that this is in fact the stationary age distribution along a lineage). Both alleles would also have to be in the same individual in age class a, which occurs with probability $1/M_a$. Taken together, we find that the probability of coalescence per-year in age class a is

$$2 \cdot \left(\frac{p_a}{G}\right) \cdot \left(\frac{\sum_{j\geq a} p_j}{G}\right)\left(\frac{1}{M_a}\right) - \left(\frac{p_a}{G}\right)^2 \left(\frac{1}{M_a}\right),$$

where we multiplied by 2, because either allele could have been the newborn in the previous generation, but subtracted the probability that they both were, because this event should be counted only once. The probability of coalescence per generation and corresponding effective population size follow from multiplying these probabilities by the generation time, summing them over age classes, and rearranging terms:

$$\frac{1}{N_e} = \frac{1}{G}\sum_a \frac{w_a}{M_a}, \qquad (1)$$

where $w_a = p_a^2 + 2\sum_{j>a} p_a p_j$. Note that the (a, j)-term in w_a is proportional to the probability that coalescence in age class a occurs in an individual that fathered a newborn carrying one of the alleles at age a, and a newborn carrying the other allele at age j. Moreover, the w_a terms add up to one, allowing us to define the effective age class size as the weighted harmonic mean

$$\frac{1}{M} = \sum_a \frac{w_a}{M_a}. \tag{2}$$

The effective population size can then be viewed as the product of the effective age-class size and the effective number of age-classes, which is simply the expected generation time *G*, i.e.,

$$N_e = G \cdot M. \tag{3}$$

Next, we extend the model to incorporate endogenous reproductive variance (as opposed to the variance introduced by stochastic mortality and birth). To this end, we assume that each newborn is assigned a vector $\vec{r}$ describing its age-dependent, relative reproductive success, such that its probability of being chosen as a parent among the individuals of age class a is r_a/M_a (i.e., r_a corresponds to the expected, rather than actual, reproductive success of the individual at age a). We further assume that the proportion of individuals with a given vector $\vec{r}$ that reach age a, $f_a(\vec{r})$, can vary with age, in effect allowing for dependencies between age-specific mortality and reproductive success. Thus, the model is set up to allow for dependencies between reproductive success, fecundity and longevity, examples of which have been observed in natural populations (e.g., between the age of first reproduction and longevity (39, 40), or between reproductive success and longevity (39, 41)).

The model thus extended can be solved along the same lines as described above (see SI Section 1.2). Specifically, the coalescence rate per generation and corresponding effective population size take a similar form:

$$\frac{1}{N_e} = \frac{1}{G}\sum_a \frac{w_a}{M_a}, \tag{4}$$

but in this case

$$w_a = p_a^2 W_{a,a} + 2\sum_{j>a} p_a p_j W_{a,j}, \tag{5}$$

where $W_{a,j} \equiv E_{\vec{r}}\left(r_a \cdot r_j \middle| \text{individual reaches age} \geq j\right)$ for $j \geq a$. As in the simpler case, the (a, j)-term in w_a is proportional to the probability that coalescence in age class a occurs in an individual that fathered a newborn carrying one of the alleles at age a, and a newborn carrying the other allele at age j; but in this case, the coefficients $W_{a,j}$ factor in the effect of endogenous reproductive variance. In contrast to the simpler case, the w_a terms do not necessarily add up to 1. We therefore introduce a normalization by $W = \sum_a w_a$, and define the effective age class size as

$$\frac{1}{M} = \frac{1}{G}\sum_a \frac{w_a/W}{M_a}. \tag{6}$$

In these terms, the effective population size takes the form

$$N_e = G \cdot M / W. \tag{7}$$

To provide some intuition, we consider the special case in which relative reproductive success is independent of age and of mortality rates. Namely, each newborn is assigned a relative reproductive success r at birth, and its probability of being chosen as a parent among the individuals of age class a is r/M_a. The distribution of r values then has expectation 1 (by construction) and we denote its variance by V_r ($V_r = 0$ when there is no endogenous reproductive variance). In this case, the coalescence rates in any given age class are increased by a factor $1 + V_r$, and therefore the effective population size is

$$N_e = \frac{1}{1+V_r} G \cdot M. \tag{8}$$

Thus, $W = 1 + V_r$ and it can be interpreted as the reproductive variance caused by the Poisson sampling of parents, which contributes 1, and by the endogenous reproductive variance, which contributes V_r. In turn, the contribution of age-structure to reproductive variance is reflected in $G \cdot M$.

More generally, the results of the full model can be recast in terms of the total reproductive variance. First consider a haploid Wright-Fisher process, i.e., with non-overlapping generations, with endogenous reproductive variance, modeled similarly to the example considered above. In this case, Wright (42) showed that the effective population size is

$$N_e = N/V, \tag{9}$$

where N is the census population size and $V = 1 + V_r$ is the total reproductive variance. Second, in the case with age-structure but without endogenous reproductive variance, Hill showed that the effective population size can also be written as

$$N_e = G \cdot M_1 / V, \tag{10}$$

where $G \cdot M_1$ is the number of newborns per generation, and V is the reproductive variance introduced by age-structure (33). Comparing Eqs. 7 and 10, we can express this variance as $V = M_1/M$ (≥ 1). This expression makes intuitive sense, as this would be the reproductive variance in the case considered by Wright, if the next generation were randomly chosen from a reproductive pool including only M out of M_1 individuals in the population. In SI Section 1.3, we show that Eq. 10 also holds for the general model with age-structure and endogenous reproductive variance. In this case, the total reproductive variance is

$$V = (M_1/M) \cdot W, \tag{11}$$

where the first term in this product, M_1/M, is the variance introduced by stochasticity in mortality and birth, and the second term, W, reflects the contribution of endogenous reproductive variance. Thus, we conclude (i.e., in Eq. 10) that all the effects of age-structure and endogenous reproductive variance (as well as any dependence between them) on the effective population size can be summarized in terms of the generation time, G, number of newborns per year, M_1, and total reproductive variance, V.

The model for X and autosomes. The diploid model with two sexes is more elaborate, but it is defined and solved along the same lines that we have described for the haploid model. In SI Section 2.2, we describe the model formally, and here, for brevity, we only describe how it differs from the haploid model. Notably, in this case, we allow for age-dependent mortality, fecundity, and endogenous reproductive variance to vary between the sexes. We further assume equal sex ratios at birth (although we also consider the general case in SI Section 2), and accommodate X and autosomal modes of inheritance (i.e., X linked loci in males always descend from females, whereas in all other cases the sex of parents is randomly chosen with probability ½). In SI Section 2.3, we solve this model for the stationary distributions of sex and age and corresponding coalescence rates, by extending Pollack's results for age-structured populations with two sexes (38) to account for endogenous reproductive variance. Here we describe the main results relying on the intuition gained from the haploid model.

By analogy with the haploid case (Eq. 10), we can express the effective population sizes for X and autosomes in terms of the reproductive variance of males and females. To this end, we denote the number of newborns per year, including both sexes, by M_1 (see Table 2 for a summary of notations). We further define the generation times for X and autosome linked loci as averages of the male and female generation times, G_M and G_F, weighted by the proportions of generations that they spend in each sex, i.e.,

$$G_X = \frac{2}{3}G_F + \frac{1}{3}G_M \text{ and } G_A = \frac{1}{2}(G_M + G_F). \tag{12}$$

To draw the analogy with the haploid case, we consider the effective population size in terms of alleles rather than individuals (they are equivalent for haploids). Specifically, we define the reproductive success of an allele as the number of an individual's offspring carrying that allele, and denote the reproductive variance associated with X and autosome linked alleles by V_X^* and V_A^*. In these terms, the effective population sizes for X and autosomes are analogous to Eq. 10:

$$\frac{3}{2}\cdot N_e^X = G_X\left(\frac{3}{2}\cdot M_1\right)/V_X^* \text{ and } 2\cdot N_e^A = G_A(2\cdot M_1)/V_A^*, \tag{13}$$

where the factor of 3/2 for X and 2 for autosomes follow from considering the effective size of alleles rather than individuals on the left-hand side, and the number of newborn alleles rather than individuals on the right-hand side. Expressing these results in terms of the reproductive variance of males, V_M, and of females, V_F, in SI Section 2.5, we show that

$$V_X^* = \frac{1}{4}\left(2 + \frac{1}{3}V_M + \frac{2}{3}V_F\right) \text{ and } V_A^* = \frac{1}{4}\left(2 + \frac{1}{2}V_M + \frac{1}{2}V_F\right), \tag{14}$$

where the weights reflect the proportion of generations spent in males and females and the additive factor 2 results from ploidy. Substituting these expressions into Eq. 13 we find that

$$N_e^X = \frac{4G_X M_1}{2+\frac{1}{3}V_M+\frac{2}{3}V_F} \text{ and } N_e^A = \frac{4G_A M_1}{2+\frac{1}{2}V_M+\frac{1}{2}V_F}. \tag{15}$$

Notation	Definition
M_1	Number of male and female newborns per-year
G_M, G_F	Average generation times in males and females
G_X, G_A	Average generation times for X and autosomes
V_X^*, V_A^*	Reproductive variance for X and autosome linked alleles
V_M, V_F	Reproductive variance of males and females
μ_M, μ_F	Average mutation rate per generation in males and females
μ_X, μ_A	Average mutation rate on the X and autosomes

Table 2: Notation for the diploid model with two sexes.

Polymorphism levels on X and autosomes. Having solved for the effective population sizes, we now turn to polymorphism levels. To this end, we assume that the expected mutation rates in males and females depend linearly on age (27, 43), and that their expectations for generation times G_M and G_F are μ_M and μ_F respectively (see SI Section 3 for general dependency). The expected mutation rates per generation on X and autosome linked lineages are weighted averages of these expected rates,

$$\mu_A = \frac{1}{2}(\mu_M + \mu_F) \text{ and } \mu_X = \frac{2}{3}\mu_F + \frac{1}{3}\mu_M. \tag{16}$$

The expected heterozygosity on X and autosomes then follows from the standard forms:

$$E(\pi_A) = 4N_e^A \mu_A \text{ and } E(\pi_X) = 3N_e^X \mu_X. \tag{17}$$

Note that these expressions are usually derived assuming that the genealogical and mutational processes are independent (44). This assumption is violated here, because both the coalescence and mutation rates depend on the ages along a lineage. In SI Section 3, we show that the standard forms hold nonetheless.

We can now combine our results to derive an expression for the X to autosome ratio of polymorphism levels. From Eqs. 15 and 17 and rearranging terms, we find that the ratio of expected heterozygosity is

$$\frac{E(\pi_X)}{E(\pi_A)} = \frac{3}{4} \cdot \frac{f(\mu_M/\mu_F) \cdot f(G_M/G_F)}{f\left(\frac{2+V_M}{2+V_F}\right)}, \qquad (18)$$

where $f(x) = \frac{2x+4}{3x+3}$ (see Eq. S86 for the case in which the sex ratio at birth differs from 1). When the mutation rates, generation times, and reproductive variance are identical in both sexes, this expression reduces to the naïve neutral expectation of ¾. When these factors differ between sexes, Eq. 18 provides a simple expression for the effect of each factor. Notably, the effects of age-structure and endogenous reproductive variance reduce to effects of male to female ratios of generation times and of reproductive variances (more precisely $(V_M + 2)/(V_F + 2)$).

Comparing ratios of polymorphism and divergence. We can now ask how changes in these factors affect X to autosome ratios of polymorphism and divergence. More precisely, considering the numbers of substitutions rather than divergence (i.e., ignoring multiple hits) and excluding the contribution of ancestral, the equivalent expression to Eq. 18 is

$$\frac{E(K_X)}{E(K_A)} = \frac{f(\mu_M/\mu_F)}{f(G_M/G_F)} \qquad (19)$$

(35, 45). While a higher ratio of male to female mutation rates, $\alpha = \mu_M/\mu_F$, has the same effect on X to autosome ratios of polymorphism and divergence, a higher ratio of male to female generation times (G_M/G_F) has opposite effects, decreasing the ratio for polymorphism but increasing it for divergence (Fig. 1). The difference arises because one way in which the male to female ratio of generation times affects polymorphism levels is by changing the relative rates of coalescence on X and autosomes, whereas there is no such effect on divergence levels, for which coalescence times on both X and autosomes equal the species split time (again, neglecting the contribution of ancestral polymorphism). By the same token, reproductive variances affect only the coalescence rates, and thus polymorphism but not divergence levels. Importantly, the different effects of life history factors on polymorphism and divergence complicate the interpretation of X to autosome polymorphism ratios that are divided by divergence to an outgroup (e.g., (5, 7, 9, 46)) (see Discussion).

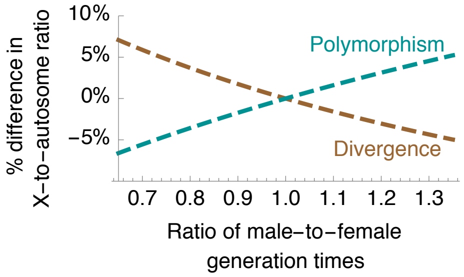

Figure 1: Generation time effects on X to autosome divergence and polymorphism ratios.

X to autosome polymorphism ratios in humans and other hominids. Thus far, we considered general results for the effects of life history on the X to autosome ratio of polymorphism levels, and now we turn to the effects in human populations and in other hominid species. To this end, we rely on a model that approximates the dependence of hominid mutation rates on sex and age (26, 43). We assume that mutations accumulate linearly with the number of germ-cell divisions, and that the rate per division varies at different stages of development (43). In females, the number of cell divisions does not depend on age because all oogonial mitotic divisions occur before birth (47). In males, germ-cell division exhibits two main phases, with an approximately constant number until puberty, followed by a linear rate of division in adult testis during spermatogenesis (48, 49). We previously (45) fitted the corresponding mutation model to the relationship between the number of mutations and the parents' sex and age observed in human pedigrees (27), to obtain the following approximations for female and male mutation rates per bp per generation:

$$\mu_F = 5.42 \cdot 10^{-9} \text{ and } \mu_M = 6.13 \cdot 10^{-9} + 3.33 \cdot 10^{-10}(G_M - P)/\tau, \qquad (20)$$

where P is the puberty age in males and τ is the duration of the seminiferous epithelial cycle (during spermatogenesis). Recent pedigree studies indicate that maternal age also affect mutation rates (e.g. (29)), more generally suggesting that some mutations are in fact spontaneous rather than replication driven (43, 50). At present, however, the studies are not sufficiently precise for us to incorporate maternal age and spontaneous effects in our mutational model. Once they are, it should be straight forward to study how they affect the results that we report below.

Given our mutational model and relying on prior knowledge about life history parameters, we consider how each of the factors detailed in Eq. 18 affects deviation of the X to autosome ratio from the naïve expectation of ¾, and how they affect variation in the ratio among populations and species. We begin with the effect of sex differences in reproductive variance (Fig. 2). This effect has been noted by many (e.g., (24, 51)) and follows intuitively from the fact that a higher male variance results in higher coalescence rates on autosomes than on the X, thus increasing π_X/π_A. In theory, we know that higher reproductive variance in males can increase π_X/π_A by as much as 50% (with an upper bound of $^9/_8$ compared to $^3/_4$), but in practice, the effect is expected to be much smaller. Consider, for example, a model in which only a fraction p of the males can reproduce and the number of offspring for each follows a Poisson distribution with mean $2/p$, while all females reproduce with equal probabilities following a Poisson with mean 2, resulting in a ratio $(V_M + 2)/(V_F + 2) = p^{-1}$. In this case, assuming that only $p = 0.3$ of males can reproduce would result in an X to autosome polymorphism ratio that exceeds ¾ by only 22% (i.e., much lower than the theoretical limit). Unfortunately, little is known about the plausible range of reproductive variances in humans. In five hunter-gatherer groups in which this variance was measured in small samples, it was found to be 1.7-4.2 folds higher in males (25), corresponding to an increase of 6-20% relative to ¾, which translates into a relative difference of ~13% among populations.

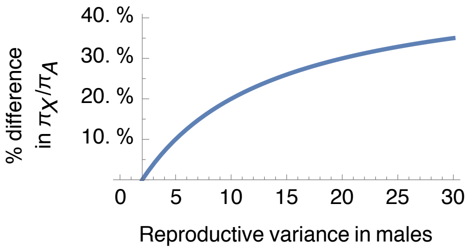

Figure 2: The effect of male reproductive variance on the ratio of X to autosome polymorphism levels, assuming the female reproductive variance equals 2.

Next, we consider the effect of the ratio of male to female generation times (Fig. 3A). In seven hunter-gatherer groups in which these were measured, the mean generation times vary between 25 and 33 years and the ratio of male to female generation times between 1.03 and 1.37 (30, 45). Accounting for both mutational and genealogical effects, this variation corresponds to an 18%-24% reduction, and an ~8% relative difference in X to autosome polymorphism ratios. Intriguingly, the relative effect is on the order of the variation in observed X to autosome ratios among human populations, which are 6%-9% lower in Europeans and 8%-9% lower in Asians compared to Africans (11). However, the effect of generation times on the X to autosome polymorphism ratio should be averaged over the period that gave rise to extant polymorphism levels. Thus, while these estimates suggest that higher ratios of male to female generation times in non-Africans may have contributed substantially to observed differences, a more precise account of this contribution should factor in the shared evolutionary history among populations, as well as subsequent lineage-specific variation in generation times.

Differences in ratios of male to female generation times, have also been reported among groups of wild chimpanzees, albeit based on small samples (31). Specifically, mean generation times have been estimated to vary between 22 and 27 years and the ratio of male to female generation times between 0.8 and 1.2, corresponding to a 14%-24% reduction and a ~13% relative difference in X to autosome polymorphism ratios. Assuming such differences in generation times have persisted during the millions of years over which neutral genetic variation accumulated, they would have therefore contributed substantially to the variation in polymorphism ratios among hominid populations and species.

As we already noted, paternal and maternal generation times affect the X to autosome ratios of polymorphism and divergence differently. Based on our mutational model, we can now ask about the combined mutational and genealogical effect of paternal and maternal generation times in hominids. Changing the paternal or maternal generation times has a genealogical effect of the same magnitude (but not direction; see Eq. 18), but changing the paternal generation time has a greater effect on the mutation rate per generation than changing the maternal one. Thus, overall, changes in paternal generation time have a greater effect on the X to autosome polymorphism ratio than the maternal one (Fig. 3A). In contrast, as we have shown previously ((45)), changes in maternal generation times have a greater effect on the divergence ratio (Fig. 3B), because the paternal age has opposing effects on the mutation rate per generation and on the number of generations along a lineage.

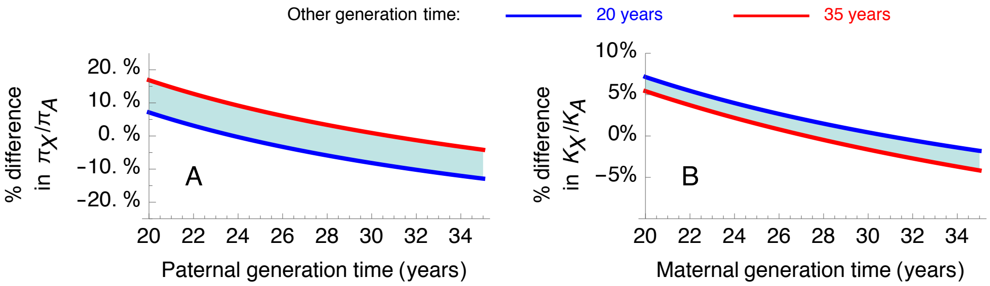

Figure 3: The effects of maternal and paternal generation times on X to autosome ratios of polymorphism (A) and divergence (B) in humans. Note that the range on the y-axis differs between (A) and (B), reflecting the added genealogical effect on polymorphism but not on divergence.

Lastly, the degree to which mutation is male-biased depends on the age of puberty in males and on the seminiferous epithelial cycle length, suggesting that changes to these parameters would also affect the X to autosome polymorphism ratio (Fig. 4). For example, varying the age of puberty between 13.5 years, the value estimated for humans (52), and 7.5 years, the value estimated for chimpanzees (53), while holding other life history parameter at their values in humans, decreases X to autosome polymorphism ratios by 3%. Similarly, varying the seminiferous epithelial cycle length from 16 days (humans (54)) to 14 days (chimpanzees (55)) decreases polymorphism ratios by 1%. Differences in these parameters between humans and more remotely related species may be greater, resulting in a more substantial effect on polymorphism and divergence ratios (45). For example, using ages of puberty in cercopithecoids, estimated to be between 3.5-6 years (56), leads to a reduction of 3-4% in polymorphism ratios compared to humans, and assuming the seminiferous epithelial cycle lengths in cercopithecoids is between 9.5-11.6 days (57, 58) leads to a reduction of 3-5%.

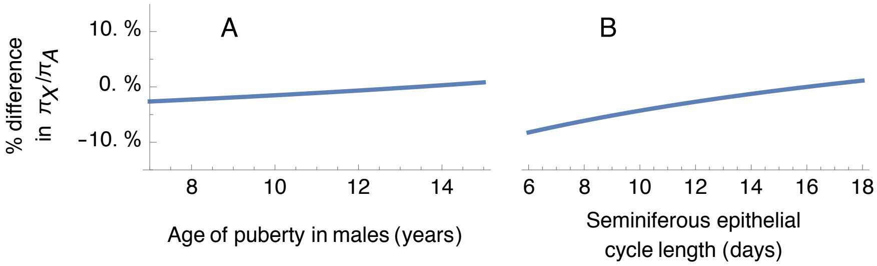

Figure 4: The effects of male puberty age (A) and seminiferous epithelial cycle length (B) on the polymorphism ratio. Other parameters are fixed for their estimates in humans, i.e., generation times of $G_M = 33.8$ and $G_F = 26.9$ years (estimated in hunter-gatherers (30, 45)), male age of puberty of $P = 13.5$ years (52), and seminiferous epithelial cycle length of $\tau = 16/365$ years (54).

Discussion

We have derived general expressions for the effects of life history factors on X and autosome polymorphism levels, and explored their plausible effects on levels observed in humans and closely related species. The expression for the ratio (Eq. 18) clarifies the effect of each factor, and in particular, shows that the genealogical effect of life history factors reduces to the effect of the ratios of male to female generation times, G_M/G_F, and of reproductive variances, $(V_M + 2)/(V_F + 2)$. These results apply to any species with X and autosomes, and can be readily generalized to species with Z and autosomes. We focused on hominid species, because more is known about the factors affecting their polymorphism levels. Specifically, by considering a hominid model for the effects of life history on mutation rates, we have shown that plausible variation in G_M/G_F within human populations and among closely related species can have a substantial effect on levels of polymorphism on the X compared to the autosomes. It therefore seems plausible that evolutionary changes in life history contributed substantially to the variation in X to autosome ratios among human populations and hominid species, and to their seemingly ubiquitous deviation from the naïve expectation of ¾.

This generation time effect acts jointly with myriad other factors that have been highlighted previously (e.g., 24, 5, 8, 11). Notably, the ratio of male to female reproductive variances, $(V_M + 2)/(V_F + 2)$, which tends to be greater than 1 and varies considerably among human populations and closely related species, is likely to have been an important explanatory factor (24, 5). Another likely contributor is the different responses of polymorphism levels on X and autosomes to changes in population size, and to bottlenecks in particular (22). Notably, changes in population size (most notably, the Out-of-Africa bottleneck) have been estimated to contribute 36%-47% of the observed reduction in X to autosome ratios in Asians and 38%-60% in Europeans compared to African populations, where the ranges correspond primarily to the use of different demographic models (11). Importantly, however, these estimates assume that the X to autosomes ratio of effective population sizes in the absence of bottlenecks is ¾, while consideration of sex-specific generation times and reproductive variances point to a ratio closer to 1, and thus suggest a weaker effect of population size changes. Finally, X to autosome ratios have been shown to vary substantially along the genome due to variation in the effects of linked selection, e.g., with an estimated average reduction of 25-35% near compared to far from genes (11). Taken together, these factors are likely to explain most of the variation in ratios observed among human populations and along the genome. In considering variation among hominid species, differences in other factors affecting the male bias of mutation rates, $\alpha = \mu_M/\mu_F$, such as the male age of puberty and the length of the spermatogenetic cycle (43), may have also contributed substantially.

Ultimately, we would like to quantify the contributions of each factor to observed X to autosome polymorphism ratios. At present, however, a number of practical obstacles stand in the way. Most importantly, despite numerous papers characterizing X to autosome ratios, we still lack reliable estimates of their absolute values, and thus of their deviations from the naïve expectation of ¾. The main reason being that current estimates are based on dividing polymorphism levels by divergence to an outgroup (e.g., to macaques (7, 9), chimpanzees (5, 7), orangutans (5, 7), gorillas (7), marmosets (7), olive baboons (46) or, when non-human species are considered, to humans (46)), in an attempt to control for differences in mutation rates on X and autosomes. The "normalized" ratios are assumed to reflect only genealogical differences between X and autosomes, but this assumption is wrong. Notably, changes to life history traits should lead to changes in the X to autosome ratio of divergence over evolutionary time (45), and indeed this ratio has been found to vary substantially with the choice of outgroup (59, 60).

Our results make the problematic nature of this normalization procedure quite clear, by allowing us to write down an explicit form for the normalized ratio:

$$\frac{\pi_X/K_X}{\pi_A/K_A} = \left[\frac{f(\mu_M/\mu_F)}{f(\mu_M^*/\mu_F^*)} \cdot f(G_M^*/G_F^*)\right] \cdot \frac{3}{4} \frac{f(G_M/G_F)}{f((2+V_M)/(2+V_F))}, \tag{21}$$

where parameters corresponding to the outgroup lineage are denoted by "*". The second term (on the right-hand side) in this expression reflects the genealogical component of the X to autosome ratio, and will henceforth be referred to as "the genealogical ratio", whereas the first (in brackets) includes the mutational effect on the ratio and the terms introduced by the normalization. For the normalization to fulfill its purpose of canceling out the mutational effects, the first term must equal 1. As noted, however, the male bias in mutation rates ($\alpha = \mu_M/\mu_F$) has evolved substantially over phylogenetic time scales (59, 61), and therefore the mutational terms $f(\mu_M/\mu_F)/f(\mu_M^*/\mu_F^*)$ do not cancel out. Even if they did, the different dependencies of polymorphism and divergence ratios on the ratio of male to female generation times introduces the additional term $f(G_M^*/G_F^*)$ into the "normalized" polymorphism ratios. If we assume that generation times in males are greater than in females, as observed in extant hunter-gatherers (30), gorillas and some chimpanzee groups (31), this term would introduce a downward bias (e.g., up to 6%, for estimated generation times in different hunter-gatherer groups). The combined effect of male mutation bias and generation times on the normalization in hominoids, i.e., the term $f(G_M^*/G_F^*)/f(\mu_M^*/\mu_F^*)$, is dominated by the maternal generation time G_F^* (45). Varying G_F^* along the lineage to the outgroup from 18 to 30, while fixing other parameters to human values, results in a 10% increase in the normalized ratio. Taken together, these factors make existing, "normalized" estimates of the X to autosome ratio of polymorphism levels (e.g., (11)) difficult to interpret. Specifically, they cannot be interpreted as estimates of the genealogical X to autosome ratios, and are thus not comparable with the naïve expectation of ¾.

These difficulties suggest that it may be preferable to consider X to autosome polymorphism ratios without normalization, and find alternative means to account for differences in the mutability of X and autosomes. We can begin by separating polymorphism ratios by types of mutation and genomic contexts, e.g., look at the ratio of C to T mutations in an ancestral ACA context (62). If we knew the sex specific mutation rates corresponding to these types and contexts during the genealogical history of extant samples, then we could divide out the mutational effects on these polymorphism ratios (i.e., $f(\mu_M/\mu_F)$) to obtain estimates of the genealogical X to autosome ratio. In principle, we can learn about these mutation rates from the extant mutational spectrum, and specifically from the dependency of the rates of different types of mutations on sex and life history parameters (see below). While we currently lack a sufficiently accurate characterization of the sex and age specific mutational spectrum to do so, it may very well be within reach in the near future, at least in humans, owing to the increasing power of pedigree studies of mutation (29). Meanwhile, focusing on specific categories of mutations, the rates of which have been fairly stable over phylogenetic timescales (e.g., CpG transitions and ACA->ATA (62, 63)), may provide initial approximations for the genealogical X to autosome ratio.

One can also apply new approaches to learn about the relative contribution of the different factors that shaped the genealogical ratio. Notably, as different types of mutations have distinct dependencies on life history parameters but X to autosome ratios corresponding to these types share the same genealogical history, one can make inferences about life history parameters (and the corresponding mutational spectrum) by requiring the same genealogical X to autosome ratio for different types of mutations (Gao, Moorjani et al., in preparation). Specifically, this approach may allow one to learn about the effects of the ratio of male to female generation times on the genealogical ratio. In addition, one could imagine using pairwise sequentially Markovian coalescent (PSMC) based inferences about historical effective population sizes on X and autosomes, to learn about the effects of bottlenecks, e.g., the OoA bottleneck, on the genealogical ratio directly rather than through simulations (cf. (64, 11)). Moreover, it may be possible to extend PSMC to use data from X and autosomes jointly in order to infer historical changes in female and male effective population sizes over time, and examine, for example, if they exhibited dramatic changes during the Out of Africa bottleneck (i.e., to examine the hypothesis of (8)).

More generally, all the factors that are thought to affect neutral polymorphism levels on X and autosomes can now be integrated into a coalescent framework, providing a natural way to quantify how their effects combine, and an efficient model from which to simulate. Our results describe how life history factors affect the effective population sizes on X and autosomes in the Kingman coalescent. The integration of other factors then follows from their known descriptions in terms of the coalescent, starting from these effective population sizes. For example, the effects of changes in population size and generation times can thus be integrated as appropriate changes to the effective population sizes on X and autosomes backwards in time. Moreover, the effects of linked selection on relative diversity levels along the genome also follow from their description in terms of the coalescent (e.g., (65-67)), starting from these effective population sizes. Lastly, the effects of life history on mutation can be incorporated in terms of mutational models akin to our hominid model. With the theoretical framework in place, and the data needed to infer the effects of each factor either already present or around the corner, a comprehensive, quantitative understanding of diversity levels on X and autosomes in hominoids and other taxa should be within reach.

Acknowledgements. We thank M. Przeworski for many helpful discussions and comments on the manuscript.

Bibliography

1. Webster TH & Wilson Sayres MA (2016) Genomic signatures of sex-biased demography: progress and prospects. *Curr. Opin. Genet. Dev.* 41:62-71.
2. Aquadro CF, Begun DJ, & Kindahl EC (1994) Selection, recombination, and DNA polymorphism in Drosophila. *Non-neutral evolution: Theories and Molecular Data*, ed B G (Springer, Dordrecht), pp 46-56.
3. Ellegren H (2009) The different levels of genetic diversity in sex chromosomes and autosomes. *Trends Genet.* 25(6):278-284.
4. Leffler EM*, et al.* (2012) Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* 10(9):e1001388.
5. Hammer MF, Mendez FL, Cox MP, Woerner AE, & Wall JD (2008) Sex-biased evolutionary forces shape genomic patterns of human diversity. *PLoS Genet* 4(9):e1000202.
6. Keinan A, Mullikin JC, Patterson N, & Reich D (2009) Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nat. Genet.* 41(1):66-70.
7. Hammer MF*, et al.* (2010) The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nat. Genet.* 42(10):830-831.
8. Keinan A & Reich D (2010) Can a sex-biased human demography account for the reduced effective population size of chromosome X in non-Africans? *Mol. Biol. Evol.* 27(10):2312-2321.
9. Gottipati S, Arbiza L, Siepel A, Clark AG, & Keinan A (2011) Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. *Nat. Genet.* 43(8):741-743.
10. Prado-Martinez J*, et al.* (2013) Great ape genetic diversity and population history. *Nature* 499(7459):471-475.
11. Arbiza L, Gottipati S, Siepel A, & Keinan A (2014) Contrasting X-linked and autosomal diversity across 14 human populations. *The American Journal of Human Genetics* 94(6):827-844.
12. Veeramah KR, Gutenkunst RN, Woerner AE, Watkins JC, & Hammer MF (2014) Evidence for increased levels of positive and negative selection on the X chromosome versus autosomes in humans. *Mol. Biol. Evol.* 31(9):2267-2282.
13. Nam K*, et al.* (2015) Extreme selective sweeps independently targeted the X chromosomes of the great apes. *Proceedings of the National Academy of Sciences* 112(20):6413-6418.
14. Haldane JBS (1924) A mathematical theory of natural and artificial selection. . *Trans. Camb. Philos. Soc.* Part I.(23):19-41.
15. Charlesworth B, Coyne JA, & Barton NH (1987) The Relative Rates of Evolution of Sex-Chromosomes and Autosomes. *American Naturalist* 130(1):113-146.

16. Charlesworth B (2012) The Effects of Deleterious Mutations on Evolution at Linked Sites. *Genetics* 190(1):5-22.
17. McVicker G, Gordon D, Davis C, & Green P (2009) Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* 5(5):e1000471.
18. Hernandez RD*, et al.* (2011) Classic Selective Sweeps Were Rare in Recent Human Evolution. *Science* 331(6019):920-924.
19. Fay JC & Wu CI (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol. Biol. Evol.* 16(7):1003-1005.
20. Hey J & Harris E (1999) Population bottlenecks and patterns of human polymorphism. *Mol. Biol. Evol.* 16(10):1423-1426.
21. Wall JD, Andolfatto P, & Przeworski M (2002) Testing models of selection and demography in Drosophila simulans. *Genetics* 162(1):203-216.
22. Pool JE & Nielsen R (2007) Population size changes reshape genomic patterns of diversity. *Evolution* 61(12):3001-3006.
23. Bryc K*, et al.* (2010) Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proceedings of the National Academy of Sciences* 107(Supplement 2):8954-8961.
24. Charlesworth B (2001) The effect of life-history and mode of inheritance on neutral genetic variability. *Genet. Res.* 77(2):153-166.
25. Betzig L (2012) Means, variances, and ranges in reproductive success: comparative evidence. *Evolution and Human Behavior* 33(4):309-317.
26. Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* 1(1):40-47.
27. 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.
28. Venn O*, et al.* (2014) Strong male bias drives germline mutation in chimpanzees. *Science* 344(6189):1272-1275.
29. Wong WS*, et al.* (2016) New observations on maternal age effect on germline de novo mutations. *Nat Commun* 7:10486.
30. Fenner JN (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* 128(2):415-423.
31. Langergraber KE*, et al.* (2012) Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci. U. S. A.* 109(39):15716-15721.
32. Felsenstein J (1971) Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* 68(4):581-597.
33. Hill WG (1972) Effective size of populations with overlapping generations. *Theor. Popul. Biol.* 3(3):278-289.
34. Johnson DL (1977) Inbreeding in populations with overlapping generations. *Genetics* 87(3):581-591.
35. Charlesworth B (1980) *Evolution in age-structured populations* (Cambridge University Press, Cambridge).
36. Orive ME (1993) Effective population size in organisms with complex life-histories. *Theor. Popul. Biol.* 44(3):316-340.

37. Sagitov S & Jagers P (2005) The coalescent effective size of age-structured populations. *Annals of Applied Probability* 15(3):1778-1797.
38. Pollak E (2011) Coalescent theory for age-structured random mating populations with two sexes. *Math. Biosci.* 233(2):126-134.
39. Westendorp RG & Kirkwood TB (1998) Human longevity at the cost of reproductive success. *Nature* 396(6713):743.
40. Pettay JE, Kruuk LE, Jokela J, & Lummaa V (2005) Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proc. Natl. Acad. Sci. U. S. A.* 102(8):2838-2843.
41. Thomas F, Teriokhin A, Renaud F, De Meeûs T, & Guégan J-F (2000) Human longevity at the cost of reproductive success: evidence from global data. *J. Evol. Biol.* 13(3):409-414.
42. Wright S (1939) Statistical genetics in relation to evolution, Vol. 802. *Hermann et Cie.: Paris*.
43. Segurel L, Wyman MJ, & Przeworski M (2014) Determinants of mutation rate variation in the human germline. *Annu Rev Genomics Hum Genet* 15:47-70.
44. Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology* 7(1):44.
45. Amster G & Sella G (2016) Life history effects on the molecular clock of autosomes and sex chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 113(6):1588-1593.
46. Evans BJ, Zeng K, Esselstyn JA, Charlesworth B, & Melnick DJ (2014) Reduced representation genome sequencing suggests low diversity on the sex chromosomes of tonkean macaque monkeys. *Mol. Biol. Evol.* 31(9):2425-2440.
47. Franchi L, Mandl AM, & Zuckerman S (1962) The development of the ovary and the process of oogenesis. *The Ovary*, ed Zuckerman S (Academic Press, London), Vol 1, pp 1-88.
48. Drost JB & Lee WR (1995) Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environ. Mol. Mutagen.* 25(S2):48-64.
49. Ehmcke J, Wistuba J, & Schlatt S (2006) Spermatogonial stem cells: questions, models and perspectives. *Hum. Reprod. Update* 12(3):275-282.
50. Gao Z, Wyman MJ, Sella G, & Przeworski M (2015) Interpreting the dependence of mutation rates on age and time. *ArXiv e-prints.* 1507:6890.
51. Charlesworth B & Charlesworth D (2010) *Elements of evolutionary genetics* (Roberts And Company Publishers, Greenwood Village, Colorado).
52. Nielsen CT*, et al.* (1986) Onset of the Release of Spermatozia (Supermarche) in Boys in Relation to Age, Testicular Growth, Pubic Hair, and Height. *J. Clin. Endocrinol. Metab.* 62(3):532-535.
53. Marson J, Meuris S, Cooper R, & Jouannet P (1991) Puberty in the male chimpanzee: progressive maturation of semen characteristics. *Biol. Reprod.* 44(3):448-455.
54. Heller CG & Clermont Y (1963) Spermatogenesis in man: an estimate of its duration. *Science* 140(3563):184-186.
55. Smithwick EB, Young LG, & Gould KG (1996) Duration of spermatogenesis and relative frequency of each stage in the seminiferous epithelial cycle of the chimpanzee. *Tissue Cell* 28(3):357-366.
56. Dixson AF (2009) *Sexual selection and the origins of human mating systems* (Oxford Univ, Press, Oxford).
57. Barr AB (1973) Timing of spermatogenesis in four nonhuman primate species. *Fertil. Steril.* 24(5):381-389.

58. Clermont Y & Antar M (1973) Duration of the cycle of the seminiferous epithelium and the spermatogonial renewal in the monkey Macaca arctoides. *Am. J. Anat.* 136(2):153-165.
59. Presgraves DC & Yi SV (2009) Doubts about complex speciation between humans and chimpanzees. *Trends Ecol. Evol.* 24(10):533-540.
60. Sayres MAW, Venditti C, Pagel M, & Makova KD (2011) Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. *Evolution* 65(10):2800-2815.
61. Sayres MAW & Makova KD (2011) Genome analyses substantiate male mutation bias in many species. *Bioessays* 33(12):938-945.
62. Hwang DG & Green P (2004) Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc. Natl. Acad. Sci. U. S. A.* 101(39):13994-14001.
63. Moorjani P, Amorim CEG, Arndt PF, & Przeworski M (2016) Variation in the molecular clock of primates. *Proceedings of the National Academy of Sciences* 113(38):10607-10612.
64. Li H & Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493-496.
65. Kaplan NL, Hudson RR, & Langley CH (1989) The Hitchhiking Effect Revisited. *Genetics* 123(4):887-899.
66. Hudson RR & Kaplan NL (1995) Deleterious background selection with recombination. *Genetics* 141(4):1605-1617.
67. Nordborg M, Charlesworth B, & Charlesworth D (1996) The effect of recombination on background selection. *Genet. Res.* 67(2):159-174.