Y-chromosome degeneration due to speciation and founder effect

Nianqin Zhang[1,2] and Yongjun Zhang[3]

[1]Department of Cardiology, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.
[2] Southern Medical University, Guangzhou, China.
[3]Science college, Liaoning Technical University, Fuxin, China.

Abstract

The Y chromosome in the XY sex-determination system is often shorter than its X counterpart, a condition attributed to degeneration after Y recombination ceases. Contrary to the traditional view of continuous, gradual degeneration, our study reveals stabilization within large mating populations. In these populations, we demonstrate that both mutant and active alleles on the Y chromosome can reach equilibrium through a mutation-selection balance. However, the emergence of a new species, particularly through the founder effect, can disrupt this equilibrium. Specifically, if the male founders of a new species carry only a mutant allele for a particular Y-linked gene, this allele becomes fixed, leading to the loss of the corresponding active gene on the Y chromosome. Our findings suggest that the rate of Y-chromosome degeneration may be linked to the frequency of speciation events associated with single-male founder events.

Keywords: sex chromosome, equilibrium, evolution, founder effect, speciation

1 Introduction

Sex is a fundamental characteristic of life that not only contributes to the diversity of offspring but also facilitates chromosome recombination, enabling

the accumulation of beneficial genes. However, the suppression of recombination can lead to the degeneration of sex chromosomes (Bachtrog et al 2014; Furman et al 2020).

Sex-determination mechanisms are incredibly diverse, with the XY system being a notable example. In this system, females have two X chromosomes, while males have one X and one Y chromosome. These sex chromosomes originated from autosomes, with one evolving into the Y chromosome upon acquiring a sex-determining gene—commonly SRY in mammals. The structural differences between the sex chromosomes contribute to the suppression of recombination, leading to the degradation of the Y chromosome. For instance, the human Y chromosome retains only three percent of its ancestral genes (Skaletsky et al 2003). A similar phenomenon is observed in the ZW sex-determination system, where the ZW genotype produces females and the ZZ genotype produces males, and the W chromosome tends to be shorter than the Z chromosome (Zhou et al 2014).

The mechanisms underlying the degeneration of the Y chromosome due to recombination suppression are still a subject of investigation. This can be examined at two levels. The first level focuses on the evolution of individual genes on the Y chromosome, disregarding the influence of other genes. The second level explores the interactions among different genes on the same Y chromosome.

At the second level, several models have been proposed. Muller's ratchet (Muller 1964; Ohno 1967; Charlesworth 1978) suggests that each loss of the class of Y chromosomes with the fewest deleterious mutations, known as a *click* of Muller's ratchet, is an irreversible event due to the absence of recombination (Bachtrog 2013). The background selection model suggests that the Y chromosome loses genetic variation as neutral mutations are eliminated along with deleterious mutations linked to them (Charlesworth et al 1993). The genetic hitchhiking model proposes that the Y chromosome gradually loses genetic activity as the frequency of deleterious mutations increases when they are in close proximity to beneficial mutations, such as SRY (Rice 1987). These mechanisms may act in combination (Bachtrog 2008).

To fully understand Y-chromosome degeneration, it is important to first study it independently at the first level before considering its combination with the second level. Lenormand *et al.* (Lenormand et al 2020) proposed a mechanism at the first level, demonstrating that the instability and divergence of *cis*-regulatory sequences in the non-recombining region may accelerate Y-chromosome degeneration. The examination of a single gene on the Y chromosome may also involve other models, such as the Hardy-Weinberg law (Johnston et al 2018) and Haldane's approach (Haldane 1935). However, these models have not been directly associated with Y-chromosome degeneration.

Traditionally, it has been assumed that Y-chromosome degeneration occurs gradually over time or generations (Charlesworth 2021). This assumption has led to concerns about the eventual disappearance of the human Y chromosome (Graves 2006). However, this assumption fails to explain the pattern

where degeneration is rarely observed in cold-blooded vertebrates, in contrast to birds and mammals. For example, a group of European tree frogs has shown no divergence between the X and Y chromosomes despite evolving for millions of years (Stöck et al 2011). Furthermore, the human Y chromosome has not lost any genes since the divergence of the human and chimpanzee lineages approximately 6 million years ago (Hughes et al 2012). These observations challenge the notion that Y-chromosome degeneration is a universal and inevitable process. There is evidence suggesting that degeneration can be reversed periodically, as indicated by theories such as sex chromosome turnover (Schartl 2004; Miura 2018) and occasional X-Y recombination (Perrin 2009; Stöck et al 2013; Rodrigues et al 2018). Furthermore, the rate of Y-chromosome degeneration on the second level has been shown to slow down over generations and potentially approach an equilibrium state. This phenomenon has been demonstrated through simulations that incorporate Muller's ratchet, background selection, and genetic hitchhiking (Bachtrog 2008). It is also worth investigating the existence of equilibrium on the first level, focusing on the behavior of individual genes.

One way to study the equilibrium for a single gene is by using the Hardy-Weinberg law (Johnston et al 2018). This model allows for the calculation of genotype frequencies from allele frequencies under the assumptions of a large population, random mating, and negligible mutation, selection, and migration. Although the Hardy-Weinberg law does not explicitly incorporate mutation and natural selection, it can still yield reasonable results when extended to include these factors.

Consider a locus with two alleles, G and g, and denote their equilibrium frequencies as p and q, respectively. In the case of an autosomal locus, the population frequencies of genotypes GG, Gg, and gg are given by p^2, $2pq$, and q^2, respectively. When the relative fitnesses of these genotypes are 1, 1, and 0, respectively, mutation-selection balance leads to $q = \sqrt{\mu}$, where μ is the mutation rate from G to g per generation. For an X-linked locus, where males have the genotypes G and g and their corresponding fitnesses are 1 and 0, respectively, the equilibrium frequency q is approximately 3μ.

Another approach to study the equilibrium for a single gene is through Haldane's approach (Haldane 1935), proposed in 1935. This approach does not assume the Hardy-Weinberg population and allows genotype frequencies to be determined directly by mutation-selection balance.

Y-chromosome degeneration exhibits species-specific patterns (Gschwend et al 2012) and cannot be simply characterized as a process occurring strictly over generations (Kratochvíl et al 2021), suggesting a potential link with speciation. In large populations, genetic stability typically prevents degeneration, but speciation may originate from founder events (Templeton 1980; Provine 2004; Joly 2011), involving only a few individuals. The founder effect, a cornerstone of population genetics, emphasizes the reduction of genetic diversity in new populations originating from small ancestral groups. This effect can disrupt genetic equilibrium and lead to the loss of active alleles, potentially

contributing to Y-chromosome degeneration through speciation. However, there is a lack of study on the role of the founder effect in contributing to Y-chromosome degeneration during speciation.

Our research delves into the dynamics between species-specific equilibrium and speciation driven by founder effects. We begin by employing Haldane's approach to evaluate genetic equilibria in X-linked and autosomal alleles, laying the groundwork for a detailed exploration of sex-linked alleles. Subsequently, we shift our focus to examine the impact of population size on Y-chromosome degeneration. This is followed by an analysis of how founder events influence speciation, thereby providing insights into the complexities of Y-chromosome evolution.

2 X-linked locus

Consider an X-linked locus with alleles G and g. Let μ represent the mutation rate from G to g per generation, assuming no difference between males and females and no back mutation. Suppose the population is large and mating is random. For females, the genotypes are GG, Gg, and gg, with relative fitnesses of 1, 1, and 0, respectively. For males, the genotypes are G and g, with fitnesses of 1 and 0, respectively. In other words, the allele G is the functional wild-type allele, which is not dosage-sensitive, and the non-functional mutant allele g is recessive lethal. Let us denote the frequency of genotype Gg in females by a. Female reproduction is shown in Table 1, where no mutation is considered.

Table 1 Female reproduction concerning an X-linked locus with alleles G and g, assuming no mutation. For males, the relative fitnesses of genotypes G and g are 1 and 0, respectively. For females, the relative fitnesses of genotypes GG, Gg, and gg are 1, 1, and 0, respectively. The frequencies of genotypes GG and Gg in the parental generation are $1-a$ and a, respectively.

	GG $1-a$	Gg a	gg
G 1	{ 2GG } $1-a$	{ GG, Gg } a	-
g 0	-	-	

When considering mutations, we need to make the following substitutions

$$\mathrm{GG} \rightarrow \begin{cases} \mathrm{GG}, & (1-\mu)^2, \\ \mathrm{Gg}, & 2\mu(1-\mu), \\ \mathrm{gg}, & \mu^2, \end{cases} \tag{1}$$

$$\mathrm{Gg} \rightarrow \begin{cases} \mathrm{Gg}, & (1-\mu), \\ \mathrm{gg}, & \mu. \end{cases} \tag{2}$$

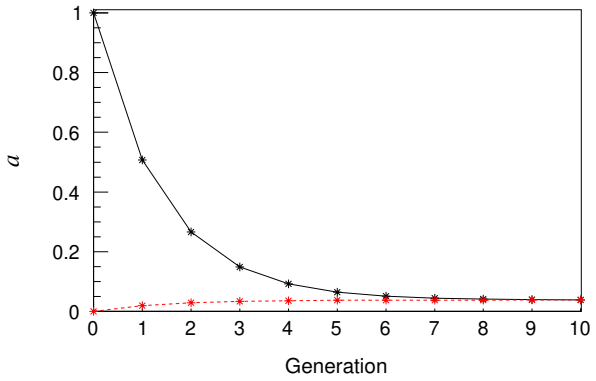

Fig. 1 Two examples of the evolving of a for an X-linked locus, with initial values of 1 and 0, respectively, where a represents the frequency of genotype Gg. Both examples converge to the equilibrium $\frac{1+4\mu-\sqrt{1+8\mu}}{2\mu} \approx 4\mu$, where μ is the mutation rate from allele G to allele g per generation. We use $\mu = 10^{-2}$ for demonstration purposes.

By plugging Eqs. (1) and (2) into Table (1), we obtain

$$\begin{aligned} N_{\mathrm{GG}} &\propto \left[(1-a) + \tfrac{a}{2}\right](1-\mu)^2, \\ N_{\mathrm{Gg}} &\propto \left[(1-a) + \tfrac{a}{2}\right] 2\mu(1-\mu) + \tfrac{a}{2}(1-\mu), \end{aligned} \tag{3}$$

where N_{GG} and N_{Gg} are the numbers of female offspring individuals with genotypes GG and Gg, respectively. The variable a may take different values for different generations. Thus, we introduce a_n for the parents and $a_{n+1} = N_{\mathrm{Gg}}/(N_{\mathrm{GG}} + N_{\mathrm{Gg}})$ for the offspring. Then we have the recursion equation

$$a_{n+1} = \frac{4\mu + a_n - 2\mu a_n}{2 + 2\mu - \mu a_n}. \tag{4}$$

The frequency a converges over generations, as demonstrated in Fig. 1. When the equilibrium is reached, there is an equation

$$a = \frac{4\mu + a - 2\mu a}{2 + 2\mu - \mu a}, \tag{5}$$

which reflects mutation-selection balance.

Under the constraint $0 \leq a \leq 1$, the equation has only one solution,

$$a = \frac{1 + 4\mu - \sqrt{1 + 8\mu}}{2\mu}, \tag{6}$$

which is approximately $a \approx 4\mu$ when $\mu \ll 1$.

If male individuals of genotype g are alive but infertile, their corresponding frequency would be $\frac{a}{2} + \left[\frac{a}{2} + (1-a)\right]\mu$, or approximately $2\mu + \mu$, where 2μ represents inherited mutations and μ represents *de novo* mutations. This result aligns with the findings of Haldane (Haldane 1935).

The frequency of the g allele can also be calculated and is approximately $\frac{4}{3}\mu$, in contrast to the result of 3μ derived from the Hardy-Weinberg equilibrium (Johnston et al 2018). This discrepancy arises because the assumption of a Hardy-Weinberg population does not hold when considering mutations.

3 Autosomal locus

Consider an autosomal locus with alleles G and g. Let μ represent the mutation rate from G to g per generation, assuming no difference between males and females and no back mutation. Suppose the population is large and mating is random. The genotypes are GG, Gg, and gg. Let their relative fitnesses be 1, 1, and 0, respectively. Let us denote the frequency of genotype Gg by a, with no difference between males and females. The reproduction, without considering any mutation, is shown in Table 2.

Table 2 Reproduction concerning an autosomal locus with alleles G and g, with no mutation considered. The relative fitnesses of genotypes GG, Gg, and gg are 1, 1, and 0, respectively. The frequencies of genotypes GG and Gg for the parental generation are $1-a$ and a, respectively.

	GG $1-a$	Gg a	gg
GG $1-a$	{ 4GG } $(1-a)^2$	{ 2GG, 2Gg } $(1-a)a$	-
Gg a	{ 2GG, 2Gg } $a(1-a)$	{ GG, 2Gg, gg } a^2	-
gg	-	-	-

To take into account mutation, we substitute Eqs. (1) and (2) into Table (2). Thus, we obtain the following equations

$$\begin{aligned} N_{\mathrm{GG}} &\propto \left[(1-a)^2 + a(1-a) + \tfrac{a^2}{4}\right](1-\mu)^2, \\ N_{\mathrm{Gg}} &\propto \left[(1-a)^2 + a(1-a) + \tfrac{a^2}{4}\right] 2\mu(1-\mu) \\ &\quad + \left[a(1-a) + \tfrac{a^2}{2}\right](1-\mu), \end{aligned} \tag{7}$$

where N_{GG} and N_{Gg} are the numbers of offspring individuals with genotypes GG and Gg, respectively. Let us introduce a_n for the parents and $a_{n+1} =$

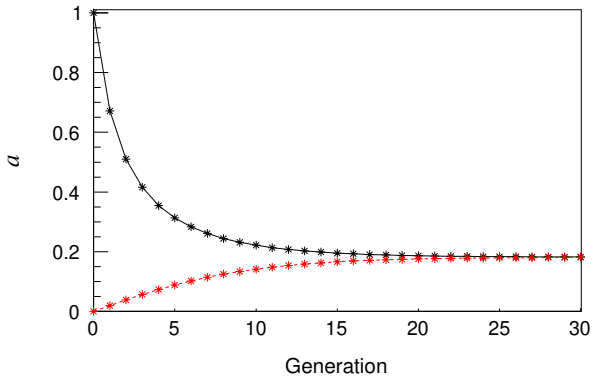

Fig. 2 Two examples of the evolving of a for an autosomal locus, with initial values of 1 and 0, respectively, where a is the frequency of genotype Gg. Both examples converge to the equilibrium $2\frac{\sqrt{\mu}-\mu}{1-\mu} \approx 2\sqrt{\mu}$. We use $\mu = 10^{-2}$ for demonstration purposes.

$N_{\mathrm{Gg}}/(N_{\mathrm{GG}} + N_{\mathrm{Gg}})$ for the offspring. Then we have the recursion equation

$$a_{n+1} = \frac{8\mu + 4a_n - 8\mu a_n - 2a_n^2 + 2\mu a_n^2}{4 + 4\mu - 4\mu a_n - a_n^2 + \mu a_n^2} = \frac{4\mu + 2(1-\mu)a_n}{2 + 2\mu + (1-\mu)a_n} \tag{8}$$

Frequency a will converge, as demonstrated in Fig. 2, and an equilibrium will be reached, satisfying the equation

$$a = \frac{4\mu + 2(1-\mu)a}{2 + 2\mu + (1-\mu)a} \tag{9}$$

Under the constraint of $0 \le a \le 1$, the equation has only one solution,

$$a = 2\frac{\sqrt{\mu} - \mu}{1 - \mu}. \tag{10}$$

We can further obtain the frequency of allele g as $\frac{\sqrt{\mu}-\mu}{1-\mu}$. When $\mu \ll 1$, it approximates to $\sqrt{\mu}$, which is consistent with the result obtained based on the Hardy-Weinberg law (Johnston et al 2018).

4 X-degenerate Y chromosome gene

The X chromosome may play a crucial role in Y-chromosome degeneration. Engelstädter (Engelstädter 2008) demonstrated through simulation that mutations occurring on the X chromosome exert strong selection against mutations on the Y chromosome, thereby slowing down the process of Y-chromosome degeneration. Hence, it is important to study the X and Y chromosomes together.

Table 3 Reproduction concerning an X-degenerate gene with alleles G and g, with no mutation considered. X and Y denote sex chromosomes with G while x and y denote sex chromosomes with g. For males, the relative fitnesses of genotypes XY, Xy, xY, and xy are 1, 1, 1, and 0, respectively. For females, the relative fitnesses of genotypes XX, Xx, and xx are 1, 1, and 0, respectively. The frequencies of genotypes XY, Xy, and xY are $1-b-c$, b, and c, respectively. The frequencies of genotypes XX and Xx are $1-a$ and a, respectively.

	XY $1-b-c$	Xy b	xY c	xy
XX $1-a$	{ 2XX, 2XY} $(1-a)(1-b-c)$	{ 2XX, 2Xy } $(1-a)b$	{ 2Xx, 2XY } $(1-a)c$	-
Xx a	{ XX, Xx, XY, xY} $a(1-b-c)$	{ XX, Xx, Xy, xy } ab	{ Xx, xx, XY, xY } ac	-
xx	-	-	-	-

In our analysis, we extend Haldane's approach to a locus that is non-sex-specific and located within a non-recombining region on both the X and Y chromosomes. Let's consider a locus with alleles G and g, where G represents the allele on the X and Y chromosomes, and g represents the mutant allele. We assume a random mating process in a large population, with a mutation rate of μ from G to g per generation. We will use the following symbols:

- X – X chromosome with G,
- x – X chromosome with g,
- Y – Y chromosome with G,
- y – Y chromosome with g,
- $a = \frac{N_{\mathrm{Xx}}}{N_{\mathrm{XX}}+N_{\mathrm{Xx}}}$,
- $b = \frac{N_{\mathrm{Xy}}}{N_{\mathrm{XY}}+N_{\mathrm{Xy}}+N_{\mathrm{xY}}}$,
- $c = \frac{N_{\mathrm{xY}}}{N_{\mathrm{XY}}+N_{\mathrm{Xy}}+N_{\mathrm{xY}}}$,

where N_{Xx} refers to the number of individuals with the Xx genotype and so on. The fitness of genotypes XX, Xx, XY, Xy, and xY is 1, while the fitness of genotypes xx and xy is 0. The reproduction is shown in Table 3, where no mutation is considered.

We plug the following into Table 3 to account for mutation.

$$\mathrm{XX} \rightarrow \begin{cases} \mathrm{XX}, & (1-\mu)^2, \\ \mathrm{Xx}, & 2\mu(1-\mu), \\ \mathrm{xx}, & \mu^2; \end{cases} \tag{11}$$

$$\mathrm{XY} \to \begin{cases} \mathrm{XY}, & (1-\mu)^2, \\ \mathrm{Xy}, & \mu(1-\mu), \\ \mathrm{xY}, & \mu(1-\mu), \\ \mathrm{xy}, & \mu^2; \end{cases} \tag{12}$$

$$\mathrm{Xx} \to \begin{cases} \mathrm{Xx}, & (1-\mu), \\ \mathrm{xx}, & \mu; \end{cases} \tag{13}$$

$$\mathrm{Xy} \to \begin{cases} \mathrm{Xy}, & (1-\mu), \\ \mathrm{xy}, & \mu; \end{cases} \tag{14}$$

$$\mathrm{xY} \to \begin{cases} \mathrm{xY}, & (1-\mu), \\ \mathrm{xy}, & \mu. \end{cases} \tag{15}$$

Let us introduce a_n, b_n, c_n for the parents and a_{n+1}, b_{n+1}, c_{n+1} for offspring. Then we have

$$\begin{aligned} a_{n+1} &= \tfrac{4\mu+a_n-2\mu a_n+2c_n-4\mu c_n-2a_nc_n+2\mu a_nc_n}{2+2\mu-\mu a_n-2\mu c_n-a_nc_n+\mu a_nc_n}, \\ b_{n+1} &= \tfrac{2\mu-\mu a_n+2b_n-2\mu b_n-a_nb_n+\mu a_nb_n}{2+2\mu-\mu a_n-2\mu b_n-a_nb_n+\mu a_nb_n}, \\ c_{n+1} &= \tfrac{2\mu+a_n-\mu a_n-2\mu b_n-a_nb_n+\mu a_nb_n}{2+2\mu-\mu a_n-2\mu b_n-a_nb_n+\mu a_nb_n}. \end{aligned} \tag{16}$$

These equations describe the change in variables (a, b, c) from one generation to the next. Over successive generations, the variables (a, b, c) converge to certain values. Equilibria exist, satisfying

$$\begin{cases} a = \tfrac{4\mu+a-2\mu a+2c-4\mu c-2ac+2\mu ac}{2+2\mu-\mu a-2\mu c-ac+\mu ac}, \\ b = \tfrac{2\mu-\mu a+2b-2\mu b-ab+\mu ab}{2+2\mu-\mu a-2\mu b-ab+\mu ab}, \\ c = \tfrac{2\mu+a-\mu a-2\mu b-ab+\mu ab}{2+2\mu-\mu a-2\mu b-ab+\mu ab}. \end{cases} \tag{17}$$

Under the constraints $0 \le a, b, c \le 1$, the equation has two solutions, giving rise to two equilibria. The first equilibrium is

$$(a, b, c) = \left(2\frac{\sqrt{\mu}-\mu}{1-\mu},\ \frac{\sqrt{\mu}-\mu}{1-\mu},\ \frac{\sqrt{\mu}-\mu}{1-\mu}\right). \tag{18}$$

Given that $\mu \ll 1$, the approximation is

$$(a, b, c) \approx (2\sqrt{\mu},\ \sqrt{\mu},\ \sqrt{\mu}). \tag{19}$$

The second equilibrium is

$$(a, b, c) = \left(\frac{1+4\mu-\sqrt{1+8\mu}}{2\mu},\ 1,\ 0\right). \tag{20}$$

Given that $\mu \ll 1$, the approximation is

$$(a, b, c) \approx (4\mu,\ 1,\ 0). \tag{21}$$

The first equilibrium corresponds to the equilibrium (10) in the scenario of an autosomal locus, while the second equilibrium corresponds to the equilibrium (6) in the scenario of an X-linked locus.

When the G allele is dosage-sensitive, it influences the Xy genotype and can decrease its frequency. Nevertheless, equilibria can still be attained in such cases. It has been proposed (Charlesworth 1978) that the presence of the Xy genotype could trigger dosage compensation. However, dosage compensation may benefit Xy but have negative effects on XY individuals. In the first equilibrium scenario, the XY genotype frequency, $1 - b - c$, is close to 1, while the Xy genotype frequency, b, is considerably lower, roughly $\sqrt{\mu} \ll 1$. This substantial difference highlights the widespread prevalence of the XY genotype, potentially impeding the development of dosage compensation mechanisms. Conversely, in the second equilibrium scenario, the Xy genotype is the only male genotype and therefore faces no competition, creating an environment conducive to the development of dosage compensation in males.

For a given set of variables (a_n, b_n, c_n), the subsequent generations will be determined by Eq. (16) and converge to the second equilibrium only when $b_n = 1$. In all other cases, regardless of the values of a_n and c_n, the population will always converge to the first equilibrium. In other words, if the G allele is initially present on the Y chromosome, it will persist in subsequent generations. This conclusion is supported by both analysis and simulations using Eq. (16). Setting $\mu = 0$ simplifies the analysis without altering the evolutionary trend. In this case, Eq. (16) simplifies to

$$\begin{aligned} a_{n+1} &= \tfrac{a_n + 2c_n - 2a_n c_n}{2 - a_n c_n}, \\ b_{n+1} &= \tfrac{2b_n - a_n b_n}{2 - a_n b_n}, \\ c_{n+1} &= \tfrac{a_n - a_n b_n}{2 - a_n b_n}. \end{aligned} \tag{22}$$

When $a_n > 0$, it is clear that b_{n+1} consistently decreases in the range $b_n \in (0, 1)$. This implies that unless the initial value of b is set to 1, i.e., $b_1 = 1$, it will inevitably converge to 0. In this scenario, reaching 0 corresponds to reaching the first equilibrium. This pattern is also evident in Table 3, where genotype xy, with a fitness of 0, is purged, along with the associated y. As a result, y is unable to effectively compete with Y and consistently experiences negative selection pressure. Thus, when $\mu = 0$, the outcome is either the exclusive presence of y throughout the entire process or the exclusive presence of Y at the end. However, when $\mu \neq 0$, certain Y mutate into y, guaranteeing the continual presence of at least a small fraction of y within the population.

5 Finite population

The existence of equilibria, particularly the first equilibrium, implies that Y-chromosome degeneration does not occur as long as the population size is large. However, these equilibria can be easily disrupted in small populations, potentially leading to Y-chromosome degeneration. To determine the applicable population size range for our model, we examine finite populations via a two-step analysis. In the first step, we establish the initial generation's genotype distribution based on equilibrium frequencies (Eq. 18) through random assignments. In other words, each individual of the first generation is randomly selected from an infinitely large population at the first equilibrium. In the second step, we use the relations in Table 3 to simulate the population's evolution through binomial sampling with a mutation rate of $\mu = 10^{-4}$, as detailed in the appendix.

Y-chromosome evolution varies with population size. Fig. 3 illustrates that in smaller populations, fluctuations play a prominent role, and the equilibrium positions are less stable. Conversely, Fig. 4 reveals that in very small populations, genetic drift can lead to the fixation of mutations, potentially resulting in the loss of active genes on the Y chromosome. Such events may have occurred in species with small-sized populations, as exemplified by frogs from Caribbean islands (Ma and Veltsos 2021). However, Fig. 3 highlights that larger populations have a greater capacity to maintain active genes on the Y chromosome. A population size exceeding 10,000 can be considered effectively infinite, allowing our model to be applicable. Notably, humans (Tomaszkiewicz et al 2016), being a species with a large population, should be able to retain their remaining genes on the Y chromosome.

6 Founder Effect and Speciation

The founder effect (Templeton 1980; Provine 2004; Joly 2011) may influence Y-chromosome evolution. Consider a single-male founder scenario in which one male establishes a new, large population. This population will reach one of the two possible equilibria, either the first equilibrium, i.e., Eq. (18), or the second equilibrium, i.e., Eq. (20). In cases of exponential population growth, genotype frequency fluctuations have limited time to cause mutation fixation before the population becomes too large for their influence. Thus, reaching which equilibrium is mainly determined by the founding male's genotype, randomly selected from the original population at the first equilibrium. The genotype of the founding female, or the genotypes of founding females if there are many, has no effect on this. If the founding male has the genotype Xy (i.e., $b_1 = 1$), the population will reach the second equilibrium; otherwise, with genotypes XY or xY, it will default to the first equilibrium. The likelihood of a randomly chosen male having genotype Xy—and thus the population reaching the second equilibrium—is roughly $\sqrt{\mu}$ (for a mutation rate μ of 10^{-4}, this value is

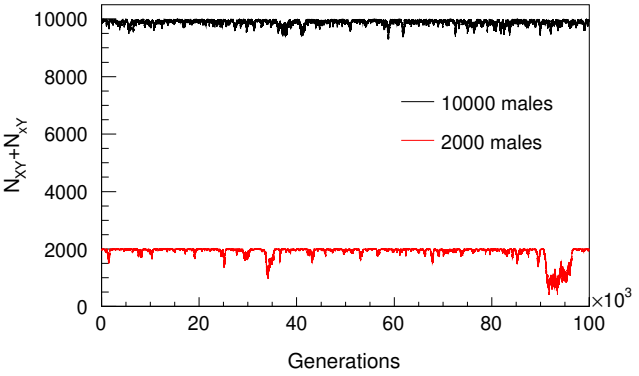

Fig. 3 Two examples of the Y chromosome evolving, with 10,000 and 2,000 males participating in reproduction, respectively. The vertical axis represents the total number of individuals with genotypes XY and xY. Both examples demonstrate fluctuations, but they also exhibit equilibrium positions. The simulation employs the binomial sampling method, which is commonly used in the Wright-Fisher model (Muirhead 2016). For illustrative purposes, a mutation rate of $\mu = 10^{-4}$ is employed. It is important to note that there exists a competition between mutation-selection balance and sampling randomness, resulting in frequent switching between the equilibrium and fluctuation windows. One wide fluctuation window is apparent on the lower curve near the right end. Lower mutation rates, such as $\mu = 10^{-5}$ or $\mu = 10^{-6}$, would result in wider fluctuation windows.

0.01). Scenarios involving multiple founding males have negligible probabilities of simultaneously having the genotype Xy—and thus reaching the second equilibrium—and are therefore not the focus of this study.

Let us focus on a single-male founder event that gives rise to a new population at the second equilibrium. On an evolutionary timescale, such an event has the potential to induce speciation. It is important to note that before the completion of speciation, gene flow may occur; for example, when males migrate from the original population to the new one. In this scenario, the founder effect can be neutralized, allowing for the recovery of the lost active allele and the restoration of the original equilibrium. Conversely, if speciation is finalized before gene flow can occur, the lost active allele becomes irretrievable, leading to Y-chromosome degeneration. The frequency of such single-male founder events, when associated with speciation, may influence the extent of Y-chromosome degeneration. To illustrate, consider a series of peripatric speciation events (Colvin 2018), each initiated by a single male. After 70 such events, the Y chromosome has approximately a 50% chance of losing the active allele, as calculated by $0.99^{70} \approx 0.5$.

Founder events, when coupled with speciation, offer a potential explanation for the observed disparities in Y-chromosome lengths across species, including European tree frogs (Stöck et al 2011) and humans (Skaletsky et al 2003). Similarly, bottleneck events (Weaver et al 2021), marked by significant population reductions, can influence Y-chromosome degeneration, as potentially seen in Caribbean island frogs (Ma and Veltsos 2021). It's important to note that

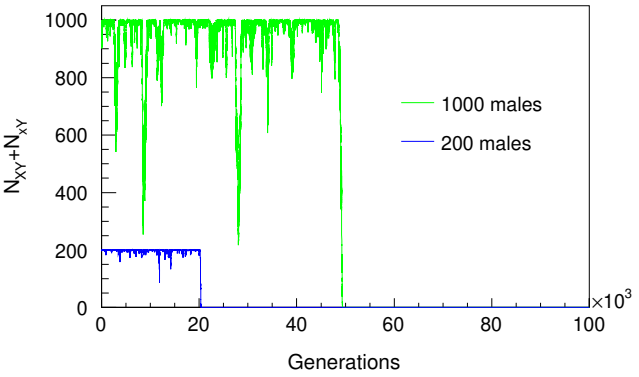

Fig. 4 Two examples of the Y chromosome evolving, with 1,000 and 200 males participating in reproduction, respectively. The mutation rate is $\mu = 10^{-4}$ for illustrative purposes. Significant fluctuations are observed, and mutation fixation occurs at generations 20,360 and 49,369, respectively.

while founder effects typically align with longer evolutionary timescales and speciation processes, bottleneck effects can manifest over much shorter durations, independent of speciation. When gene flow is effectively curtailed, the bottleneck effect becomes equivalent to the founder effect. By designing controlled environments where gene flow is eliminated, one may simulate founder scenarios to empirically validate their impact on Y-chromosome degeneration.

In our model, the framework is extendable to scenarios involving multiple loci. Consider a case with 1,000 non-sex-specific active alleles located in non-recombining regions of both the X and Y chromosomes, and a mutation rate $\mu = 10^{-4}$. A single-male founder event could result in the loss of a substantial number of active alleles from the Y chromosome, specifically 10, as calculated by $1000\sqrt{\mu}$. In other words, this event induces fixed mutations at 10 loci. When the number of active alleles is reduced to 100, a single-male founder event, on average, leads to one fixed mutation on the Y chromosome. When the number of active alleles further decreases to 10, it takes, on average, 10 single-male founder events to cause one fixed mutation on the Y chromosome. As the number of active alleles decreases, the rate of Y-chromosome degeneration experiences an exponential deceleration.

7 Conclusions

We have formulated an analytical model to scrutinize the evolution of individual genes, emphasizing the founder effect's impact on Y-chromosome degeneration. We show that in large mating populations, a genetic equilibrium can be established for non-sex-specific alleles in non-recombining regions of both X and Y chromosomes, effectively halting Y-chromosome degeneration. However, this stability also maintains a subset of males with inactive alleles. If such a male becomes the founder of a new species, the species will inherit only

the inactive Y-chromosome allele, setting the stage for Y-chromosome shortening. This dual nature of genetic stability—preventing degeneration within large populations while enabling it during speciation events, especially those with single-male founders—adds complexity to our understanding of genetic diversity and evolution.

Statements and Declarations

Simulation code repository: https://github.com/zyjlntu/Y-degeneration

References

Bachtrog D (2008) The temporal dynamics of processes underlying Y chromosome degeneration. Genetics 179(3):1513–1525. https://doi.org/10.1534/genetics.107.084012

Bachtrog D (2013) Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet 14(2):113–124. https://doi.org/10.1038/nrg3366

Bachtrog D, Mank JE, Peichel CL, et al (2014) Sex Determination: Why So Many Ways of Doing It? PLoS Biol 12(7)(7):e1001,899. https://doi.org/10.1371/journal.pbio.1001899

Charlesworth B (1978) Model for evolution of Y chromosomes and dosage compensation. Proc Natl Acad Sci U S A 75(11):5618–5622. https://doi.org/10.1073/pnas.75.11.5618

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. Genetics 134(4):1289–1303. https://doi.org/10.1093/genetics/134.4.1289

Charlesworth D (2021) The timing of genetic degeneration of sex chromosomes. Philos Trans R Soc B 376(1832). https://doi.org/10.1098/RSTB.2020.0093

Colvin AZ (2018) Peripatric speciation. WikiJournal Sci 1(2):8. https://doi.org/10.15347/wjs/2018.008

Engelstädter J (2008) Muller's ratchet and the degeneration of Y chromosomes: A simulation study. Genetics 180(2):957–967. https://doi.org/10.1534/genetics.108.092379

Furman BLS, Metzger DCH, Darolti I, et al (2020) Sex Chromosome Evolution: So Many Exceptions to the Rules. Genome Biol Evol 12(6):750–763. https://doi.org/10.1093/gbe/evaa081

Graves JA (2006) Sex chromosome specialization and degeneration in mammals. Cell 124(5):901–914. https://doi.org/10.1016/j.cell.2006.02.024

Gschwend AR, Weingartner LA, Moore RC, et al (2012) The sex-specific region of sex chromosomes in animals and plants. Chromosom Res 20(1):57–69. https://doi.org/10.1007/s10577-011-9255-y

Haldane JB (1935) The rate of spontaneous mutation of a human gene. J Genet (2004) 83(3):235–244. https://doi.org/10.1007/BF02717892

Hughes JF, Skaletsky H, Brown LG, et al (2012) Strict evolutionary conservation followed rapid gene loss on human and rhesus y chromosomes. Nature 483(7387):82–87. https://doi.org/10.1038/nature10843

Johnston HR, Keats BJB, Sherman SL (2018) Population genetics. In: Emery Rimoin's Princ. Pr. Med. Genet. Genomics Found. Elsevier, p 359–373, https://doi.org/10.1016/B978-0-12-812537-3.00012-3

Joly E (2011) The existence of species rests on a metastable equilibrium between inbreeding and outbreeding. An essay on the close relationship between speciation, inbreeding and recessive mutations. Biol Direct 6(1):1–55. https://doi.org/10.1186/1745-6150-6-62

Kratochvíl L, Stöck M, Rovatsos M, et al (2021) Expanding the classical paradigm: what we have learnt from vertebrates about sex chromosome evolution. Philos Trans R Soc B 376(1833). https://doi.org/10.1098/rstb.2020.0097

Lenormand T, Fyon F, Sun E, et al (2020) Sex Chromosome Degeneration by Regulatory Evolution. Curr Biol 30(15):3001. https://doi.org/10.1016/j.cub.2020.05.052

Ma WJ, Veltsos P (2021) The Diversity and Evolution of Sex Chromosomes in Frogs. Genes (Basel) 12(4):483. https://doi.org/10.3390/GENES12040483

Miura I (2018) Sex determination and sex chromosomes in Amphibia. Sex Dev 11(5-6):298–306. https://doi.org/10.1159/000485270

Muirhead CA (2016) Genetic Drift, Models of Random. Encycl Evol Biol pp 136–143. https://doi.org/10.1016/B978-0-12-800049-6.00024-X

Muller HJ (1964) The relation of recombination to mutational advance. Mutat Res 1(1):2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Ohno S (1967) Sex Chromosomes and Sex-Linked Genes, Monographs on Endocrinology, vol 1. Springer, Berlin, https://doi.org/10.1007/978-3-642-88178-7

Perrin N (2009) Sex reversal: A fountain of youth for sex chromosomes? Evolution 63(12):3043–3049. https://doi.org/10.1111/j.1558-5646.2009.00837.x

Provine WB (2004) Ernst Mayr: Genetics and speciation. Genetics 167(3):1041–1046. https://doi.org/10.1093/genetics/167.3.1041

Rice WR (1987) Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics 116(1):161–167. https://doi.org/10.1093/genetics/116.1.161

Rodrigues N, Studer T, Dufresnes C, et al (2018) Sex-chromosome recombination in common frogs brings water to the fountain-of-youth. Mol Biol Evol 35(4):942–948. https://doi.org/10.1093/molbev/msy008

Schartl M (2004) Sex chromosome evolution in non-mammalian vertebrates. Curr Opin Genet Dev 14(6):634–641. https://doi.org/10.1016/j.gde.2004.09.005

Skaletsky H, Kuroda-Kawaguchl T, Minx PJ, et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423(6942):825–837. https://doi.org/10.1038/nature01722

Stöck M, Horn A, Grossen C, et al (2011) Ever-young sex chromosomes in European tree frogs. PLoS Biol 9(5):e1001,062. https://doi.org/10.1371/journal.pbio.1001062

Stöck M, Savary R, Betto-Colliard C, et al (2013) Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic green toads (Bufo viridis subgroup). J Evol Biol 26(3):674–682. https://doi.org/10.1111/jeb.12086

Templeton AR (1980) The theory of speciation via the founder principle. Genetics 94(9):1011–1038. https://doi.org/10.1093/genetics/94.4.1011

Tomaszkiewicz M, Rangavittal S, Cechova M, et al (2016) A time- and cost-effective strategy to sequence mammalian Y chromosomes: An application to the de novo assembly of gorilla Y. Genome Res 26(4):530–540. https://doi.org/10.1101/gr.199448.115

Weaver SC, Forrester NL, Liu J, et al (2021) Population bottlenecks and founder effects: implications for mosquito-borne arboviral emergence. https://doi.org/10.1038/s41579-020-00482-8

Zhou Q, Zhang J, Bachtrog D, et al (2014) Complex evolutionary trajectories of sex chromosomes across bird taxa. Science 346:1246,338. URL https://www.science.org/doi/10.1126/science.1246338

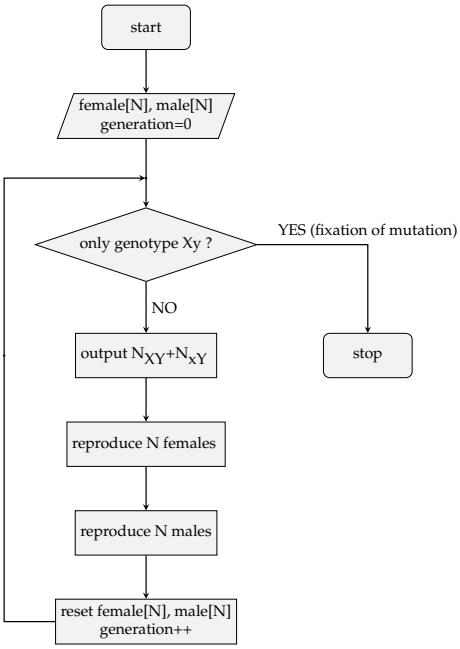

Fig. A1 Flowchart of the main program (C++) for simulation. In the first generation, N males and N females are randomly selected from the first equilibrium (18). The program then checks if all N males have the genotype Xy. If they do, the program stops. Otherwise, the next generation is reproduced, and the program continues. The program stops when mutation fixation occurs. The output $N_{XY}+N_{xY}$ is used to generate Fig. 3 and Fig. 4.

Appendix A Simulation

Simulation is necessary when applying the relations in Table 3 to a finite population. Fig. A1 presents the flowchart of the main program, outlining the process involved in the simulation. Fig. A2 provides a subroutine flowchart specifically for reproducing female offspring. The process of reproducing males follows a similar procedure.

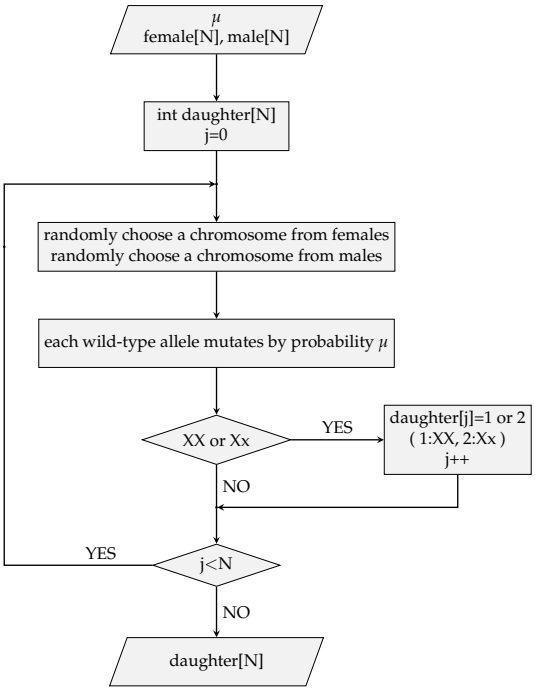

Fig. A2 Flowchart of the program for reproducing females in the simulation. The input consists of N females and N males, and the program generates N female offspring as the output. The same process can be applied to reproduce males. The program utilizes the binomial sampling method, which is commonly used in the Wright-Fisher model to study fluctuations in the absence of mutation and selection.