

Article

# Limiting Genotype Frequencies of Y-Linked Genes with a Mutant Allele in a Two-Sex Population

Miguel González <sup>1,†</sup> , Cristina Gutiérrez <sup>2,\*,†</sup>  and Rodrigo Martínez <sup>3,†</sup> 

<sup>1</sup> Department of Mathematics and ICCAEx, Faculty of Sciences, University of Extremadura, Avda. Elvas s/n, 06006 Badajoz, Spain; mvelasco@unex.es

<sup>2</sup> Department of Mathematics and ICCAEx, Faculty of Business, Finance and Tourism, University of Extremadura, Avda. de la Universidad s/n, 10071 Cáceres, Spain

<sup>3</sup> Department of Mathematics and ICCAEx, University Center of Plasencia, University of Extremadura, Avda. Virgen del Puerto 2, 10600 Plasencia, Spain; rmartinez@unex.es

\* Correspondence: cgutierrez@unex.es; Tel.: +34-927-257-000

† These authors contributed equally to this work.

**Abstract:** A two-type two-sex branching process is considered to model the evolution of the number of carriers of an allele and its mutations of a Y-linked gene. The limiting growth rates of the different types of couples and males (depending on the allele, mutated or not, that they carry on) on the set of coexistence of both alleles and on the fixation set of the mutant allele are obtained. In addition, the limiting genotype of the Y-linked gene and the limiting sex frequencies on those sets are established. Finally, the main results have been illustrated with simulated studies contextualized in problems of population genetics.

**Keywords:** two-sex Galton–Watson branching process; genotype frequencies; mutant allele; growth rate; sex ratio



**Citation:** González, M.; Gutiérrez, C.; Martínez, R. Limiting Genotype Frequencies of Y-Linked Genes with a Mutant Allele in a Two-Sex Population. *Mathematics* **2021**, *9*, 131. <https://doi.org/10.3390/math9020131>

Received: 11 November 2020

Accepted: 7 January 2021

Published: 9 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In many populations, the sex of an individual is determined by a pair of chromosomes X and Y. The females are homozygous and carry XX chromosomes, the males are heterozygous and carry XY chromosomes. Sex-linkage occurs if the phenotypic expression of an allele is related to the chromosomal sex of the individual. Here we are interested in Y-linkage. Although there are much less Y-linked traits than X-linked traits, recent studies have shown the significance of Y-linkage in the biology of humans and other animals, see e.g., [1]. A mathematical model for the propagation of Y-linked genes in two-sex populations was introduced by González et al. [2] and further studied in [3,4]. The genes occur in two allelic forms, called R and r, where the latter one cannot mutate, the mating is assumed to be monogamous (perfect fidelity mating) and with blind choice, that is, females ignore or do not care about the genotype of the chosen partner.

In this paper we consider the model just outlined but with the additional possibility of mutations. Such a model was introduced in González et al. [5]. More precisely, in this paper it is assumed that R represents the nonmutant allelic form of the gene which can mutate and thus give rise to new allelic forms, all of them denoted by r and called mutant alleles. Although the latter can also mutate this is ignored in this case because we do not further distinguish between possible types of mutation, and a reversion, i.e., a mutation from r back to R, is not allowed. In such paper, conditions for the extinction and survival of those alleles were studied and also further biological information that explain the relevance of mutations in this context were provided. Furthermore, inference on the main parameters of the model was made in [6].

The aim of this work is to find the allele growth rates as well as the limiting genotype and sex frequencies for such a model in the supercritical case (in terms of the usual

nomenclature used in branching processes theory). Particularly, the growth rate of the mutant allele on its fixation set and the one of the nonmutant allele on the coexistence set are obtained. Also on that set of coexistence, the growth rate of the mutant allele is studied. However, in this case, mutant allele survival depends on the behaviour of the nonmutant allele, and this fact makes this study more complex. The obtained results depend on the relation between the means of the offspring distributions of the different genotypes (with mutant or nonmutant alleles) and the probability of mutation. Finally, we illustrate the attained results by way of simulated studies contextualised in problems of population genetics, such as the evolution of infertility problems in males, the existence of degeneration of Y chromosome or the possibility of diversity of species without natural selection.

We have split the paper into eight sections. In Section 2 we provide a mathematical description of the model. Then, in Section 3, we research the limiting growth rate of the nonmutant allele. Next, we found the rate of growth of the mutant allele on the coexistence set and also on its fixation set, in Section 4. As consequence, in Section 5, we deal with the limiting genotype and sex frequencies. Illustrative examples are shown in Section 6. Finally, in Section 7 some concluding remarks are given and the paper ends with an Appendix which contains the proofs of the results provided throughout the paper.

### 2. Model Description

A detailed description of the mathematical model is provided next: individuals are called females, R-males (males with R-genotype) and r-males (males with r-genotype), and these can mate with females to form R- or r-couples depending on whether the male is of type R or r, respectively. In accordance with the rules of genetic inheritance and by taking the possibility of mutation into account, an R-couple can give birth to females, R-males, and r-males, whereas, given the assumption of no backmutation, an r-couple gives birth to females and r-males.

Assuming nonoverlapping generations and given the number of R- and r-couples in generation  $n$ , denoted by  $Z_n^R$  and  $Z_n^r$ , respectively, the numbers of females, males, and couples of each genotype in the next generation are determined by the following two-stage procedure of reproduction and mating:

In the *reproduction phase*, couples produce offspring independent of each other and in accordance with an offspring law which does not depend on the generation, is the same for a given genotype, but may vary for different genotypes since the mutation could affect the reproductive capacity. Formally, the numbers of females and males produced by the couples of type R and r are determined by sequences of independent, identically distributed (i.i.d.) random vectors, viz.

$$(F_{ni}^R, M_{ni}^R, M_{ni}^{R \rightarrow r}) \text{ and } (F_{ni}^r, M_{ni}^{r \rightarrow r}), (n, i) \in \mathbb{N}_0 \times \mathbb{N},$$

respectively. Here,  $F_{ni}^x$  denotes the number of female offspring of the  $i$ th  $x$ -couple in generation  $n$  with  $x \in \{R, r\}$ , while  $M_{ni}^R$  and  $M_{ni}^{r \rightarrow r}$  denote the numbers of nonmutant male offspring of the  $i$ th R and r-couples, respectively, in generation  $n$ , and  $M_{ni}^{R \rightarrow r}$  denotes the number of mutant males produce by the  $i$ th R-couple in generation  $n$ . All variables are assumed to have finite mean and variance, and we denote

$$m_R = E[F_{ni}^R + M_{ni}^R + M_{ni}^{R \rightarrow r}] \text{ and } m_r = E[F_{ni}^r + M_{ni}^{r \rightarrow r}].$$

Furthermore, the conditional law of  $(F_{ni}^R, M_{ni}^R, M_{ni}^{R \rightarrow r})$  given  $F_{ni}^R + M_{ni}^R + M_{ni}^{R \rightarrow r} = k$ ,  $k \geq 0$ , is assumed to be multinomial with parameters  $k$ ,  $\alpha$ ,  $(1 - \alpha)(1 - \beta)$  and  $(1 - \alpha)\beta$  and the conditional law of  $(F_{ni}^r, M_{ni}^{r \rightarrow r})$  given  $F_{ni}^r + M_{ni}^{r \rightarrow r} = k$ ,  $k \geq 0$ , to be multinomial with parameters  $k$ ,  $\alpha$ ,  $(1 - \alpha)$ . This means that  $\alpha$  equals the probability that an offspring is female, while  $\beta$  equals the probability that a newborn male of an R-couple has the mutant allele. At the end of the reproduction phase, one has  $F_{n+1}$ ,  $M_{n+1}^R$ , and  $M_{n+1}^r$ , the total numbers of females, R-males, and r-males, respectively, which constitute the generation  $n + 1$  and they are determined by the following relations (with the empty sum defined as 0):

$$\begin{aligned}
 F_{n+1} &= \sum_{i=1}^{Z_n^R} F_{ni}^R + \sum_{j=1}^{Z_n^r} F_{nj}^r, \\
 M_{n+1}^R &= \sum_{i=1}^{Z_n^R} M_{ni}^R \quad \text{and} \quad M_{n+1}^r = M_{n+1}^{R \rightarrow r} + M_{n+1}^{r \rightarrow r},
 \end{aligned} \tag{1}$$

where

$$M_{n+1}^{R \rightarrow r} = \sum_{i=1}^{Z_n^R} M_{ni}^{R \rightarrow r} \quad \text{and} \quad M_{n+1}^{r \rightarrow r} = \sum_{j=1}^{Z_n^r} M_{nj}^{r \rightarrow r}.$$

Thus,  $M_{n+1}^{R \rightarrow r}$  and  $M_{n+1}^{r \rightarrow r}$  denote the total number of  $r$ -males stemming from  $R$ - and  $r$ -couples in generation  $n$ , respectively. Moreover, we denote  $Z_n = Z_n^R + Z_n^r$  to the total number of couples in generation  $n$ .

In the *mating phase* the number of couples of each genotype, that is  $Z_{n+1}^R$  and  $Z_{n+1}^r$ , are determined as follows, given the total number of females,  $R$ -males and  $r$ -males in generation  $n + 1$ . We assume monogamous mating which means that the total number of couples equals the minimum of  $F_{n+1}$  and  $M_{n+1} = M_{n+1}^R + M_{n+1}^r$ . In the case when  $F_{n+1} \geq M_{n+1}$ , this means that each male finds a female to mate with, i.e.,  $Z_{n+1}^R = M_{n+1}^R$  and  $Z_{n+1}^r = M_{n+1}^r$ . In the case when  $F_{n+1} < M_{n+1}$ , females must choose a partner. We assume that this is done without preference of the genotype (blind choice). As a consequence, the conditional law of the total number of  $x$ -couples  $Z_{n+1}^x$  given  $(F_{n+1}, M_{n+1}^R, M_{n+1}^r)$  is hypergeometric with parameters  $(F_{n+1}, M_{n+1}, M_{n+1}^x)$  with  $x \in \{R, r\}$ .

The bivariate sequence  $\{(Z_n^R, Z_n^r)\}_{n \geq 0}$  describing the evolution of the number of couples for each type of males over generations is called *Y-linked two-sex branching process with mutations, fidelity mating and blind choice of males*. As shown in [5], it is a temporally homogeneous multitype Markov chain and exhibits the extinction-explosion dichotomy which is usual for branching processes with independent reproduction.

Notice that, since the empty sum is assumed to be zero, if in some generation there are no mating units of type  $R$  then, from this generation on, the couples and males of that type as well as mutant-males coming from them no longer exist, that is, if  $Z_n^R = 0$  for some  $n > 0$ , then  $Z_k^R = 0, M_k^R = 0$  and  $M_k^{R \rightarrow r} = 0$  for all  $k > n$ . Also, if  $Z_n^R = 0$  and  $Z_n^r = 0$  for some  $n > 0$ , then  $Z_k^r = 0$  and  $M_k^r = 0$  for all  $k > n$ . However, this behaviour is different for the  $r$ -allele when  $Z_n^R \neq 0$ . Indeed, despite  $Z_n^r = 0$ , it could happen that some  $R$ -couple gives birth to males whose corresponding allele has suffered a mutation and some of these males could mate forming couples of type  $r$ . Hence, if  $Z_n^R \neq 0$ , one can find that  $M_k^r > 0$  and  $Z_k^r > 0$ , for some  $k > n$ , even being  $Z_n^r = 0$ . Taking into account this fact, in [5] was shown that the survival of the population over generations is determined by the two events  $\{Z_n^R \rightarrow 0, Z_n^r \rightarrow \infty\}$ , termed  $r$ -fixation, and  $\{Z_n^R \rightarrow \infty, Z_n^r \rightarrow \infty\} = \{Z_n^R \rightarrow \infty\}$  almost surely (a.s.), termed simultaneous survival of both genotypes or coexistence, that is, the survival of  $R$ -genotype implies also the survival of  $r$ -genotype.

The following sections are devoted to the study of asymptotic growth of each genotype on survival events (in the usually called supercritical case). In all of them we shall write  $P_{(i,j)}(\cdot)$  for  $P(\cdot | Z_0^R = i, Z_0^r = j)$  or even the index  $(i, j)$  will be dropped in the notation if there is no ambiguity. To end this section, it is worth mentioning that detailed investigation of some particular two-type branching processes beyond the general theory has been also developed in other settings as, for example, in [7].

### 3. Nonmutant Allele Growth Rate

Next, we assume that  $(\alpha \wedge (1 - \alpha))(1 - \beta) m_R > 1$ , where, recall that  $(1 - \alpha)(1 - \beta) m_R$  is the mean number of  $R$ -males per  $R$ -couple and  $\alpha(1 - \beta) m_R$  can be interpreted as the mean number of females produced by  $R$ -couples who mate with  $R$ -males. Under this condition, it was proved in [5] that the survival of  $R$ -genotype (equivalently, simultaneous survival of both genotypes) occurs with positive probability. It then makes, to determine in

this case, the limiting growth rates for the number of R-couples and R-males. Answer is provided by the following result.

**Theorem 1.** Let  $\tau_R = (\alpha \wedge (1 - \alpha))(1 - \beta) m_R$ . If  $\tau_R > 1$ , then there exists a random variable  $W_R$  which is positive and finite on  $\{Z_n^R \rightarrow \infty\}$ , such that

$$\lim_{n \rightarrow \infty} \frac{Z_n^R}{\tau_R^n} = W_R \text{ and } \lim_{n \rightarrow \infty} \frac{M_n^R}{\tau_R^n} = \frac{(1 - \alpha)(1 - \beta) m_R}{\tau_R} W_R \text{ a.s. on } \{Z_n^R \rightarrow \infty\}.$$

Intuitively speaking, this theorem establishes that the numbers of R-couples and R-males grow geometrically at the rate given by the mean number of nonmutant males generated by an R-couple or by the mean number of females stemming from an R-couple who mate with nonmutant males, respectively, depending on whether  $\alpha$  is higher than 0.5 or not.

#### 4. Mutant Allele Growth Rate

Firstly, we consider the study of the rate of growth of the process  $\{Z_n^r\}_{n \geq 0}$  on the event of coexistence, that is on  $\{Z_n^R \rightarrow \infty\}$ . This study turns out to be more difficult than the previous case because of the possible dependency of the survival of the  $r$ -allele on the behaviour of the R-allele. In fact, from the previous result, it is easy to deduce that  $r$ -males stemming from R-couples grow geometrically at the same rate as R-males. To carry out this study, we again assume that  $(\alpha \wedge (1 - \alpha))(1 - \beta) m_R > 1$  and distinguish two cases: when  $\alpha > 0.5$  and when  $\alpha \leq 0.5$ , in order to better understanding. Finally, we consider the study on  $r$ -fixation set.

##### 4.1. On Coexistence Set: Case $\alpha > 0.5$

When  $\alpha > 0.5$ , the number of females always exceeds the number of males from some generation onwards on the set of survival of both genotypes (see Corollary A.1 in [5]). Then, eventually, the number of each type of couple equals the number of males of each genotype. Therefore, the process behaves essentially as a two-type Galton–Watson process where one of the types (R-allele) gives birth the two types of individuals existing in the population (R-males and  $r$ -males) while the other type ( $r$ -allele) only produces individuals of its own type ( $r$ -males). For more details, the reader is referred to the state-space representation in page 10 of the Appendix A. The limit theorems for a reducible multitype Galton–Watson process (as it is our case) are well studied in [8] and can be applied here to obtain the following theorem.

**Theorem 2.** If  $\alpha > 0.5$  and  $(1 - \alpha)(1 - \beta) m_R > 1$ , then there exists a random variable  $W_r$  which is positive and finite on  $\{Z_n^R \rightarrow \infty\}$ , such that, a.s. on  $\{Z_n^R \rightarrow \infty\}$ ,

- (i) if  $m_r > (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{((1 - \alpha) m_r)^n} = \lim_{n \rightarrow \infty} \frac{M_n^r}{((1 - \alpha) m_r)^n} = W_r$ ,
- (ii) if  $m_r = (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{n((1 - \alpha) m_r)^n} = \lim_{n \rightarrow \infty} \frac{M_n^r}{n((1 - \alpha) m_r)^n} = \frac{\beta}{1 - \beta} W_r$ ,
- (iii) if  $m_r < (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{((1 - \alpha)(1 - \beta) m_R)^n} = \lim_{n \rightarrow \infty} \frac{M_n^r}{((1 - \alpha)(1 - \beta) m_R)^n} = \gamma W_r$ ,

with  $W_R$  as in Theorem 1 and  $\gamma = \frac{\beta m_R}{(1 - \beta) m_R - m_r}$ .

Notice that, in the previous results it is shown that the rate of growth of  $\{Z_n^r\}_{n > 0}$  (and also the rate of growth of  $\{M_n^r\}_{n > 0}$ ) changes depending on the relation between  $m_r$  (the mean number of offspring per  $r$ -couple) and  $(1 - \beta) m_R$  (the mean number of offspring per R-couples related with R-allele). In fact, when  $m_r > (1 - \beta) m_R$  its asymptotic growth is geometric being the rate of growth the mean number of males stemming from  $r$ -couples. On the other hand, when  $m_r = (1 - \beta) m_R$ , the normalized sequence is  $\{n((1 -$

$\alpha) m_r)^n\}_{n>0} = \{n((1 - \alpha)(1 - \beta) m_R)^n\}_{n>0}$ . However, when  $m_r < (1 - \beta) m_R$ , the growth of the  $r$ -allele is mainly (at least in part or totally when  $m_r = 0$ ) due to the mutations. For that, its asymptotic growth is geometric with rate given by the mean number of males produced by  $R$ -couples (the same rate as  $R$ -allele (see Theorem 1)). Therefore, when  $m_r \geq (1 - \beta) m_R$ , the  $r$ -allele is the dominant one in the population (in the sense that eventually there are more males in the population with  $r$ -genotype than with  $R$ -genotype).

4.2. On Coexistence Set: Case  $\alpha \leq 0.5$

Now, when  $\alpha < 0.5$ , the number of males is eventually higher than the number of females (see Corollary A.1 in [5]). Then, in this case, the total numbers of couples of each type are distributed according to hypergeometric distributions. Therefore, the process  $\{(Z_n^R, Z_n^r)\}_{n \geq 0}$  cannot be seen as a multitype branching process as in the previous case. Moreover, in the boundary case,  $\alpha = 0.5$ , we have an oscillating situation where we cannot assert that eventually the number of females (or males) is higher than the number of males (or females) from one generation onward. These two statements make the case  $\alpha \leq 0.5$  to be more complicated to study from a mathematical point of view than the case  $\alpha > 0.5$ , although we can establish similar results to such given in Theorem 2.

In particular, it can be proved that when  $\alpha \leq 0.5$  and  $m_r < (1 - \beta) m_R$ , the asymptotic growth of the total number of  $r$ -couples is geometric being the growth rate the mean number of females stemming from  $R$ -couples who have mated with  $R$ -males, that is  $\alpha(1 - \beta) m_R$  (the same rate as  $R$ -allele, see Theorem 1), while than, when  $m_r > (1 - \beta) m_R$ , it is also geometric, but now being the rate of growth the mean number of females stemming from  $r$ -couples, that is  $\alpha m_r$ . The case  $m_r = (1 - \beta) m_R$  is special, being the normalised sequence  $\{n(\alpha m_r)^n\}_{n>0} = \{n(\alpha(1 - \beta) m_R)^n\}_{n>0}$ , since in each generation  $R$ -couples can produce  $r$ -males and this partial pedigree grows geometrically at rate given by  $\alpha m_r$  (the mean number of females stemming from  $r$ -couple).

**Theorem 3.** *If  $\alpha \leq 0.5$  and  $\alpha(1 - \beta) m_R > 1$ , then there exists a random variable  $W_r^*$  which is positive and finite on  $\{Z_n^R \rightarrow \infty\}$ , such that, a.s. on  $\{Z_n^R \rightarrow \infty\}$ ,*

- (i) *if  $m_r > (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{(\alpha m_r)^n} = \frac{\alpha}{1 - \alpha} \lim_{n \rightarrow \infty} \frac{M_n^r}{(\alpha m_r)^n} = W_r^*$ ,*
- (ii) *if  $m_r = (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{n(\alpha m_r)^n} = \frac{\alpha}{1 - \alpha} \lim_{n \rightarrow \infty} \frac{M_n^r}{n(\alpha m_r)^n} = \frac{\beta}{1 - \beta} W_R$ ,*
- (iii) *if  $m_r < (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{(\alpha(1 - \beta) m_R)^n} = \frac{\alpha}{1 - \alpha} \lim_{n \rightarrow \infty} \frac{M_n^r}{(\alpha(1 - \beta) m_R)^n} = \gamma W_R$ ,*

*with  $W_R$  as in Theorem 1 and  $\gamma$  as in Theorem 2.*

4.3. On  $r$ -Fixation Set

Finally, we study the growth rate of  $r$ -allele on  $r$ -fixation set, that is on  $\{Z_n^R \rightarrow 0, Z_n^r \rightarrow \infty\}$ . It was shown in [5] that the process  $\{Z_n^r\}_{n \geq 0}$  evolves as a two-sex Galton–Watson branching process on the event  $\{Z_n^R \rightarrow 0\}$  (at least from one  $n$  on for each path); therefore, the asymptotic properties established by [9] can be applied here and we deduce the following result:

**Theorem 4.** *Let  $\tau_r = (\alpha \wedge (1 - \alpha)) m_r$ . If  $\tau_r > 1$ , then there exists a random variable  $\tilde{W}_r$  which is positive and finite on  $\{Z_n^R \rightarrow 0, Z_n^r \rightarrow \infty\}$ , such that, a.s. on  $\{Z_n^R \rightarrow 0, Z_n^r \rightarrow \infty\}$ ,*

$$\lim_{n \rightarrow \infty} \frac{Z_n^r}{\tau_r^n} = \tilde{W}_r \text{ and } \lim_{n \rightarrow \infty} \frac{M_n^r}{\tau_r^n} = \frac{(1 - \alpha) m_r}{\tau_r} \tilde{W}_r.$$

Intuitively, this theorem states that, if  $R$ -couples have disappeared when the number of  $r$ -couples explodes to infinity, then this number as well as the number of  $r$ -males grow geometrically at the rate given by the minimum between the mean number of males or females stemming from those  $r$ -couples. Notice that this behaviour on  $r$ -fixation set is the same as the one obtained on the coexistence set when  $m_r > (1 - \beta) m_R$ .



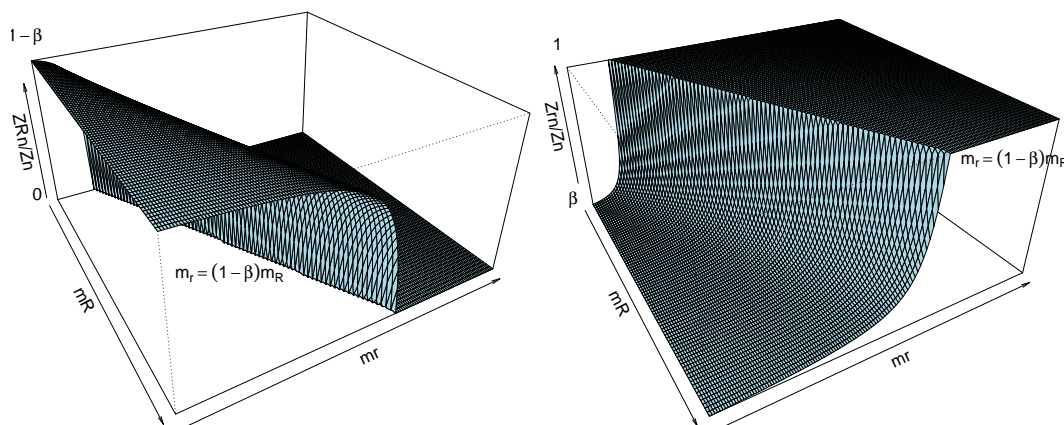
### 5. Limiting Genotype and Sex Frequencies

From the previous study, the following results relative to the limiting genotype and the sex frequencies are easily deduced and therefore their proofs are omitted. Recall that  $Z_n = Z_n^R + Z_n^r$  denotes the total number of mating units in generation  $n$ .

**Theorem 5.** *If  $(\alpha \wedge (1 - \alpha))(1 - \beta) m_R > 1$ , then, a.s. on  $\{Z_n^R \rightarrow \infty\}$ ,*

- (i) *if  $m_r \geq (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^R}{Z_n} = 0$  and  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{Z_n} = 1$ .*
- (ii) *if  $m_r < (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^R}{Z_n} = \frac{(1 - \beta) m_R - m_r}{m_R - m_r}$  and  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{Z_n} = \frac{\beta m_R}{m_R - m_r}$ .*

The same result can be established in terms of males. Anyway, notice that the limiting R-genotype frequency is a constant less than or equal to  $1 - \beta$ , achieving this maximum value when  $m_r = 0$  and the minimum (equals to 0) when  $m_r \geq (1 - \beta) m_R$ . Hence, for the r-genotype, the limiting frequency is always positive, higher than or equal to  $\beta$ , achieving this minimum value when  $m_r = 0$  and the maximum (equals to 1) when  $m_r \geq (1 - \beta) m_R$  (see Figure 1). We conclude thus that, independently of the mutation rate, the higher reproduction capacity of the r-genotype, the lower the limiting R-genotype frequency. Moreover, there is no dominant genotype in the case  $m_r < (1 - \beta) m_R$ , which implies an important difference in comparison with the results described in [3] for the Y-linked model without mutations, where the limiting genotype frequency is null or one. Notice that this 0–1 duality is also obtained on r-fixation set, that is, Theorem 5 (i) holds on  $\{Z_n^R \rightarrow 0, Z_n^r \rightarrow \infty\}$ , when  $(\alpha \wedge (1 - \alpha)) m_r > 1$ .



**Figure 1.** Limiting genotype frequency for R-allele (left) and for r-allele (right), depending on  $m_R$  and  $m_r$ , for fixed  $\beta$ .

Finally, we give a general result related to the limiting sex frequency. One can observe that, in all cases, the limiting sex frequency in the population only depends on the probability of an offspring to be female.

**Theorem 6.** *It is verified that*

$$\lim_{n \rightarrow \infty} \frac{F_n}{F_n + M_n} = \alpha$$

- (i) *a.s. on  $\{Z_n^R \rightarrow \infty\}$ , if  $(\alpha \wedge (1 - \alpha))(1 - \beta) m_R > 1$ , and*
- (ii) *a.s. on  $\{Z_n^R \rightarrow 0, Z_n^r \rightarrow \infty\}$ , if  $(\alpha \wedge (1 - \alpha)) m_r > 1$ .*

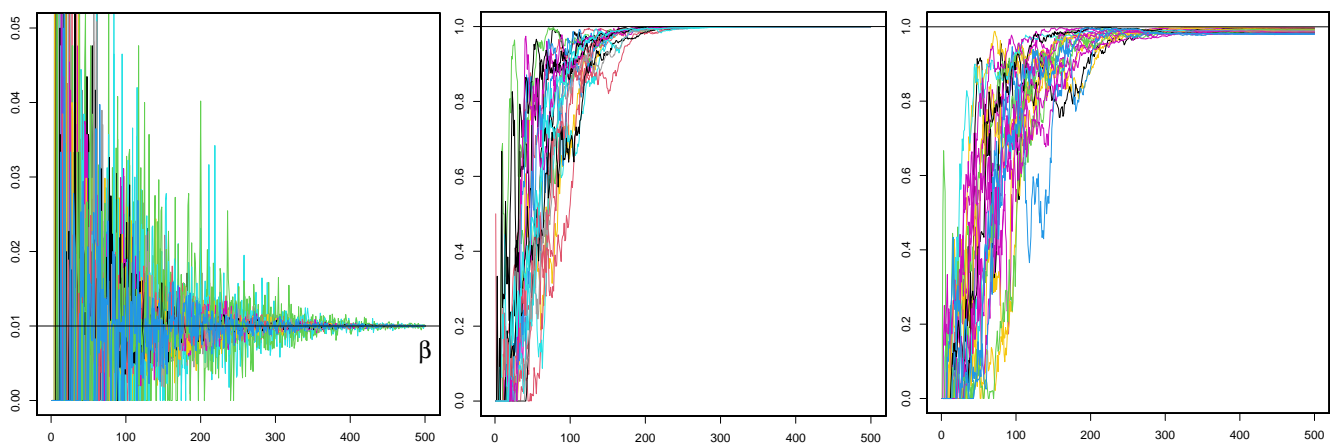
### 6. Illustrative Examples

In this section, the different asymptotic behaviours for an allele of a Y-linked gene and its mutations obtained in the previous sections are illustrated by means of a series of simulated examples. First, we justify the values of the parameters chosen in the simulations that follow. The sex ratio is well-known not to be balanced but closed to 0.5, being less that

0.5 in some situations (see for example [10,11]) and greater than 0.5 in others (see [12–14]). We fix  $\alpha = 0.45$ , since the case  $\alpha < 0.5$  is mathematically more interesting. For the mutation rate, we take  $\beta$  equal to 0.01. Although such value seems to be inflated related to actual values in nature as can be seen in [15], we consider that is adequate for the convenience of modelling the process. Moreover, with respect to the number of mating units at initial generation, we consider  $Z_0^R = 1$ , which could represent Y-chromosomal most recent common ancestor with the nonmutant original allele R (see for example [16–18]), and  $Z_0^r = 0$ , since mutations appear randomly along time. Finally, Poisson distributions are considered as reproduction laws. This type of distribution is frequently used as offspring distribution (see for example [19–24]). The reproduction mean for R-genotype is fixed taking a value of  $m_R = 2.3$  which is included in usual ranges for mammalian, for example in human (see the web page <https://datos.bancomundial.org/indicador/SP.DYN.TFRT.IN>). Those parameters verified that  $\alpha < 0.5$  and  $\alpha(1 - \beta)m_R > 1$ , and then there exists a positive probability of coexistence of both genotypes. In order to illustrate the different behaviours on this event, we consider the following three specific and real scenarios depending on  $m_r$ :

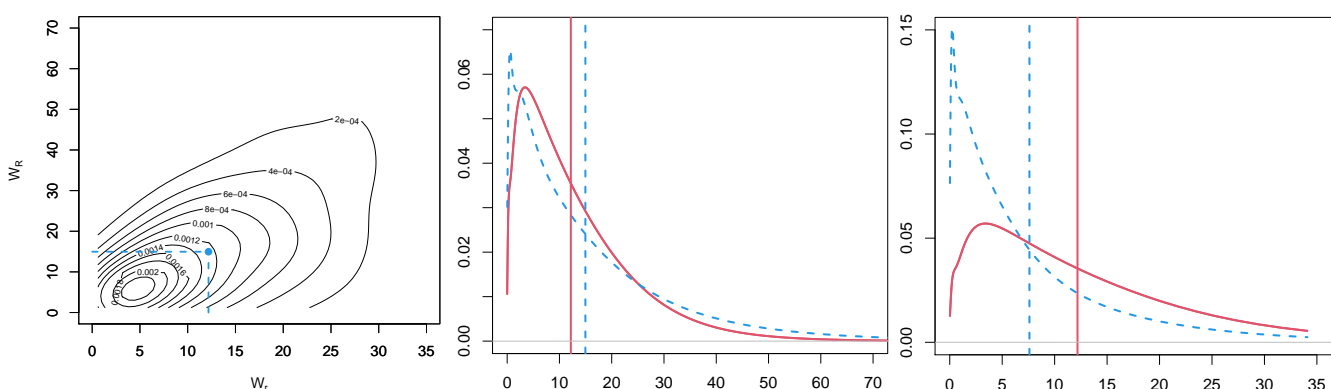
- In the first scenario, we consider  $m_r = 0$ , that is, the r-allele does not reproduce and, then, only appears in the population from R-mating units via mutation. This is the case of infertility in humans, where most of the cases turn out from new random deletions on the Y chromosome in the azoospermia factor regions in an affected individual's father who is not himself infertile (see [25] and the web page <https://medlineplus.gov/genetics/condition/y-chromosome-infertility/>). Notice that in this scenario,  $m_r < (1 - \beta)m_R$ .
- On the other hand, in the next two scenarios, we consider that  $m_r \neq 0$ , and then the r-allele may be transmitted by both types of mating units.
  - In the second scenario, we consider that accumulated mutations do not affect the reproductive capacity of r-mating units, and therefore  $m_r = m_R$ . This situation happens, for example, when the diversity of species is generated by accumulated mutations and not via natural selection (see [26,27]). Notice that in this scenario,  $m_r > (1 - \beta)m_R$ .
  - In the third scenario, mutations of mutations are considered deleterious, since accumulated mutations may drive to degeneration of Y chromosome (see [28–31]). Hence, only first mutations (may be different in every stance) of males stemming from R-type mating units are archived as r-allele and are transmitted to descendant. Notice that in this scenario it is reasonable to assume that  $m_r = (1 - \beta)m_R$  by considering that couples with males stemming from a first mutation of the R-allele have the same capacity of producing nonmutant males as an R-couple and therefore  $(1 - \alpha)m_r = (1 - \alpha)(1 - \beta)m_R$ .

In all three scenarios we simulated 20 paths of the process  $\{(Z_n^R, Z_n^r)\}_{n \geq 0}$  in which both genotypes have survived until generation 500. Figure 2 illustrates the limiting genotype frequency of r-allele, being equal to  $\beta$  for the first scenario and equal to one for last two scenarios. Then, we conclude that infertile males (first scenario) are present in male population as a proportion given by the mutation rate (see [32]). On the other hand, the original nonmutant allele is negligible with respect to the mutant allele in the other two scenarios. This statement is a mathematical explanation of the evolution and diversity of species without natural selection (see [26]) and the beginning of degeneration of Y chromosome (see [33]), where nonmutant original allele disappears along time.



**Figure 2.** Plot of  $Z_n^r / Z_n$ , for several paths on coexistence set of a process when  $m_r < (1 - \beta) m_R$  (left),  $m_r > (1 - \beta) m_R$  (middle) and  $m_r = (1 - \beta) m_R$  (right).

In the second scenario we also studied the relation between the limiting random variables  $W_R$ ,  $W_r^*$ , and  $\tilde{W}_r$ . As previously, we considered the process starting with  $(Z_0^R, Z_0^r) = (1, 0)$  couples and a parameter vector  $(\alpha, \beta, m_R, m_r)$  with values  $(0.45, 0.01, 2.3, 2.3)$ . To approximate the joint probability distribution of the random vector  $(W_r^*, W_R)$  we simulated 2000 paths of the process until generation 500 belonging to the coexistence set (i.e., both alleles were alive in the last observed generation). The distribution of the random vector of  $(Z_{500}^r, Z_{500}^R)$  with each coordinate normalised by the corresponding growth rate (to the 500th power) is considered a good enough approximation of the distribution of  $(W_r^*, W_R)$ . Figure 3, left and middle graphs, shows the estimates of the joint distribution of  $(W_r^*, W_R)$  and its marginal densities, respectively, obtained using kernel density estimators, as well as the mean values of those distributions,  $(12.178, 14.963)$ . We obtain a positive correlation between both random variables (Pearson’s correlation coefficient 0.664,  $p$ -value  $< 2.2 \times 10^{-16}$ ) and a statistically significant difference between their means (paired t-Student test:  $p$ -value  $< 2.2 \times 10^{-16}$ ). To estimate the distribution of  $\tilde{W}_r$ , we simulated again 2000 paths of the process until generation 500 but now belonging to the  $r$ -fixation set (i.e., such that  $Z_{500}^R = 0$  and  $Z_{500}^r > 0$ ). Figure 3 (right) compares the kernel density estimates of the distributions of  $\tilde{W}_r$  (dashed line) and  $W_r^*$  (solid line) as well as their mean values ( $E[\tilde{W}_r] \approx 7.597$  and  $E[W_r^*] \approx 12.178$ ). In this case we also found statistically significant differences between the mean values of both distributions (Welch’s test:  $p$ -value  $< 2.2 \times 10^{-16}$ ).



**Figure 3.** (Left) Contour plot showing the joint probability distribution of  $(W_r^*, W_R)$  and its mean vector. (Middle) Comparison of the kernel density estimates of the density functions of  $W_R$  (dashed line) and  $W_r^*$  (solid line), jointly with their mean values (vertical lines). (Right) Comparison of the kernel density estimates of the distributions of  $W_r^*$  (solid line) and  $\tilde{W}_r$  (dashed line), jointly with their mean values (vertical lines).



## 7. Concluding Remarks

This work deals with the problem of determining the limiting growth rates and the limiting genotype frequencies of two alleles named R and r of a Y-linked gene, allowing the possibility that the R-allele mutates into the r-form assuming a two-sex (male-female) population. Moreover, the limiting sex ratio of this population is also obtained. The model studied in this communication is the so-called Y-linked bisexual branching processes with mutations and blind choice of males defined in [5]. This model studies the evolution of the number of carriers of two alleles in a two-sex monogamous population under the assumption that the considered gene has no effect on the mating process and also considering the possibility that one of the allelic forms of the gene can mutate into the other form without allowing the possibility of reversion.

We derived the limiting growth rates of surviving genotypes as functions of the mean numbers of females and males generated by a couple and the mutation rate. In particular, on the r-fixation set such genotype evolves as a two-sex Galton–Watson branching process and therefore there exists a positive probability of survival of this genotype if the mean number of female and male offspring per r-couple are both greater than unity. In this case, the numbers of couples and males grow geometrically at the same growth rate which is given by the mean number of females (if the probability  $\alpha$  for an offspring to be female is less than or equal to 0.5) or males (if  $\alpha > 0.5$ ) per r-couple (that is,  $\alpha m_r$  and  $(1 - \alpha) m_r$ , respectively).

On the coexistence set, both, R and r-genotypes have a positive probability of survival if the mean number of female offspring per R-couple who mate with R-males and the mean number of male offspring per R-couple are both greater than unity. Our results show in this case that, the growth rates for the number of R-couples and males with R-genotype coincide on the event of survival. In particular, both variables grow geometrically, and the limiting growth rate equals the mean number of female offspring per R-couple who mate with R-males if  $\alpha \leq 0.5$  (that is,  $\alpha(1 - \beta) m_R$ , with  $\beta$  being the probability of mutation), whereas it equals the mean number of R-male offspring per R-couple if  $\alpha > 0.5$  (that is,  $(1 - \alpha)(1 - \beta) m_R$ ).

The growth rate for the number of couples and males of r-genotype on the coexistence set turns out to be more complicated due to the mutations and it is necessary to take into account the relation between the mean number of offspring per r-couple,  $m_r$ , and the mean number of offspring per R-couple,  $m_R$ , multiplied by  $(1 - \beta)$ . Notice that both quantities appeared in the previous growth rates, for r and R-allele, respectively, independently on  $\alpha$ . So, in the case that  $m_r > (1 - \beta) m_R$ , the numbers of r-couples and r-males grow geometrically at a growth rate given by the mean number of females or males (depending on whether  $\alpha \leq 0.5$  or  $\alpha > 0.5$ , respectively) per r-couple. Notice that, this behaviour is the same as the one obtained on the r-fixation set. On the other hand, in the case  $m_r < (1 - \beta) m_R$ , the growth rate of the numbers of r-couples and r-males equals the growth rate of the R-genotype, that is,  $\alpha(1 - \beta) m_R$  if  $\alpha \leq 0.5$  or  $(1 - \alpha)(1 - \beta) m_R$  if  $\alpha > 0.5$ . Finally, the case  $m_r = (1 - \beta) m_R$  is a boundary case in which we have proved that the number of couples and males with r-genotype along time is normalised by the sequence  $\{n(\min\{\alpha, (1 - \alpha)\} m_r)^n\}$ .

As a consequence of such results, r-genotype is the dominant one with limiting frequency unity on the event of coexistence if  $m_r \geq (1 - \beta) m_R$ , while  $m_r < (1 - \beta) m_R$  entails a balanced situations in the sense that there is not a dominant genotype and the limiting frequencies are constants which depend on  $m_R$ ,  $m_r$  and  $\beta$ . In particular, the higher the reproduction capacity of the r-genotype, the lower the limiting R-genotype frequency, being the maximum value given by the rate of nonmutation (that is  $1 - \beta$ ) when r-couples do not generate any descendants (that is,  $m_r = 0$ ). Finally, we found that the limiting female frequency equals the probability of being female in any case, and thus it does not depend on the Y-linked gene.

From a practical point of view, these mathematical results could give a scientific answer to questions of population genetics, as for example, the evolution of infertility in males

(see [25,32]), the existence of degeneration of Y chromosome (see [28–31]) or the possibility of evolution and diversity of species without natural selection (see [26,27]). In particular, we conclude that although infertility conditions not to be inherited ( $m_r = 0$ ), the frequency of this mutation in the population is not negligible along time (see [34–36]), and it is given by the mutation rate,  $\beta$  (this is also true when the males are not infertile, but their reproductive capacity is very small, i.e.,  $m_r = (1 - \beta) m_R$ ). Another conclusion is that the original nonmutant allele is negligible in the population with respect to mutations (the dominant allele) along time, when the reproductive capacity is not affected by mutations (i.e., natural selection does not work). Hence, accumulative mutations (mutation of mutation is archived,  $m_r = m_R$ ) may explain the possibility of diversity of species without natural selection (see [26]), while deleterious mutations (mutation of mutation is not archived,  $m_r = (1 - \beta) m_R$ ) could explain the beginning of the degeneration of Y chromosome (see [33,37]). Of course, the obtained results also explain the phenomenon of evolution by natural selection, since when there is a differential reproduction ( $m_r > m_R$ ), the r-allele is the dominant allele along time.

In conclusion, the limiting behaviour of Y-linked genes in a two-sex branching process with blind choice of males and mutations shown here turns out to be the main novelty of the present work. We note that this behaviour may be different from those obtained for other two-sex branching models in a genetic context (see, for example, [3] or [38]) and also different from those obtained in classical genetic models, in which no dominant genotype exists in the population due to the Hardy-Weinberg law. Moreover, the results of the paper are expected to be a valuable contribution to the theory of two-sex branching processes and to clarify well-known phenomena of population genetics in a mathematically rigorous way. The mathematical tools used to obtain the results of this paper also allow to solve the conjectures proposed in [39] relative to the limiting growth rates for an X-linked two-sex branching process. Such a model describes the evolution of the number of carriers of the genotypes and phenotypes defined by the alleles of an X-linked gene with two alleles when one of them is considered recessive and responsible for a serious disorder of a severe disease as could be hemophilia, Duchenne muscular dystrophy, or other genetic diseases (see [40] for background information).

**Author Contributions:** All Conceptualization, M.G., C.G. and R.M.; methodology, M.G., C.G. and R.M.; software, M.G., C.G. and R.M.; validation, M.G., C.G. and R.M.; formal analysis, M.G., C.G. and R.M.; investigation, M.G., C.G. and R.M.; resources, M.G., C.G. and R.M.; data curation, M.G., C.G. and R.M.; writing—original draft preparation, M.G., C.G. and R.M.; writing—review and editing, M.G., C.G. and R.M.; visualization, M.G., C.G. and R.M.; supervision, M.G., C.G. and R.M.; project administration, M.G., C.G. and R.M.; funding acquisition, M.G., C.G. and R.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministerio de Economía y Competitividad [grant MTM2015-70522-P], the Ministerio de Ciencia e Innovación [grant PID2019-108211GB-I00] and the Junta de Extremadura [grants IB16103 and GR18103] and the Fondo Europeo de Desarrollo Regional.

**Acknowledgments:** Gutiérrez would like to thank Gerold Alsmeyer for valuable discussions during her stay in the University of Münster. The authors are also very grateful to the anonymous referees for their comments and suggestions which have considerably improved this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Proofs

In order to prove the results of this paper, we introduce the following notation. We consider the filtrations  $\mathcal{G}_n = \sigma(Z_0^R, Z_0^r, F_k^R, M_k^R, F_k^r, M_k^r, Z_k^R, Z_k^r, k = 1, \dots, n)$ ,  $n \geq 1$  ( $\mathcal{G}_0 = \sigma(Z_0^R, Z_0^r)$ ) and  $\mathcal{F}_n = \sigma(\mathcal{G}_{n-1}, F_n^R, M_n^R, F_n^r, M_n^r)$ ,  $n \geq 1$ . Then, it is verified that  $\mathcal{G}_{n-1} \subseteq \mathcal{F}_n \subseteq \mathcal{G}_n$ ,  $n \geq 1$ . In addition, we write  $P_{(i,j)}(\cdot)$  for  $P(\cdot | Z_0^R = i, Z_0^r = j)$  and  $E_{(i,j)}[\cdot]$  for  $E[\cdot | Z_0^R = i, Z_0^r = j]$ , for any  $i, j \geq 0$ .

**Proof of Theorem 1.**

One infers from Lemmas A.1 and A.5 in [5] and Theorem 7.28 in [41] that

$$0 < \prod_{n=0}^{\infty} \frac{Z_{n+1}^R}{\tau_R Z_n^R} < \infty \text{ a.s. on } \{Z_n^R \rightarrow \infty\},$$

which gives the desired result, since for each  $N \geq 1$ ,

$$\frac{Z_N^R}{\tau_R^N} = Z_0^R \prod_{n=0}^{N-1} \frac{Z_{n+1}^R}{\tau_R Z_n^R}.$$

Moreover, taking into account Lemma A.3 in [5], the result holds also true for  $M_n^R$ , which concludes the proof.  $\square$

**Proof of Theorem 2.**

As  $\alpha > 0.5$  and  $(1 - \alpha)(1 - \beta) m_R > 1$ , by Corollary A.1 in [5] and the definition of the model, we have that

$$\{Z_n^R \rightarrow \infty\} \subset \{M_n < F_n \text{ eventually}\} \subset \{Z_n^R = M_n^R, Z_n^r = M_n^r \text{ eventually}\} \text{ a.s.}$$

Therefore, for almost all paths in  $\{Z_n^R \rightarrow \infty\}$ , there exists a constant depending on path,  $K \in \mathbb{N}$ , such that, a.s. on  $\{Z_n^R \rightarrow \infty\}$ , we can write (1) in the following way,

$$(M_{n+1}^R, M_{n+1}^r) = \left( \sum_{i=1}^{M_n^R} M_{ni}^R, \sum_{i=1}^{M_n^R} M_{ni}^{R \rightarrow r} + \sum_{j=1}^{M_n^r} M_{nj}^{r \rightarrow r} \right), \text{ for } n > K.$$

Then, defining the variables  $M_{ni}^{R \rightarrow R} = M_{ni}^R$  and  $M_{ni}^{r \rightarrow R} = 0$ , for all  $(n, i) \in \mathbb{N}_0 \times \mathbb{N}$ , we can rewrite

$$(M_{n+1}^R, M_{n+1}^r) = \sum_{x \in \{R, r\}} \sum_{i=1}^{M_n^x} (M_{ni}^{x \rightarrow R}, M_{ni}^{x \rightarrow r}), \text{ for } n > K.$$

Hence,  $\{(Z_n^R, Z_n^r)\}_{n > K}$  turns out to be like a path of a reducible two-type branching process where the first component can give birth individuals of two types while the second component can only give birth individuals of its own type. Taking into account the relation between both types, the mean matrix associate with this multitype process takes the following form:

$$(1 - \alpha) \begin{pmatrix} (1 - \beta) m_R & \beta m_R \\ 0 & m_r \end{pmatrix}.$$

This is a reducible matrix with two associated eigenvalues  $\lambda_1 = (1 - \alpha)(1 - \beta) m_R$ —with left and right eigenvectors  $v_1 = (1, \gamma)$  and  $v_2 = (1, 0)'$ , respectively—and  $\lambda_2 = (1 - \alpha) m_r$ —with left and right eigenvectors  $\mu_1 = (0, 1)$  and  $\mu_2 = (-\gamma, 1)'$  respectively, with  $\gamma = \beta m_R ((1 - \beta) m_R - m_r)^{-1}$ , when  $(1 - \beta) m_R \neq m_r$ .

Therefore, the result follows directly applying Theorem 2.1 (for (i) and (iii)) and Theorem 2.3 (for (ii)) in [42]  $\square$

**Auxiliary results for the proof of Theorem 3.**

Let  $\alpha \leq 0.5$  and  $\alpha(1 - \beta) m_R > 1$ . Given  $\varepsilon > 0$ , we define

$$\begin{aligned} \eta_1 &= \alpha(1 - \beta)(m_R - \varepsilon)(1 - 3\varepsilon/m), \text{ with } m = (m_R \wedge m_r)(1 - \delta_{m_r,0}) + m_R \delta_{m_r,0} + \varepsilon \\ \eta_2 &= \alpha(1 - \beta)(m_R + \varepsilon)(1 + 3\varepsilon/m'), \text{ with } m' = (m_R \wedge m_r)(1 - \delta_{m_r,0}) + m_R \delta_{m_r,0} - \varepsilon \\ \eta_3 &= (1 - \delta_{m_r,0})\alpha(m_r - \varepsilon)(1 - 3\varepsilon/m) \\ \eta_4 &= \alpha(\beta m_R - \varepsilon)(1 - 3\varepsilon/m) \\ \eta_5 &= (1 - \delta_{m_r,0})\alpha(m_r + \varepsilon)(1 + 3\varepsilon/m') \\ \eta_6 &= \alpha(\beta m_R + \varepsilon)(1 + 3\varepsilon/m'), \end{aligned}$$

where  $\delta_{m,r,0}$  is the Kronecker delta. Trivially, there exists  $a > 0$  such that, for every  $\varepsilon \in (0, a)$ , one can easily verify

$$1 < \eta_1 < \eta_2, \quad 0 \leq \eta_3 \leq \eta_5 \quad \text{and} \quad 0 < \eta_4 < \eta_6.$$

For any of those  $\varepsilon$ , we define, for  $n = 0, 1, \dots$

$$A_n^\varepsilon = \{\eta_1 Z_n^R < Z_{n+1}^R < \eta_2 Z_n^R, \eta_3 Z_n^r + \eta_4 Z_n^R < Z_{n+1}^r < \eta_5 Z_n^r + \eta_6 Z_n^R\}.$$

□

**Theorem A1.** *Let  $\alpha \leq 0.5$  and  $\alpha(1 - \beta) m_R > 1$ . Fixed  $\varepsilon > 0$  such that  $\eta_1 > 1, \eta_3 \geq 0$  and  $\eta_4 > 0$ , it is verified that*

$$\liminf_{n \rightarrow \infty} A_n^\varepsilon = \{Z_n^R \rightarrow \infty\} \text{ a.s.}$$

**Proof.**

Since  $\eta_1 > 1$ , then  $\liminf_{n \rightarrow \infty} A_n^\varepsilon \subseteq \{Z_n^R \rightarrow \infty\}$  a.s., and therefore it is enough to prove that  $P(\liminf_{n \rightarrow \infty} A_n^\varepsilon) \geq P(Z_n^R \rightarrow \infty)$ . To this end, for any  $N \geq 1$ , we define the stopping time  $T(N) = \min\{n : Z_n^r \wedge Z_n^R \geq N\}$ , where  $T(N) = \infty$  if  $Z_n^R < N$  or  $Z_n^r < N$  for every  $n \geq 0$ . Obviously,  $\{Z_n^R \rightarrow \infty\} \subseteq \{T(N) < \infty\}$  a.s., for each  $N \geq 1$ , since  $\{Z_n^R \rightarrow \infty\} = \{Z_n^R \rightarrow \infty, Z_n^r \rightarrow \infty\}$  a.s.

Given  $k \geq 0$ , note that  $\{T(N) = k\} = \{Z_k^r \wedge Z_k^R \geq N, Z_n^r \wedge Z_n^R < N, n = 0, \dots, k-1\}$ , and taking into account that  $\{(Z_n^R, Z_n^r)\}_n$  is a Markov chain, one has that

$$\begin{aligned} P\left(\bigcap_{n=k}^{\infty} A_n^\varepsilon \mid T(N) = k\right) &= P\left(\bigcap_{n=k}^{\infty} A_n^\varepsilon \mid Z_k^r \wedge Z_k^R \geq N\right) \\ &\geq \inf_{i,j \geq N} P\left(\bigcap_{n=k}^{\infty} A_n^\varepsilon \mid Z_k^R = i, Z_k^r = j\right) \\ &= \inf_{i,j \geq N} P\left(\bigcap_{n=0}^{\infty} A_n^\varepsilon \mid Z_0^R = i, Z_0^r = j\right) = \inf_{i,j \geq N} P_{(i,j)}\left(\bigcap_{n=0}^{\infty} A_n^\varepsilon\right). \end{aligned}$$

Therefore, one obtains that

$$\begin{aligned} P(\liminf_{n \rightarrow \infty} A_n^\varepsilon) &= P\left(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n^\varepsilon\right) \geq P\left(\bigcap_{n=T(N)}^{\infty} A_n^\varepsilon\right) \\ &= \sum_{k=0}^{\infty} P\left(\bigcap_{n=k}^{\infty} A_n^\varepsilon \mid T(N) = k\right) P(T(N) = k) \\ &\geq \inf_{i,j \geq N} P_{(i,j)}\left(\bigcap_{n=0}^{\infty} A_n^\varepsilon\right) P(T(N) < \infty) \\ &\geq \inf_{i,j \geq N} P_{(i,j)}\left(\bigcap_{n=0}^{\infty} A_n^\varepsilon\right) P(Z_n^R \rightarrow \infty). \end{aligned}$$

Hence, to conclude, it is enough to prove that, for all  $n \geq 0$ ,

$$P((A_n^\varepsilon)^c \mid \mathcal{G}_n) \leq Ce^{-BZ_n^R} + \frac{K}{Z_n^R}, \text{ a.s. on } \{Z_n^R > M\}, \tag{A1}$$

for some suitable positive constants  $C, B, K$  and  $M$ . Indeed,

$$\begin{aligned} P_{(ij)}\left(\bigcup_{n=0}^{\infty} (A_n^\varepsilon)^c\right) &= \sum_{n=0}^{\infty} P_{(ij)}\left((A_n^\varepsilon)^c \cap A_{n-1}^\varepsilon \cap \dots \cap A_0^\varepsilon\right) \\ &= \sum_{n=0}^{\infty} E_{(ij)}\left[I_{A_{n-1}^\varepsilon \cap \dots \cap A_0^\varepsilon} P\left((A_n^\varepsilon)^c \mid \mathcal{G}_n\right)\right] \\ &\leq C \sum_{n=0}^{\infty} e^{-B i \eta_1^n} + \frac{K}{i} \sum_{n=0}^{\infty} \eta_1^{-n}, \end{aligned}$$

and therefore  $\lim_{i,j \rightarrow \infty} P_{(ij)}\left(\bigcup_{n=0}^{\infty} (A_n^\varepsilon)^c\right) = 0$ , since  $\eta_1 > 1$ .

In order to prove (A1), we decompose the set  $(A_n^\varepsilon)^c$  as follows,

$$\begin{aligned} (A_n^\varepsilon)^c &= \{Z_{n+1}^R \leq \eta_1 Z_n^R\} \cup \{Z_{n+1}^R \geq \eta_2 Z_n^R\} \\ &\cup \{Z_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R\} \cup \{Z_{n+1}^r \geq \eta_5 Z_n^r + \eta_6 Z_n^R\}. \end{aligned}$$

Then, it is enough to prove that the probabilities of each one of these sets are upper bounded by the same bound given in (A1). Since the study of each one of these probabilities can be developed in the same manner, we confine ourselves to analyse  $P(Z_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R \mid \mathcal{G}_n)$ , as both genotypes are implied. For that, we define, for  $n = 1, 2, \dots$

$$\begin{aligned} A_{F_n} &= \{ |F_n - (\alpha m_R Z_{n-1}^R + \alpha m_r Z_{n-1}^r)| \leq \alpha \varepsilon (Z_{n-1}^R + (1 - \delta_{m_r,0}) Z_{n-1}^r) \}, \\ A_{M_n} &= \{ |M_n - ((1 - \alpha) m_R Z_{n-1}^R + (1 - \alpha) m_r Z_{n-1}^r)| \leq (1 - \alpha) \varepsilon (Z_{n-1}^R + (1 - \delta_{m_r,0}) Z_{n-1}^r) \}, \\ A_{M_n^r} &= \{ |M_n^r - ((1 - \alpha) \beta m_R Z_{n-1}^R + (1 - \alpha) m_r Z_{n-1}^r)| \leq (1 - \alpha) \varepsilon (Z_{n-1}^R + (1 - \delta_{m_r,0}) Z_{n-1}^r) \}, \end{aligned}$$

and write

$$\begin{aligned} \{Z_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R\} &\subseteq \{Z_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R, M_{n+1}^r > \eta_3 Z_n^r + \eta_4 Z_n^R, M_{n+1} \leq F_{n+1}\} \\ &\cup \{M_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R\} \cup D_1^c \\ &\cup (\{Z_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R, M_{n+1} > F_{n+1}\} \cap D_1), \end{aligned} \tag{A2}$$

with  $D_1 = A_{F_{n+1}} \cap A_{M_{n+1}} \cap A_{M_{n+1}^r}$ . By the definition of the model, it is clear that the first set has null probability. The probability of the second set is bounded taking into account that  $\alpha \leq (1 - \alpha)$ , the reproduction laws are assumed to have finite variances and the conditional Chebyshev's inequality:

$$\begin{aligned} &P(M_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R \mid \mathcal{G}_n) \\ &\leq P(M_{n+1}^r - (1 - \delta_{m_r,0})(1 - \alpha)m_r Z_n^r - (1 - \alpha)\beta m_R Z_n^R \leq (1 - \delta_{m_r,0})\alpha(m_r - \varepsilon)(1 - 3\varepsilon/m)Z_n^r \\ &\quad - (1 - \delta_{m_r,0})\alpha m_r Z_n^r + \alpha(\beta m_R - \varepsilon)(1 - 3\varepsilon/m)Z_n^R - \alpha\beta m_R Z_n^R \mid \mathcal{G}_n) \\ &\leq P(|M_{n+1}^r - (1 - \delta_{m_r,0})(1 - \alpha)m_r Z_n^r - (1 - \alpha)\beta m_R Z_n^R| \geq b((1 - \delta_{m_r,0})Z_n^r + Z_n^R) \mid \mathcal{G}_n) \\ &\leq \frac{\text{Var}(M_{n+1}^r \mid \mathcal{G}_n)}{b^2((1 - \delta_{m_r,0})Z_n^r + Z_n^R)^2} \leq \frac{K_1}{Z_n^R} \text{ a.s. on } \{Z_n^R > 0\}, \end{aligned}$$

for certain positive constants  $b$  and  $K_1$ . In a similar way,  $P(A_{F_{n+1}}^c \mid \mathcal{G}_n)$ ,  $P(A_{M_{n+1}}^c \mid \mathcal{G}_n)$  and  $P(A_{M_{n+1}^r}^c \mid \mathcal{G}_n)$  are bounded, and therefore we deduce that  $P(D_1^c \mid \mathcal{G}_n)$  is also bounded a.s. on  $\{Z_n^R > 0\}$  by  $K_2 / Z_n^R$ , for some positive constant  $K_2$ .

Finally, we deal with the probability of the last set in (A2). Notice that,  $Z_{n+1}^r \mid \mathcal{F}_{n+1} \sim \text{Hyper}(F_{n+1}, M_{n+1}, M_{n+1}^r)$  on  $\{M_{n+1} > F_{n+1}\}$ , therefore  $E[Z_{n+1}^r \mid \mathcal{F}_{n+1}] = \frac{F_{n+1}}{M_{n+1}} M_{n+1}^r$  a.s. on that set. Moreover, it is verified a.s. on  $D_1$  that,



$$\begin{aligned} & \frac{F_{n+1}}{M_{n+1}} M_{n+1}^r \\ & \geq \alpha((\beta m_R - \varepsilon) Z_n^R + (1 - \delta_{m_r,0})(m_r - \varepsilon) Z_n^r) \left( 1 - \frac{2\varepsilon(Z_n^R + (1 - \delta_{m_r,0}) Z_n^r)}{(m_R + \varepsilon) Z_n^R + (m_r + \varepsilon)(1 - \delta_{m_r,0}) Z_n^r} \right) \\ & \geq \alpha((\beta m_R - \varepsilon) Z_n^R + (1 - \delta_{m_r,0})(m_r - \varepsilon) Z_n^r) \left( 1 - \frac{2\varepsilon}{m} \right), \end{aligned}$$

and then, on  $\{M_{n+1} > F_{n+1}\} \cap D_1$ , it is verified a.s. that,

$$\eta_3 Z_n^r + \eta_4 Z_n^R - E[Z_{n+1}^r | \mathcal{F}_{n+1}] \leq -\frac{\varepsilon}{m} \alpha((1 - \delta_{m_r,0})(m_r - \varepsilon) Z_n^r + (\beta m_R - \varepsilon) Z_n^R).$$

Now, applying the bounds for the tails of a hypergeometric distribution provided in [43], we deduce that, a.s. on  $\{M_{n+1} > F_{n+1}\} \cap D_1$  and  $\{Z_n^R > M\}$ , for some suitable positive constants  $C, B$  and  $M$ ,

$$\begin{aligned} & P(Z_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R | \mathcal{F}_{n+1}) \\ & \leq P\left(Z_{n+1}^r - E[Z_{n+1}^r | \mathcal{F}_{n+1}] \leq -\frac{\varepsilon}{m} \alpha((1 - \delta_{m_r,0})(m_r - \varepsilon) Z_n^r + (\beta m_R - \varepsilon) Z_n^R) | \mathcal{F}_{n+1}\right) \\ & \leq \exp\left\{-2 \frac{(\alpha\varepsilon/m((1 - \delta_{m_r,0})(m_r - \varepsilon) Z_n^r + (\beta m_R - \varepsilon) Z_n^R))^2 - 1}{M_{n+1}^r + 1}\right\} \\ & \leq \exp\left\{-2 \frac{(\alpha\varepsilon/m((1 - \delta_{m_r,0})(m_r - \varepsilon) Z_n^r + (\beta m_R - \varepsilon) Z_n^R))^2 - 1}{(1 - \alpha)((1 - \delta_{m_r,0})(m_r + \varepsilon) Z_n^r + (\beta m_R + \varepsilon) Z_n^R) + 1}\right\} \\ & \leq C e^{-B((1 - \delta_{m_r,0}) Z_n^r + Z_n^R)} \\ & \leq C e^{-B Z_n^R}. \end{aligned}$$

Then, a.s. on  $\{Z_n^R > M\}$ ,

$$\begin{aligned} & P(\{Z_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R, M_{n+1} > F_{n+1}\} \cap D_1 | \mathcal{G}_n) \\ & = E[P(Z_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R | \mathcal{F}_{n+1}) I_{\{M_{n+1} > F_{n+1}\} \cap D_1} | \mathcal{G}_n] \\ & \leq C e^{-B Z_n^R}. \end{aligned}$$

Finally, combining all the above probabilities, we obtain that,

$$P(Z_{n+1}^r \leq \eta_3 Z_n^r + \eta_4 Z_n^R | \mathcal{G}_n) \leq C e^{-B Z_n^R} + \frac{K}{Z_n^R}, \text{ a.s. on } \{Z_n^R > M\},$$

and the proof is completed.  $\square$

From previous theorem, the limit of the quotient of the total number of couples of each type is established in the following result, when  $m_r \neq (1 - \beta) m_R$ .

**Theorem A2.** *Let us assume  $\alpha \leq 0.5$  and  $\alpha(1 - \beta) m_R > 1$ .*

- (i) *If  $m_r > (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{Z_n^R} = \infty$ , a.s. on  $\{Z_n^R \rightarrow \infty\}$ .*
- (ii) *If  $m_r < (1 - \beta) m_R$ , then  $\lim_{n \rightarrow \infty} \frac{Z_n^r}{Z_n^R} = \gamma$ , a.s. on  $\{Z_n^R \rightarrow \infty\}$ ,*

*with  $\gamma$  as in Theorem 2.*

**Proof.**

For each  $k \geq 0$  and  $n \geq k$ , iterating, it is verified that a.s. on  $\bigcap_{n=k}^{\infty} A_n^\varepsilon$

$$\begin{aligned} \frac{\eta_4}{\eta_2} \sum_{l=1}^{n-k} \left(\frac{\eta_3}{\eta_2}\right)^{l-1} + \frac{\eta_3^{n-k}}{\eta_2^{n-k}} \frac{Z_k^r}{Z_k^R} &\leq \frac{\eta_4}{\eta_2} + \frac{\eta_3}{\eta_2} \frac{Z_{n-1}^r}{Z_{n-1}^R} \\ &\leq \frac{Z_n^r}{Z_n^R} \leq \frac{\eta_6}{\eta_1} + \frac{\eta_5}{\eta_1} \frac{Z_{n-1}^r}{Z_{n-1}^R} \\ &\leq \frac{\eta_6}{\eta_1} \sum_{l=1}^{n-k} \left(\frac{\eta_5}{\eta_1}\right)^{l-1} + \frac{\eta_5^{n-k}}{\eta_1^{n-k}} \frac{Z_k^r}{Z_k^R}. \end{aligned}$$

When  $m_r > (1 - \beta) m_R$ , it is possible to choose  $\varepsilon$  small enough, such that  $\eta_2 < \eta_3$ , and then the result is derived taking into account Theorem A1. Moreover, when  $m_r < (1 - \beta) m_R$ , we take  $\varepsilon$  small enough such that  $0 \leq \eta_3 \leq \eta_5 < \eta_1 < \eta_2$ . Then, from Theorem A1 we deduce that a.s. on  $\{Z_n^R \rightarrow \infty\}$ ,

$$\frac{\eta_4}{\eta_2 - \eta_3} \leq \liminf_{n \rightarrow \infty} \frac{Z_n^r}{Z_n^R} \leq \limsup_{n \rightarrow \infty} \frac{Z_n^r}{Z_n^R} \leq \frac{\eta_6}{\eta_1 - \eta_5},$$

which gives the desired result, when  $\varepsilon$  goes to zero.  $\square$

Moreover, for the case  $m_r > (1 - \beta) m_R$ , the following result is established.

**Lemma A1.** *If  $\alpha \leq 0.5$ ,  $\alpha(1 - \beta) m_R > 1$  and  $m_r > (1 - \beta) m_R$ , then*

$$\liminf_{n \rightarrow \infty} \frac{Z_{n+1}^r}{Z_n^r} > 1 \text{ and } \lim_{n \rightarrow \infty} \frac{M_{n+1}^{R \rightarrow r}}{Z_n^r} = 0 \text{ a.s. on } \{Z_n^R \rightarrow \infty\}.$$

Moreover, for each  $0 < \rho < 1/2$ , a.s. on  $\{Z_n^R \rightarrow \infty\}$ , as  $n \rightarrow \infty$ ,

$$\begin{aligned} \frac{M_{n+1}^{R \rightarrow r}}{Z_n^r} &= (1 - \alpha) m_r + O((Z_n^r)^{-\rho}), \quad \frac{Z_n^r}{M_n^r} = \frac{\alpha}{1 - \alpha} + O((Z_{n-1}^r)^{-\rho}), \\ \frac{M_{n+1}^r}{Z_n^r} &= (1 - \alpha) m_r + O((Z_n^r)^{-\rho}) \text{ and } \frac{Z_{n+1}^r}{Z_n^r} = \alpha m_r + O((Z_n^r)^{-\rho}). \end{aligned}$$

**Proof.** Since  $m_r > (1 - \beta) m_R$  and  $\alpha(1 - \beta) m_R > 1$ , it is possible to choose  $\varepsilon$  small enough, such that  $1 < \eta_3$  and  $0 < \eta_4$ . For each  $k \geq 0$  and  $n \geq k$ , iterating, it is verified that, a.s. on  $\bigcap_{n=k}^{\infty} A_n^\varepsilon$ ,

$$\frac{Z_{n+1}^r}{Z_n^r} \geq \frac{\eta_3 Z_n^r + \eta_4 Z_n^R}{Z_n^r} = \eta_3 + \eta_4 \frac{Z_n^R}{Z_n^r},$$

and the first limit is deduced from Theorems A1 and A2. In addition, we conclude the second limit by Lemma A.3 in [5] and Theorem A2, since

$$\frac{M_{n+1}^{R \rightarrow r}}{Z_n^r} = \frac{M_{n+1}^{R \rightarrow r} Z_n^R}{Z_n^R Z_n^r}.$$

Moreover, taking into account the first limit, the proofs of the third and fourth statements follow the same ideas as Lemma 5 in [3] and Lemma A.4 in [5], respectively, and they are therefore omitted. Finally, the last two statements are directly deduced, since

$$M_{n+1}^r = M_{n+1}^{R \rightarrow r} + M_{n+1}^{r \rightarrow r} \text{ and } \frac{Z_{n+1}^r}{Z_n^r} = \frac{Z_{n+1}^r}{M_{n+1}^r} \frac{M_{n+1}^r}{Z_n^r}.$$

$\square$

In order to prove the Theorem 3 for the case  $m_r = (1 - \beta) m_R$ , it is necessary a more detailed description of the model. To this end, we introduce a partition of  $Z_n^r$ , that is, a sequence  $\{(Z_n^{r,0}, \dots, Z_n^{r,n})\}_{n \geq 0}$ , such that  $Z_n^r = Z_n^{r,0} + \dots + Z_n^{r,n}$  a.s., where  $Z_n^{r,l}$  denotes the number of  $r$ -type mating units at generation  $n$  whose males have their first ancestor of  $r$ -type at generation  $l$ ,  $0 \leq l \leq n$ ,  $n \geq 0$ . Let  $Z_0^{r,0} = Z_0^r$  and let  $\{(F_{nj}^{r,l}, M_{nj}^{r \rightarrow r,l})\}_{n,l \geq 0, j \geq 1}$  be i.i.d. random vectors with the same distribution as  $(F_{01}^r, M_{01}^{r \rightarrow r})$ . Given  $n \geq 0$  and the vector  $(Z_n^{r,0}, \dots, Z_n^{r,n})$ , it is verified a.s. that

$$F_{n+1} = \sum_{i=1}^{Z_n^R} F_{ni}^R + \sum_{l=0}^n \sum_{j=1}^{Z_n^{r,l}} F_{nj}^{r,l} \quad \text{and} \quad M_{n+1}^{r \rightarrow r} = \sum_{l=0}^n M_{n+1}^{r \rightarrow r,l},$$

where  $M_{n+1}^{r \rightarrow r,l} = \sum_{j=1}^{Z_n^{r,l}} M_{nj}^{r \rightarrow r,l}$ ,  $l = 0, \dots, n$ .

Notice that  $M_{n+1}^{r \rightarrow r,l}$  denotes the number of  $r$ -type males at generation  $n + 1$  generated by  $Z_n^{r,l}$  mating units.

Given  $(F_{n+1}, M_{n+1}, M_{n+1}^{r \rightarrow r,0}, \dots, M_{n+1}^{r \rightarrow r,n}, M_{n+1}^{R \rightarrow r})$ , if  $F_{n+1} \geq M_{n+1}$ , let  $Z_{n+1}^{r,l} = M_{n+1}^{r \rightarrow r,l}$  for  $l = 0, \dots, n$ , and  $Z_{n+1}^{r,n+1} = M_{n+1}^{R \rightarrow r}$ . On the other hand, if  $F_{n+1} < M_{n+1}$ ,  $(Z_{n+1}^{r,0}, \dots, Z_{n+1}^{r,n+1})$  follows a multivariate hypergeometric distribution with parameters  $(F_{n+1}, M_{n+1}, M_{n+1}^{r \rightarrow r,0}, \dots, M_{n+1}^{r \rightarrow r,n}, M_{n+1}^{R \rightarrow r})$ . Hence, it is verified that  $Z_{n+1}^r = Z_{n+1}^{r,0} + \dots + Z_{n+1}^{r,n+1}$  a.s. Furthermore, it is easy to prove that  $\{(Z_n^R, Z_n^{r,0}, \dots, Z_n^{r,n})\}_{n \geq 0}$  is a Markov chain. In comparison with the model description of Section 2, we only relabel the vectors  $\{(F_{nj}^r, M_{nj}^{r \rightarrow r})\}_{n \geq 0, j \geq 1}$ , consider new partial sums,  $M_{n+1}^{r \rightarrow r,l}$ , for  $n \geq 0$  and  $l = 0, \dots, n$ , and count new subtypes of  $r$ -type mating units. Intuitively, we observe the model in a more precise way.

Now, let  $\varepsilon > 0$ , such that  $\eta_1 > 1$  and  $\eta_4 > 0$ . For any of these  $\varepsilon$ , we define, for  $k = 1, 2, \dots$  and  $n \geq k - 1$ ,

$$A_{n,k}^\varepsilon = \{\eta_1 Z_n^R < Z_{n+1}^R < \eta_2 Z_n^R\} \cap \{\eta_1 \bar{Z}_n^{r,k-1} \leq \bar{Z}_{n+1}^{r,k-1} \leq \eta_2 \bar{Z}_n^{r,k-1}\} \\ \bigcap_{l=k}^n \{\eta_1 Z_n^{r,l} < Z_{n+1}^{r,l} < \eta_2 Z_n^{r,l}\} \cap \{\eta_4 Z_n^R < Z_{n+1}^{r,n+1} < \eta_6 Z_n^R\},$$

where  $\bar{Z}_n^{r,k-1} = Z_n^{r,0} + \dots + Z_n^{r,k-1}$ . Notice that  $\bar{Z}_n^{r,k-1}$  may be null. Also, for each  $k = 1, 2, \dots$ , let

$$B_k^\varepsilon = \bigcap_{n=k-1}^\infty A_{n,k}^\varepsilon.$$

Since  $\bar{Z}_n^{r,k} = \bar{Z}_n^{r,k-1} + Z_n^{r,k}$ , then it is easy to deduce that  $B_k^\varepsilon \subseteq B_{k+1}^\varepsilon$  a.s. and therefore

$$\lim_{k \rightarrow \infty} B_k^\varepsilon = \bigcup_{k=1}^\infty \bigcap_{n=k-1}^\infty A_{n,k}^\varepsilon \text{ a.s.}$$

Moreover, the following result holds.

**Theorem A3.** *Let us assume  $\alpha \leq 0.5$ ,  $\alpha(1 - \beta) m_R > 1$  and  $m_r = (1 - \beta) m_R$ . Fixed  $\varepsilon > 0$  such that  $\eta_1 > 1$ , it is verified that*

$$\lim_{k \rightarrow \infty} B_k^\varepsilon = \{Z_n^R \rightarrow \infty\} \text{ a.s.}$$

**Proof.**

Fixed  $k \geq 1$ ,  $N \geq 1$  and  $\varepsilon > 0$ , such that  $\eta_1 > 1$ , defining  $T(N)$  as in the proof of Theorem A1 and taking into account that  $\bar{Z}_{k-1}^{r,k-1} = Z_{k-1}^r$  and  $\{(Z_n^R, Z_n^{r,0}, \dots, Z_n^{r,n})\}_{n \geq 0}$  is a Markov chain, then we deduce that

$$\begin{aligned}
 P(B_k^\epsilon | T(N) = k-1) &= P(B_k^\epsilon | Z_{k-1}^R \wedge Z_{k-1}^r \geq N) \\
 &\geq \inf_{i,j \geq N} P(B_k^\epsilon | Z_{k-1}^R = i, Z_{k-1}^r = j) \\
 &= \inf_{i,j \geq N} P(B_1^\epsilon | Z_0^R = i, Z_0^r = j) = \inf_{i,j \geq N} P_{(i,j)} \left( \prod_{n=0}^{\infty} A_{n,1}^\epsilon \right).
 \end{aligned}$$

Following similar arguments as in the proof of Theorem A1, we derive that

$$\begin{aligned}
 P_{(i,j)} \left( \bigcup_{n=0}^{\infty} (A_{n,1}^\epsilon)^c \right) &\leq C \sum_{n=0}^{\infty} e^{-Bi\eta_1^n} + \frac{K}{i} \sum_{n=0}^{\infty} \eta_1^{-n} + C \sum_{n=0}^{\infty} e^{-Bj\eta_1^n} + \frac{K}{j} \sum_{n=0}^{\infty} \eta_1^{-n} \\
 &\quad + C \sum_{n=0}^{\infty} (n+1)e^{-Bi\eta_1^{n-1}\eta_4} + \frac{K}{i} \sum_{n=0}^{\infty} (n+1)\eta_1^{-n+1}\eta_4^{-1},
 \end{aligned}$$

for some suitable positive constants C, B, K and M, and the proof is concluded. □

The following lemma can be obtained from the previous result and applying similar arguments as in the proof of Lemma A1.

**Lemma A2.** *If  $\alpha \leq 0.5$ ,  $\alpha(1 - \beta) m_R > 1$  and  $m_r = (1 - \beta) m_R$ , fixed  $l \geq 1$ , then*

$$\liminf_{n \rightarrow \infty} \frac{Z_{n+1}^{r,l}}{Z_n^{r,l}} > 1 \quad \text{a.s. on } \{Z_n^R \rightarrow \infty\}.$$

Moreover, for each  $0 < \rho < 1/2$ , a.s. on  $\{Z_n^R \rightarrow \infty\}$ , as  $n \rightarrow \infty$ ,

$$\begin{aligned}
 \frac{M_{n+1}^{r,l}}{Z_n^{r,l}} &= (1 - \alpha) m_r + O((Z_n^{r,l})^{-\rho}), & \frac{Z_n^{r,l}}{M_n^{r,l}} &= \frac{\alpha}{1 - \alpha} + O((Z_{n-1}^{r,l})^{-\rho}), \\
 \frac{Z_{n+1}^{r,l}}{Z_n^{r,l}} &= \alpha m_r + O((Z_n^{r,l})^{-\rho}), & \frac{Z_n^{r,n}}{M_n^{r,n}} &= \frac{\alpha}{1 - \alpha} + O((Z_{n-1}^R)^{-\rho})
 \end{aligned}$$

and

$$\frac{Z_n^{r,n}}{Z_{n-1}^R} = \frac{\beta}{1 - \beta} \alpha m_r + O((Z_{n-1}^R)^{-\rho}).$$

**Proof of Theorem 3.**

(i) First, we consider the case  $m_r > (1 - \beta) m_R$ . Since

$$\frac{Z_N^R}{(\alpha m_r)^N} = Z_0^r \prod_{n=0}^{N-1} \frac{Z_{n+1}^r}{\alpha m_r Z_n^r}$$

for each  $N \geq 1$ , one infers from Lemma A1 and Theorem 7.28 in [41] that

$$0 < \prod_{n=0}^{\infty} \frac{Z_{n+1}^r}{\alpha m_r Z_n^r} < \infty \quad \text{a.s. on } \{Z_n^R \rightarrow \infty\},$$

which gives the desired results.

The result holds also true for  $M_n^r$ , from Lemma A1 and taking into account that

$$\frac{M_{n+1}^r}{(\alpha m_r)^n} = \frac{M_{n+1}^r}{Z_n^r} \frac{Z_n^r}{(\alpha m_r)^n}.$$

- (ii) Secondly, we deal with the case  $m_r = (1 - \beta) m_R$ . Let  $\omega \in \{Z_n^R \rightarrow \infty\}$ . Then, given  $\varepsilon > 0$  such that  $\eta_1 > 1$ , from Theorem 8.3, there exists  $k \geq 1$  such that  $\omega \in B_k^\varepsilon$ . Since, for each  $n \geq k$ ,

$$Z_n^r = \bar{Z}_n^{r,k-1} + \sum_{l=k}^n Z_n^{r,l},$$

then, for  $n \geq k$  and  $\omega$  (which we do not write further for simplicity)

$$\begin{aligned} \frac{Z_n^r}{n(\alpha m_r)^n} &= \frac{1}{n} \bar{Z}_0^{r,k-1} \prod_{j=0}^{n-1} \frac{\bar{Z}_{j+1}^{r,k-1}}{\alpha m_r \bar{Z}_j^{r,k-1}} \\ &+ \frac{1}{n} \frac{\beta}{(1-\beta)} \sum_{l=k}^n Z_0^R \prod_{j=0}^{l-2} \frac{Z_{j+1}^R}{\alpha(1-\beta) m_R Z_j^R} \frac{(1-\beta) Z_l^r}{\beta \alpha m_r Z_{l-1}^R} \prod_{j=l}^{n-1} \frac{Z_{j+1}^r}{\alpha m_r Z_j^r} \end{aligned}$$

Therefore, from Lema A.3 in [5] and Lemma A2, we deduce, a.s. on  $B_k^\varepsilon$ , that

$$\frac{Z_n^r}{n(\alpha m_r)^n} = O(n^{-1}) + \frac{n-k+1}{n} \frac{\beta}{(1-\beta)} \frac{Z_{k-1}^R}{(\alpha(1-\beta) m_R)^{k-1}} \prod_{j=k}^n (1 + O((\eta_1^{n-j})^{-\rho})),$$

for some  $0 < \rho < 1/2$ , and we obtain the result from Theorem 7.28 in [41], since  $\eta_1 > 1$ .

- (iii) Finally, the case  $m_r < (1 - \beta) m_R$  is an immediate consequence of Theorems 1 and A2-(ii).

□

### References

- Kido, T.; Lau, Y.F. Roles of the Y chromosome genes in human cancers. *Asian J. Androl.* **2015**, *17*, 373–380.
- González, M.; Martínez, R.; Mota, M. Bisexual branching processes to model extinction conditions for Y-linked genes. *J. Theor. Biol.* **2009**, *258*, 478–488. [CrossRef]
- Alsmeyer, G.; Gutiérrez, C.; Martínez, R. Limiting genotype frequencies of Y-linked genes through bisexual branching processes with blind choice. *J. Theor. Biol.* **2011**, *275*, 42–51. [CrossRef]
- González, M.; Gutiérrez, C.; Martínez, R. Non-parametric Bayesian inference through MCMC method for Y-linked two-sex branching processes with blind choice. *J. Stat. Comput. Simul.* **2018**, *88*, 3565–3587. [CrossRef]
- González, M.; Gutiérrez, C.; Martínez, R. Extinction conditions for Y-linked mutant-alleles through two-sex branching processes with blind-mating structure. *J. Theor. Biol.* **2012**, *307*, 104–116. [CrossRef] [PubMed]
- González, M.; Gutiérrez, C.; Martínez, R. Bayesian inference in Y-linked two-sex branching processes with mutations: ABC approach. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, in press.
- Ispány, M.; Körmendi, K.; Pap, G. Asymptotic behavior of CLS estimators for 2-type doubly symmetric critical Galton–Watson processes with immigration. *Bernoulli* **2014**, *20*, 2247–2277. [CrossRef]
- Mode, C. *Multitype Branching Processes. Theory and Applications*; American Elsevier Publishing Co. Inc.: New York, NY, USA, 1971.
- Bagley, J. On the asymptotic properties of a supercritical bisexual branching process. *J. Appl. Probab.* **1986**, *23*, 820–826. [CrossRef]
- Fu, S.; Yan, J.; Wang, X.; Dong, J.; Chen, P.; Wang, C.; Chen, G. The audiological characteristics of a hereditary Y-linked hearing loss in a Chinese ethnic Tujia pedigree. *Int. J. Pediatr. Otorhinolaryngol.* **2011**, *75*, 202–206. [CrossRef]
- Wang, Q.; Lu, C.; Li, N.; Rao, S.; Shi, Y.; Han, D.; Li, X.; Cao, J.; Yu, L.; Li, Q.; Guan, M.; Yang, W.; Shen, Y. Y-linked inheritance of non-syndromic hearing impairment in a large Chinese family. *J. Med. Genet.* **2011**, *41*, e80. [CrossRef]
- Bonal, R.; Hernández, M.; Espelta, J.; Muñoz, A.; Aparicio, J. Unexpected consequences of a drier world: evidence that delay in late summer rains biases the population sex ratio of an insect. *R. Soc. Open Sci.* **2015**, *2*, 150–198. [CrossRef] [PubMed]
- Ellis, S.; Franks, D.; Natrass, S.; Cant, M.; Weiss, M.; Giles, D.; Balcomb, K.; Croft, D. Mortality risk and social network position in resident killer whales: sex differences and the importance of resource abundance. *Proc. R. Soc. B* **2017**, *284*, e1313. [CrossRef] [PubMed]
- Hutter, S.; Li, H.; Beisswanger, S.; De Lorenzo, D.; Stephan, W. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. *Genetics* **2007**, *177*, 469–480. [CrossRef] [PubMed]
- Helgason, A.; Einarsson, A.W.; Guðmundsdóttir, V.B.; Sigurðsson, A.; Gunnarsdóttir, E.D.; Jagadeesan, A.; Ebenesersdóttir, S.S.; Kong, A.; Stefánsson, K. The Y-chromosome point mutation rate in humans. *Nat. Genet.* **2015**, *47*, 453–457. [CrossRef]
- Cruciani, F.; Trombetta, B.; Massaia, A.; Destro-Bisol, G.; Sellitto, D.; Scozzari, R. A Revised Root for the Human Y Chromosomal Phylogenetic Tree: The Origin of Patrilineal Diversity in Africa. *Am. J. Hum. Genet.* **2011**, *88*, 814–818. [CrossRef]



17. Hammer, M. A recent common ancestry for human Y chromosomes. *Nature* **1995**, *378*, 376–378. [[CrossRef](#)]
18. Thomson, R.; Pritchard, J.; Shen, P.; Oefner, P.; Feldman, M. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7360–7365. [[CrossRef](#)]
19. Bertoin, J.; Fontbona, J.; Martínez, S. On prolific individuals in a supercritical continuous-state branching process. *J. Appl. Probab.* **2008**, *3*, 714–726. [[CrossRef](#)]
20. Blumberg, S.; Lloyd-Smith, J. Inference of  $r_0$  and transmission heterogeneity from the size distribution of stuttering chains. *PLoS Comput. Biol.* **2013**, *9*, 1–17. [[CrossRef](#)]
21. Farrington, C.; Grant, A. The distribution of time to extinction in subcritical branching processes: applications to outbreaks of infectious disease. *J. Appl. Probab.* **1999**, *36*, 771–779. [[CrossRef](#)]
22. Farrington, C.; Kanaan, M.; Gay, N. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics* **2003**, *4*, 279–295. [[CrossRef](#)] [[PubMed](#)]
23. Mode, C.; Sleemam, C. *Stochastic Processes in Epidemiology*; World Scientific: Singapore, 2000.
24. Pakes, A. Biological applications of branching processes. In *Handbook of Statistic. Stochastic Processes: Modelling and Simulation*; Shanbhag, D.N., Rao, C.R., Eds.; Elsevier: Amsterdam, The Netherlands, 2003; Volume 21, pp. 693–773.
25. Sun, S.; Heitman, J. Should Y stay or should Y go: The evolution of non-recombining sex chromosomes. *BioEssays* **2012**, *342*, 938–942. [[CrossRef](#)] [[PubMed](#)]
26. Carlin, J. Mutations are the raw materials of evolution. *Nat. Educ. Knowl.* **2011**, *3*, 10.
27. Loewe, L.; Hill, W.G. *The Population Genetics of Mutations: Good, Bad and Indifferent*; The Royal Society: London, UK, 2010.
28. Charlesworth, B.; Charlesworth, D. The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **2000**, *355*, 1563–1572. [[CrossRef](#)] [[PubMed](#)]
29. Charlesworth, B.; Charlesworth, D. Evolution: A New Idea about the Degeneration of Y and W Chromosomes. *Curr. Biol.* **2020**, *30*, R871–R873. [[CrossRef](#)]
30. Lenormand, T.; Fyon, F.; Sun, E.; Roze, D. Sex chromosome degeneration by regulatory evolution. *Curr. Biol.* **2020**, *30*, 3001–3006. [[CrossRef](#)]
31. Steinemann, S.; Steinemann, M. Y chromosomes: born to be destroyed. *BioEssays* **2005**, *27*, 1076–1083. [[CrossRef](#)]
32. Gershoni, M.; Pietrokovski, S. Reduced selection and accumulation of deleterious mutations in genes exclusively expressed in men. *Nat. Commun.* **2014**, *5*, 1–10. [[CrossRef](#)]
33. Charlesworth, B.; Charlesworth, D. Some evolutionary consequences of deleterious mutations. *Genetica* **1998**, *102*, 3. [[CrossRef](#)]
34. Hwang, K.; Yatsenko, A.; Jorgez, C.; Mukherjee, S.; Nalam, R.; Matzuk, M.; Lamb, D. Mendelian genetics of male infertility. *Ann. N. Y. Acad. Sci.* **2010**, *1214*, E1. [[CrossRef](#)]
35. Irvine, D. Epidemiology and aetiology of male infertility. *Hum. Reprod.* **1998**, *13*, 33–44. [[CrossRef](#)] [[PubMed](#)]
36. Winters, B.; Walsh, T. The epidemiology of male infertility. *Urol. Clin.* **2014**, *41*, 195–204. [[CrossRef](#)] [[PubMed](#)]
37. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **2013**, *14*, 113–124. [[CrossRef](#)]
38. González, M.; Hull, D.M.; Martínez, R.; Mota, M. Bisexual branching processes in a genetic context: The extinction problem for Y-linked genes. *Math. Biosci.* **2006**, *202*, 227–247. [[CrossRef](#)] [[PubMed](#)]
39. González, M.; Gutiérrez, C.; Martínez, R.; Mota, M. Extinction probability of some recessive alleles of X-linked genes in the context of two-sex branching processes. In *Branching Processes and Their Applications*; del Puerto, I.M., González, M., Gutiérrez, C., Martínez, R., Minuesa, C., Molina, M., Mota, M., Ramos, A., Eds.; Springer: Cham, Switzerland, 2016; Chapter 17, pp. 287–305.
40. Germain, D. General aspects of X-linked diseases. In *Fabry Disease: Perspectives from 5 Years of FOS*; Mehta, A., Beck, M., Sunder-Plassmann, G., Eds.; Oxford PharmaGenesis: Oxford, UK, 2006; Chapter 7.
41. Stromberg, K.R. *An Introduction to Classical Real Analysis*; Wadsworth, Inc.: Belmont, CA, USA, 1981.
42. Kesten, H.; Stigum, B. Limit theorem for decomposable multi-dimensional Galton-Watson processes. *J. Math. Anal. Appl.* **1967**, *17*, 309–338. [[CrossRef](#)]
43. Hush, D.; Scovel, C. Concentration of the hypergeometric distribution. *Stat. Probab. Lett.* **2005**, *75*, 127–132. [[CrossRef](#)]