



HAL
open science

Bisexual branching processes to model extinction conditions for Y-linked genes

Miguel González, Rodrigo Martínez, Manuel Mota

► **To cite this version:**

Miguel González, Rodrigo Martínez, Manuel Mota. Bisexual branching processes to model extinction conditions for Y-linked genes. *Journal of Theoretical Biology*, 2009, 258 (3), pp.478. 10.1016/j.jtbi.2008.10.034 . hal-00554526

HAL Id: hal-00554526

<https://hal.science/hal-00554526>

Submitted on 11 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Bisexual branching processes to model extinction conditions for Y-linked genes

Miguel González, Rodrigo Martínez, Manuel Mota

PII: S0022-5193(08)00573-0
DOI: doi:10.1016/j.jtbi.2008.10.034
Reference: YJTBI5349

To appear in: *Journal of Theoretical Biology*

Received date: 30 January 2008
Revised date: 27 October 2008
Accepted date: 27 October 2008

Cite this article as: Miguel González, Rodrigo Martínez and Manuel Mota, Bisexual branching processes to model extinction conditions for Y-linked genes, *Journal of Theoretical Biology* (2008), doi:[10.1016/j.jtbi.2008.10.034](https://doi.org/10.1016/j.jtbi.2008.10.034)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/jtbi

Bisexual branching processes to model extinction conditions for Y-linked genes^{*}

Miguel González Rodrigo Martínez^{**} Manuel Mota

Department of Mathematics, University of Extremadura, Badajoz 06071, Spain.

Abstract

In a two-sex monogamic population, the evolution of the number of carriers of the two alleles of a Y-linked gene is considered. To this end, a multitype bisexual branching model is presented in which it is assumed that the gene has no influence on the mating process. It is deduced from this model that the average numbers of female and male descendants per mating unit constitute the key to determining the extinction or survival of each allele. Moreover, the destiny of each allele in the population is found not to depend on the behaviour of the other.

Key words: Sex-linked inheritance, Y chromosome, two-dimensional bisexual stochastic model, perfect fidelity mating, survival.

1 Introduction

Sex determination in most mammal populations, including humans, is due to a pair of chromosomes denominated X and Y. Females have XX chromosomes, while males have two distinct chromosomes, XY. The SRY gene on the Y chromosome determines the development of the testes, and hence maleness (see Wallis et al. (2008)). The Y chromosome has very few genes in comparison with the X chromosome, but recent investigations have shown the importance of some Y-linked genes in populations of humans (see, for example, the web page www.nature.com/nature/focus/ychromosome/, Quintana-Murci and Fellous (2001), or Hughes et al. (2005)) and other animals which have the XX/XY

^{*} The research was supported by the Ministerio de Educación y Ciencia and the FEDER through the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, grant MTM2006-08891.

^{**}Corresponding author. e-mail: rmartinez@unex.es

type of sex determination system (see, for example, Bernardo et al. (2001), Gutiérrez and Teem (2006), or the review by Charlesworth et al. (2005)). Moreover this chromosome has the particularity of being male-specific and haploid, and of having a region (the non-recombining region, NRY, being 95% of the chromosome in humans – see, for example, Krausz et al. (2004) or Graves (2006)) which escapes recombination. These unique properties of the Y chromosome have important consequences for its population genetics: the NRY region passes down from father to son largely unchanged, and is therefore very useful for studying how populations have evolved. By examining the differences between modern Y chromosomes (such as DNA polymorphisms), one can attempt to reconstruct a history of paternal lineages. There have been many studies in this sense in the context of populations of humans (e.g., Hurles et al. (1998), Quintana-Murci et al. (2001), Hurles et al. (2002), or Rosa et al. (2007)) and other species (e.g., Tosi et al. (2002), Hellborg et al. (2005), or Geraldès et al. (2005)).

Another singular question associated with the Y chromosome is that of the microdeletions of this chromosome's long arm (Yq). These Yq deletions define three regions collectively known as AZF (azoospermia factors), with deletions of the AZFc region being the commonest. The Yq deletion is associated with males with fertility problems (for a review, see Krausz et al. (2003)), but many cases have been reported in which the natural transmission of this genetic defect from fathers to sons has occurred (see e.g., Saut et al. (2002), Calogero et al. (2002) or Kuhnert et al. (2004)). Related to this question, Krausz and McElreavey (2001) state: "The pathogenetic significance of Y chromosome microdeletions is spermatogenic failure and not infertility. AZF deletions can be associated with oligozoospermia which usually leads to infertility but it can also be associated with normal couple fertility. For the interpretation of any new observations about the transmission of Yq deletions we should keep in mind that fathering one or more children is the expression of the fertility status of both members of the couple". Obviously, determining the evolution of the number of males with this genetic defect in a human population is an important medical problem (see, for example, Patsalis et al. (2002) or Fitch et al. (2005)), but it has also been investigated in other species (e.g., Toure et al. (2004)).

The surname is another characteristic which can be seen as Y-linked in humans. There have been some recent studies aimed at determining the relationship between surnames and Y-chromosome lineages (e.g., King et al. (2006) or Bowden et al. (2008)).

Other organisms have a mirror image of the XX/XY sex determination system, with males being homogametic and females heterogametic. To avoid confusion, these sex chromosomes are denominated Z and W. Thus in birds, snakes, and butterflies, for example, females have ZW sex chromosomes, and males have ZZ

sex chromosomes. Now, Y-linked problems become W-linked (see, for example, Ogawa et al. (1998), Yamada et al. (2004), or Abe et al. (2008)). For simplicity in notation, throughout this paper only the XX/XY sex determination system will be referred to, although the development and the results are equally valid for the ZZ/ZW system.

Appropriate mathematical models are needed to understand the evolution of Y chromosome lineages (for example, to help solve the problem of Y-chromosomal Adam), Yq deletions, or other Y-linked genes.

Branching processes naturally come to mind in this context. These are stochastic models which arise in the description of population dynamics, being of particular use in describing the extinction/growth of populations (see Haccou et al. (2005)). Branching models have been applied to many biological problems in such fields as epidemiology, genetics, and cell dynamics. Examples include the evolution of infectious diseases (e.g., Mode and Sleemam (2000), Ball et al. (2004), or Garske and Rhodes (2008)), population genetics (e.g., Campbell (2003) or Iwasa et al. (2005)), and stem cells (e.g., Yakovlev and Yanev (2006)). Further examples are reviewed in the recent monographs of Kimmel and Axelrod (2002) and Pakes (2003), and in the communication of Caron-Lormier et al. (2006).

The simplest branching models, the Galton-Watson process and the Markov branching process, have been used to model Y-chromosome lineages and their female analogues – mitochondrial DNA lineages (see O’Connell (1995) and Neves and Moreira (2006)). But more accurate models are needed in which all the phases of sexual reproduction can be considered, including the interaction between females and males in producing offspring. In González et al. (2006), a model was presented that describes the evolution of the number of carriers of the two alleles of a Y-linked gene in a two-sex monogamic population. In that paper, it was considered that the characters controlled by such a gene can have some influence on the mating process of the species, with females having a preference for males carrying one of the alleles of the gene. It was shown that this preference can sometimes be definitive in determining the survival of the different genotypes in the population.

However, females’ choice of mate is not always conditioned by the males’ genotype. Most Y-linked characters do not appear in the phenotype of the males, or, even if they do, are not decisive at the time of mating (for example, in the case of Yq deletions or many of the NRY region genes). In these situations it seems more realistic to consider a model where females choose their mates without caring about what their genotype is, i.e., each female makes a *blind choice* of the genotype of her mate.

In this paper, we present a multitype bisexual branching process to model the

evolution of the number of carriers of each allele of a Y-linked gene or of Y chromosome lineages in a two-sex monogamic population, assuming that this gene or characteristic does not influence the mating process. We focus our attention on studying the possibility of fixation of one of the alleles of the gene, and on determining the destiny of the gene in the population.

The rest of the paper consists of four sections. In Section 2 we provide the definition of the Y-linked Bisexual Branching Process assuming that females choose the genotypes of their mates blindly. We consider Y-linked genes with two alleles. We also present some basic properties of the model. In Section 3 we study the extinction problem for both the whole population and each genotype separately, in the latter case taking into account the behaviour of the other genotype. For some critical situations, we simulate the evolution of the population and conjecture its long term behaviour. In Section 4 we provide some concluding remarks. Finally, Section 5 presents the proofs of the results.

2 The model

Consider a human or other animal diploid population with non-overlapping generations, where the sex of an individual is determined by a pair of chromosomes X and Y. A female (F) will have XX chromosomes, while a male (M) will have XY chromosomes.

Consider also a gene with a pair of alleles, R and r, linked to the Y chromosome. The character defined by this gene is exclusive to males, which will be designated by MR or Mr according to which allele they carry. We will suppose however that they both have the same phenotype, so that mating is not influenced by the gene and females choose the genotype of their mates blindly. Moreover, perfect fidelity (monogamy) is assumed at the time of mating, i.e., each individual mates with only one individual of the opposite sex, provided that some of them are still available. Monogamy is the mating option in some birds and mammals, and in particular in most human populations, and is therefore well suited to the specific problems which motivated the present work (e.g., Yq deletions, or the history of paternal lineages).

A couple will be classified as FMR or FMr according to the genotype of the male. Following the inheritance rules, FMR mating units can generate females and MR males, while FMr mating units can produce females and Mr males.

We shall model the evolution of such a population by means of a two-type bisexual branching process in discrete time. To define this model, assume that it is known how many couples there are in the present generation, labeled in the definition as generation n . The objective is to determine the number

of females, males, and mating units of each genotype in the next generation, labeled $n + 1$. To this end, two phases will be considered: reproduction and mating.

In the reproduction phase, each couple of generation n is assumed to produce offspring independently of the other couples according to a random variable. The probability distribution of these variables will be the same for all the couples with a given genotype, irrespective of the generation they belong to, and will be termed the reproduction law of that genotype. In general, FMR and FMr couples are assumed to have differences in their reproductive capacities, and consequently their respective reproduction laws are assumed to be different. This is an appropriate assumption in such problems as Yq deletions associated with male fertility problems. The important parameters will be the average numbers of offspring produced by FMR and FMr couples, denoted here by m_R and m_r , respectively.

With probability α an offspring will be a female, and consequently with probability $1 - \alpha$ it will be a male. It is assumed that the genotype has no influence on the sex determination, so that α is the same for both genotypes. Therefore, the average number of females and males per FMR couple will be αm_R and $(1 - \alpha)m_R$, respectively, and these values will be αm_r and $(1 - \alpha)m_r$, respectively, for an FMr couple.

The female offspring of all the couples in generation n gives the total number of females in generation $n + 1$, which will be denoted by F_{n+1} . On the basis of the genetic rules described above, the male offspring of all the FMR couples in generation n gives the total number of MR males in generation $n + 1$, denoted by MR_{n+1} . Analogously, the male offspring of all the FMr couples in generation n gives the total number of Mr males in generation $n + 1$, denoted by Mr_{n+1} .

In the mating phase, the number of couples of each genotype formed in generation $n + 1$ is calculated, taking into account that generations do not overlap and provided that in the population there are F_{n+1} females, MR_{n+1} males with R genotype, and Mr_{n+1} males with r genotype. We will denote by ZR_{n+1} and Zr_{n+1} the total number of FMR and FMr couples, respectively, in generation $n + 1$.

The point here is the perfect fidelity mating considered in the premises of the model. According to this assumption, the total number of couples in generation $n + 1$ is the minimum of the number of females, F_{n+1} , and the total number of males, $MR_{n+1} + Mr_{n+1}$.

If the number of females exceeds that of males, then all the males mate, and consequently the number of couples of a genotype will be equal to the number of males of that genotype, i.e., $ZR_{n+1} = MR_{n+1}$ and $Zr_{n+1} = Mr_{n+1}$. On the

contrary, if the number of females is less than or equal to the number of males, all the females mate and must *choose* the genotype of their mates. However, since all the males have the same phenotype, each of the F_{n+1} females makes a blind choice between the MR_{n+1} males with R genotype and the Mr_{n+1} males with r genotype. Thus, the number of FMR couples, ZR_{n+1} , will be given by a hypergeometric distribution of parameters F_{n+1} , $MR_{n+1} + Mr_{n+1}$, and MR_{n+1} , i.e. F_{n+1} males are selected from all males of generation $n + 1$ where MR_{n+1} males have the R genotype. The rest of the couples will have r genotype, i.e., $Zr_{n+1} = F_{n+1} - ZR_{n+1}$.

Obviously, the symmetry in the mating process means that it would be equivalent to consider Zr_{n+1} given by a hypergeometric distribution of parameters F_{n+1} , $MR_{n+1} + Mr_{n+1}$, and Mr_{n+1} , and the number of remaining couples up to F_{n+1} would be ZR_{n+1} . This symmetry is the main difference between the present model and that with preference in the mating introduced in González et al. (2006). In that model, the genetic character controlled by the gene does appear in the phenotype, and females prefer males carrying one of the alleles of the gene. The consequent lack of symmetry in the mating had a great influence on the results.

The bivariate sequence $\{(ZR_n, Zr_n)\}_{n \geq 0}$ indicating the generation-by-generation evolution of the number of mating units will be termed a Y-linked bisexual branching process.

From the definition of the model, the number of couples of each genotype in the next generation depends only on the present number of mating units, and not on the number of ancestors that belonged to past generations. Moreover, since each reproduction law remains the same over the generations, the transitions from one generation to another are homogeneous, i.e., they do not depend on the generation. The process $\{(ZR_n, Zr_n)\}_{n \geq 0}$ is therefore a homogeneous two-type Markov chain. However, the general theory of Markov chains, based on transition probabilities, does not lead to any significant results for this model. It is therefore necessary to attempt a specific methodological approach that is typical of branching processes.

The model introduced here shares some basic properties with the model with mating preference presented in González et al. (2006). These are properties which do not depend on the way the females choose their mates, and were proved in that paper. The following are four of these properties which are germane to the present work:

P 1 If in some generation there are no mating units of a particular genotype then, from this generation on, the mating units and the males of that genotype no longer exist. This is known as *extinction of this particular genotype* which implies the fixation of the allele determining the other genotype. The

extinction of the whole population occurs when there are no mating units of any genotypes.

P 2 Provided that there exist couples of both genotypes in the present, there is a positive probability of reaching any number of mating units of each genotype in future generations. In particular, the fixation of one of the genotypes or even the extinction of the whole population are possible.

P 3 Each genotype presents the dual behaviour typical of branching processes: either it becomes extinct or the number of couples of this genotype eventually reaches arbitrarily large values. The latter event is known as *explosion or indefinite growth of this particular genotype*. Consequently the whole population also presents this duality. Thus, the survival of each genotype or of the whole population is equivalent to their indefinite growth as generations go by, with the possibility that, in the long term, their size tends to be in the neighbourhood of one or more positive values having to be discarded. Although this property might seem unrealistic, it merely expresses what would be the ideal long term evolution of a population when its development is not constrained by any external bound.

P 4 If the fixation of a particular allele has occurred, the corresponding genotype evolves essentially as a Bisexual Branching Process (BBP) with perfect fidelity mating and the reproduction law of the surviving genotype. Analogously, if both genotypes have the same reproduction law, the evolution of the total number of couples can also be modeled by a BBP with perfect fidelity mating.

With respect to this last property, it has to be remarked that the present model in some way inherits some of the properties of the BBP, a model that has been extensively investigated in the literature on Branching Processes. One may cite as examples work on the extinction probability (e.g., Daley (1968), Hull (1982, 1984), Bruss (1984), Daley et al. (1986), or Alsmeyer and Rösler (2002)), the long term behaviour (e.g., Bagley (1986) or González and Molina (1996, 1997)), and inference problems (e.g., González-Fragoso (1995), Molina et al. (1998), Alsmeyer and Rösler (1996), or González et al. (2001)). Two interesting recent reviews are Hull (2003) and Haccou et al. (2005).

Before concluding this section, it is worth comparing the Y-linked bisexual branching process with the classical genetic model of Wright-Fisher with which it shares some common features. In particular, both are Markov chains and consider diploid populations with non-overlapping generations. However there are also many differences. The most important is that, in the Wright-Fisher setting, the population size remains constant over the generations, reflecting a regulation of population size by external resources. The Wright-Fisher model may therefore not be the best model for a natural population. The Y-linked

bisexual model, however, does allow indefinite growth. With respect to the mating, the Wright-Fisher model considers this to be random in the sense that every allele in a generation could come from every possible couple in the previous one owning that allele. In the Y-linked bisexual process, the number of couples owning each of the alleles, established in the mating phase, is of prime importance in determining the number of carriers of that allele in the subsequent generation.

3 The extinction problem

In this section we deal with two problems of interest in both the genetics and the population dynamics frameworks (e.g., Yq deletions, or the history of paternal lineages). The first is to obtain a necessary and sufficient condition for the whole population to become extinct almost surely. The other is to analyze the destiny of the gene in the population. To this end, we study some relevant events, such as the extinction of each genotype (mathematically, $\{ZR_n \rightarrow 0\}$ and $\{Zr_n \rightarrow 0\}$) and the survival or indefinite growth of each genotype (mathematically, $\{ZR_n \rightarrow \infty\}$ and $\{Zr_n \rightarrow \infty\}$). Survival is possible in the presence or absence of the other genotype, so we also consider the events $\{ZR_n \rightarrow \infty, Zr_n \rightarrow 0\}$ (fixation of R genotype), $\{ZR_n \rightarrow 0, Zr_n \rightarrow \infty\}$ (fixation of r genotype), and $\{ZR_n \rightarrow \infty, Zr_n \rightarrow \infty\}$ (survival of both genotypes), which are also of interest in themselves. We provide conditions depending on the parameters m_R , m_r , and α .

We will assume that at the time the population was first observed, i.e., in generation 0, there were i couples with R genotype and j couples with r genotype. In order to simplify the notation, we denote $P(\cdot | (ZR_0, Zr_0) = (i, j))$ by $P_{(i,j)}(\cdot)$. Even (i, j) will be dropped in this notation if there is no ambiguity.

The extinction of the population

It is understood that the population becomes extinct if from some generation on there are no mating units. For every (i, j) we denote *the probability of extinction* of a population starting with i FMR and j FMr mating units by $Q_{(i,j)} = P_{(i,j)}(ZR_n \rightarrow 0, Zr_n \rightarrow 0)$.

Let Q_i^R (resp. Q_j^r) be the probability of extinction of a BBP with perfect fidelity mating, reproduction law of R (resp. r) genotype, and i (resp. j) starting mating units. From property P4, it follows that, for all $i, j \geq 0$,

$$Q_{(i,0)} = Q_i^R \quad \text{and} \quad Q_{(0,j)} = Q_j^r ,$$

and therefore the results on the extinction of a BBP with perfect fidelity mating (see Daley (1968)) apply here.

When $i, j > 0$, the following result provides a satisfactory answer to the problem of extinction.

Result 1 *Let $i, j > 0$. Then $Q_{(i,j)} = 1$ if and only if $\min\{\alpha m_r, (1-\alpha)m_r\} \leq 1$ and $\min\{\alpha m_R, (1-\alpha)m_R\} \leq 1$.*

We conclude that if the average number of females or males produced by a mating unit of each type is less than or equal to one then the process becomes extinct almost surely. On the other hand, if at least one of the values $\min\{\alpha m_R, (1-\alpha)m_R\}$ or $\min\{\alpha m_r, (1-\alpha)m_r\}$ is greater than one then, applying the results of a BBP, the corresponding genotype has a positive probability of survival even without any contribution of the other genotype, and therefore the population may survive too.

The destiny of the gene in the population

After determining a necessary and sufficient condition for the process to become extinct with probability one, we focus our interest on the long term growth of the number of carriers of each allele.

However, it would seem logical for there to exist some interdependence between the behaviour of the two genotypes, so that, in the analysis of each genotype, we must distinguish between whether the other has also an indefinite growth or has become extinct. In particular, we have to deal with the events

$$\{ZR_n \rightarrow \infty, Zr_n \rightarrow 0\}, \{ZR_n \rightarrow 0, Zr_n \rightarrow \infty\}, \text{ and } \{ZR_n \rightarrow \infty, Zr_n \rightarrow \infty\}$$

which represent the possible destinies of the two alleles in the population.

First we investigate the extinction of one genotype and the survival of the other. Because of the symmetry in the role of the two genotypes, the study of the fixation of each of them, $\{ZR_n \rightarrow \infty, Zr_n \rightarrow 0\}$ and $\{ZR_n \rightarrow 0, Zr_n \rightarrow \infty\}$, leads to analogous conclusions. In both events, the surviving genotype behaves from some generation on as a BBP with perfect fidelity mating (see property P4). Hence, the theory of Daley (1968) about a BBP applies, and it is immediate to deduce the following result.

Result 2 *Let $i, j > 0$, then*

- (i) $P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow 0) > 0$ if and only if $\min\{\alpha m_R, (1-\alpha)m_R\} > 1$.
- (ii) $P_{(i,j)}(ZR_n \rightarrow 0, Zr_n \rightarrow \infty) > 0$ if and only if $\min\{\alpha m_r, (1-\alpha)m_r\} > 1$.

Intuitively, this result states that a necessary and sufficient condition for a genotype to have a positive probability of fixation is that both the female and the male descendants per couple of such a genotype are on average greater than one.

From the “if” parts of the above result, it is immediate to deduce the following corollary relative to the growth of the number of carriers of each allele.

Result 3 *Let $i, j > 0$, then*

- (i) *If $\min\{\alpha m_R, (1 - \alpha)m_R\} > 1$, then $P_{(i,j)}(ZR_n \rightarrow \infty) > 0$.*
- (ii) *If $\min\{\alpha m_r, (1 - \alpha)m_r\} > 1$, then $P_{(i,j)}(Zr_n \rightarrow \infty) > 0$.*

With respect to the simultaneous survival of both genotypes, i.e., the event $\{ZR_n \rightarrow \infty, Zr_n \rightarrow \infty\}$, one has the following result.

Result 4 *Let $i, j > 0$.*

- (i) *If $(1 - \alpha)m_R \leq 1$ or $(1 - \alpha)m_r \leq 1$, then $P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) = 0$.*
- (ii) *If $\alpha m_R < 1$ or $\alpha m_r < 1$, then $P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) = 0$.*

In Result 4 (i) we state that if the average number of male descendants per couple of either genotype is less than or equal to one, then there is null probability of the simultaneous survival of both genotypes. The key to this behaviour lies in the fact that the number of mating units of a genotype is bounded above by the number of males of that genotype.

Also, the probability of simultaneous survival of both genotypes is zero when the average number of female descendants per couple of either genotype is less than one. This is true, as indicated in Result 4 (ii), even though it could be expected that the males of this genotype would mate with females generated by mating units of the other genotype. For example, if one takes $\alpha < 0.5$, $\alpha m_R < 1$, $(1 - \alpha)m_R > 1$, and $\alpha m_r > 1$, even though the number of FMr mating units can grow indefinitely and the average number of MR male descendants per FMR mating unit is greater than one, the FMR mating units become extinct almost surely because not enough females are produced for all the population. This is an amazing fact, which in no way can be considered intuitive. To illustrate this scenario, we simulated 12 generations with $(ZR_0, Zr_0) = (50, 1)$, $\alpha = 0.3$, and the reproduction laws following Poisson distributions with $m_R = 2$ and $m_r = 4$. In Figure 1, we show a path of this process in which one observes the pattern described above.

The cases $\alpha m_R = 1$ and $\alpha m_r = 1$ are not considered in Result 4 (ii). If $\alpha \geq 0.5$, and $\alpha m_R = 1$ or $\alpha m_r = 1$, then $(1 - \alpha)m_R \leq 1$ or $(1 - \alpha)m_r \leq 1$, respectively, and Result 4 (i) guarantees that $P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) = 0$. However, if $\alpha < 0.5$, we have not proved whether there is positive or null probability

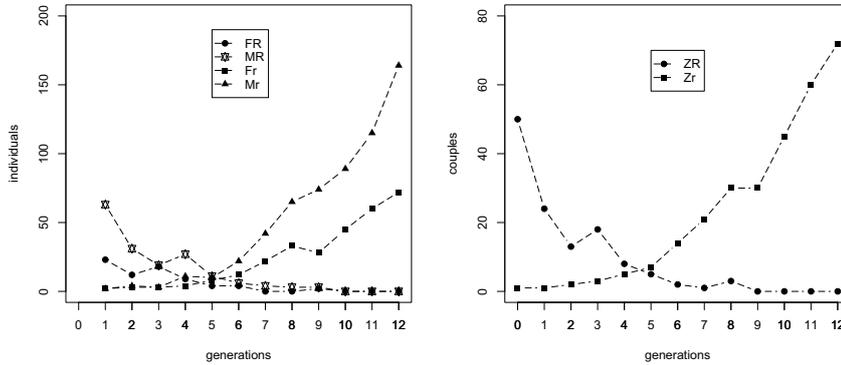


Fig. 1. Path of a process in which $\alpha m_R < 1$, $(1 - \alpha)m_R > 1$, and $\alpha m_r > 1$.

of the simultaneous survival of both genotypes. In order to conjecture the possible behaviour of the process in such situations, we performed a Monte Carlo simulation of ten batches of 10 000 processes until generation 1000, with reproduction laws following Poisson distributions with means $m_R = 2.5$ and $m_r = 2.52$, and probability for an offspring to be female $\alpha = 0.4$. Under these conditions $\alpha m_R = 1$, $(1 - \alpha)m_R = 1.5 > 1$, and $\alpha m_r = 1.008 > 1$. In all the simulated processes we took $(ZR_0, Zr_0) = (3, 3)$. Table 1 lists the results for the number of processes in each batch where both genotypes have survived by a given generation. One observes how the number of such processes decreases to zero over the course of succeeding generations, with similar records for the ten batches. Also, Figure 2 shows the proportion of processes among the 100000 simulated ones in which both genotypes have survived by a given generation. Actually, this plot provides an estimate of the probability that both genotypes survive until generation n , with $n = 1, 2, \dots$. Figure 3 shows the path of one process where both genotypes have survived until generation 1000. One observes geometric growth of the number of FMr mating units over the course of succeeding generations. The evolution of the number of FMR mating units, however, does not seem to fit any clear pattern, showing many fluctuations. However, according to property P3, it must converge either to 0 or ∞ . Since this path does not seem to have an indefinite growth, we guess that eventually this path will become extinct, although it will take a great number of generations as can be observed in the figure. We could therefore conjecture that, in the situations with $\alpha m_R = 1$ or $\alpha m_r = 1$ not covered by Result 4, $P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) = 0$.

Applying property P3, it is immediate to deduce the following corollary of Results 2 and 4, related to the extinction of each allele.

Result 5 *Let $i, j > 0$, then*

(i) *If $\min\{\alpha m_R, (1 - \alpha)m_R\} < 1$, then $P_{(i,j)}(ZR_n \rightarrow \infty) = 0$.*

batch	1	2	3	4	5	6	7	8	9	10
generation 20	233	242	221	223	219	245	199	223	229	239
generation 40	55	58	49	54	45	48	49	51	56	49
generation 60	27	29	24	25	18	22	29	30	27	27
generation 80	18	17	15	17	11	13	14	21	14	17
generation 100	13	12	11	15	7	12	11	15	12	13
generation 200	3	3	2	4	2	4	4	7	6	6
generation 300	3	2	1	1	2	4	4	3	4	5
generation 400	3	2	1	1	1	1	1	2	4	2
generation 500	3	2	0	0	1	1	1	2	4	2
generation 1000	3	1	0	0	0	0	1	1	3	1

Table 1

Records of the number of processes in each batch where both genotypes have survived over the course of succeeding generations.

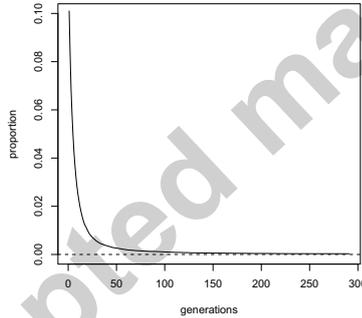


Fig. 2. Proportion over the course of succeeding generations of simulated processes in which both genotypes have survived.

(ii) If $\min\{\alpha m_r, (1-\alpha)m_r\} < 1$, then $P_{(i,j)}(Zr_n \rightarrow \infty) = 0$.

Remark 1 From Results 3 and 5, one can infer that, roughly speaking, the extinction or survival of each genotype depends only on its own reproductive abilities, without any influence of the other genotype.

Finally, our interest now is to find conditions guaranteeing a positive probability of the simultaneous survival of both alleles. Such conditions are established in the following result.

Result 6 Let $i, j > 0$. If $\alpha \neq 0.5$, $\min\{\alpha m_R, (1-\alpha)m_R\} > 1$ and $\min\{\alpha m_r, (1-\alpha)m_r\} > 1$, then $P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) > 0$.

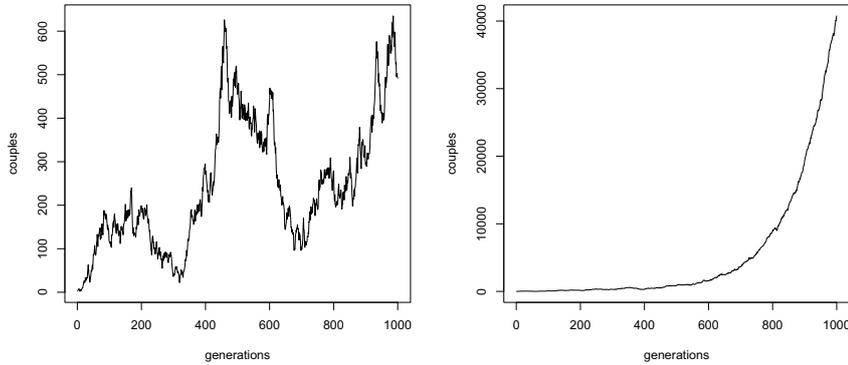


Fig. 3. Paths of the FMR (left) and FMr (right) genotypes in a process where both have survived until generation 1000.

Intuitively, this result states that, for the simultaneous survival of both genotypes, the average number of male and of female descendants of each of them must be greater than one. This happens even in the case of very unbalanced sizes of the two genotypes. Actually, when $\alpha > 0.5$ (resp. $\alpha < 0.5$) and the number of couples of each genotype grows, then the number of females (resp. males) produced by all the mating units of each genotype is almost surely greater than the number of males (resp. females) produced by these mating units. Hence, the number of mating units of each genotype is determined by the number of males (resp. females) given by all the mating units of their genotype. Therefore, as we noted in the previous remark, each genotype must develop by itself in order for there to be a positive probability of simultaneous survival. Notice that this probability has not to be equal to one, i.e., there could be some paths where both genotypes do not survive simultaneously due to random fluctuations, even being all the thresholds parameters greater than one.

When, however, $\alpha = 0.5$, the probability for the number of females to exceed that of males and the probability of there being more males than females are the same. In order to form a conjecture as to the possible behaviour of the process in this situation, where $\alpha m_R = (1 - \alpha)m_R$ and $\alpha m_r = (1 - \alpha)m_r$, we took both quantities to be greater than 1, performing a Monte Carlo simulation of ten batches of 10 000 processes until generation 200, with reproduction laws following Poisson distributions with means $m_R = 2.10$ and $m_r = 2.15$. In all the simulated processes, we took $(ZR_0, Zr_0) = (50, 50)$. Table 2 lists the results for the number of processes in each batch where both genotypes have survived by a given generation. One observes how the number of such processes decreases slowly to become stable over the course of succeeding generations, with similar records for the ten batches. Also, Figure 4 shows the proportion of processes among the 100 000 simulated ones in which both genotypes have survived by a given generation. Actually, this plot provides

batch	1	2	3	4	5	6	7	8	9	10
generation 20	9798	9809	9825	9819	9796	9803	9806	9796	9798	9809
generation 40	8112	8164	8071	8121	8084	8083	8128	8056	8063	8138
generation 60	6508	6452	6402	6516	6463	6450	6484	6408	6425	6482
generation 80	5566	5545	5553	5586	5526	5589	5634	5527	5547	5607
generation 100	5136	5150	5177	5151	5108	5183	5221	5084	5102	5203
generation 120	4953	4952	5009	4981	4937	5003	5036	4884	4931	5007
generation 140	4879	4866	4903	4895	4860	4919	4949	4808	4851	4932
generation 160	4841	4832	4873	4863	4832	4882	4907	4776	4816	4894
generation 180	4825	4818	4863	4846	4817	4869	4887	4763	4794	4888
generation 200	4820	4815	4856	4837	4813	4860	4881	4757	4790	4883

Table 2

Records of the number of processes in each batch where both genotypes have survived over the course of succeeding generations.

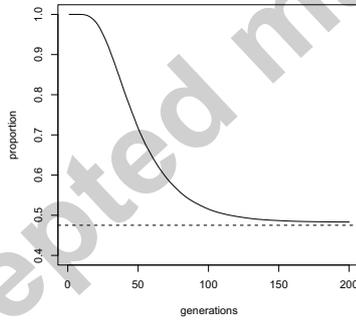


Fig. 4. Proportion over the course of succeeding generations of simulated processes in which both genotypes have survived.

an estimate of the probability that both genotypes survive until generation n , with $n = 1, 2, \dots$. This probability is close to 0.475 for n large enough. Figure 5 shows the logarithm of the number of couples of each genotype from the path of one process where both genotypes survive until generation 200. The observed linear behaviour leads one to deduce a geometric growth over the course of succeeding generations of the numbers of FMR and FMr mating units. We could therefore conjecture that, in the situations with $\alpha = 0.5$, $\alpha m_R > 1$, and $\alpha m_r > 1$, not covered by Result 6, $P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) > 0$.

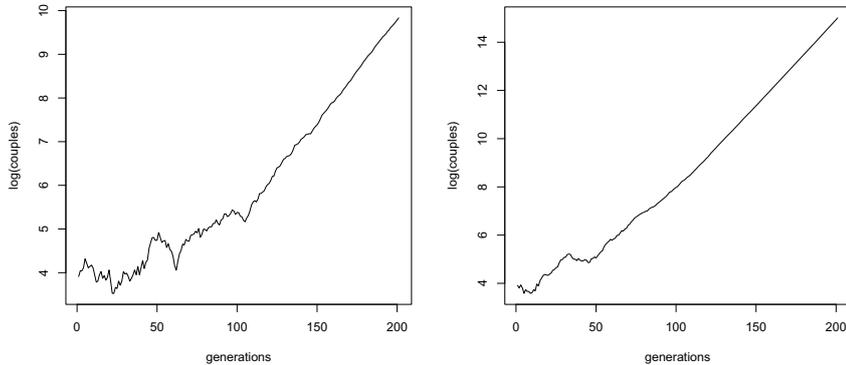


Fig. 5. Logarithm of the number of couples of the FMR (left) and FMr (right) genotypes from the path of one process where both genotypes have survived until generation 200, with $\alpha = 0.5$, $\alpha m_R > 1$, and $\alpha m_r > 1$.

4 Concluding remarks

Two biological questions have been the motivation of this communication: Yq microdeletions leading to relative male infertility, and reconstruction of paternal lineages. Both questions were addressed as the transmission of a certain Y-linked gene or mark, with two alleles or versions, controlling characters which are not expressed in the individual's phenotype. These alleles could represent the presence or the absence of the character in individuals, such as having or not having Yq deletions or certain DNA polymorphisms.

We approached both situations by constructing a two-type bisexual branching process which models the generation-by-generation evolution of the number of carriers of such a Y-linked gene or mark in a two-sex monogamic population. In this framework, only males carry the gene and individuals (females or males) can mate with no more than one individual of the opposite sex in order to produce new individuals which will be females with a given probability (sex-ratio), the same for all generations. Moreover, we considered that all males have the same phenotype, and therefore a female chooses her mate blindly without caring about which genotype he has. As we pointed out in Section 2, monogamy is the predominant mating option in some birds and mammals, and in particular in most human populations, and is therefore well suited to the specific problems which motivated the present work (i.e., Yq deletions and the history of paternal lineages). But true monogamy, as considered in this paper, is difficult to find in nature. To relax this assumption, other mating options could have been considered, such as polygamous mating or population-size-dependent number of couples. It seems likely that these other types of mating would lead to different results, as indeed is the case in the monogamous mating model with preference of females for males carrying one of the alleles of the

gene, studied in González et al. (2006), in relation to the simultaneous survival of the two genotypes.

Roughly speaking, we have shown that the extinction or survival of each genotype depends on the magnitudes of the average numbers of females and males produced by a mating unit of that genotype. Specifically, in the context of Yq deletions, this means that if the fertility rate of the couples formed by males with microdeletions is below a certain level, and there is no mutation, the males with this character will disappear from the population in the long term. On the contrary, if this fertility rate is above that level, there is a positive probability that the Yq deletions will remain in the population, even though their prevalence may be very low. This behaviour does not depend on the fertility rate of the couples formed by males without microdeletions. Therefore, although there is possible competition between the two types of males, this competition is not definitive for the presence of Yq deletions in the long term. Similar conclusions can be applied to the extinction or survival of paternal lineages. Indeed, this was the original problem which motivated work on branching processes (see Galton (1873)), although the first answers to this problem, based on the standard Galton-Watson branching model, did not take into account such biological parameters as mating option or sex ratio.

Moreover, our model has shown itself to have the advantage over the Wright-Fisher model in allowing fluctuations of the population size. Also, as pointed out above, it could be more useful than the Galton-Watson branching process for studies such as that of Neves and Moreira (2006) on Y-chromosomal Adam, because it considers a two-sex population and carefully models the mating process.

For a sex-ratio of 0.5, some conclusions were established on the basis of simulations, but not analytically. Most populations in fact do not have a sex-ratio exactly equal to 0.5, but slightly deviate from this value, and hence the analytical results of this paper apply. For a further development of this model, we could assume that the sex-ratio is not constant, but slightly biased according to the sex ratio in the population, with a tendency to move to compensate any deviation from the balance between the two sexes. In order to restore this balance, the probability for an offspring to be female in a generation could be taken, for example, as one minus the ratio between the number of females and the total number of individuals in the previous generation. This probability will be greater than 0.5 when there is an excess of males and less than 0.5 when there is an excess of females. As a result, we would obtain a population-size-dependent reproduction law. This feature has already been included in branching model studies in both the asexual version (see, for example, Klebaner (1989)) and the bisexual version (see, for example, Xing and Wang (2005)). Following the lines of those studies, it would therefore be feasible to obtain analytical results under the conditions of such a scenario.

Other questions for future mathematical research were also illustrated via simulation, such as the extinction time, for instance. In a simulated example, we showed that the time to extinction of a genotype can be extremely long when the other genotype grows indefinitely. The study of the probability distributions associated with these extinction times could be very relevant from a practical standpoint, for example in studying the problem of Y-chromosomal Adam. The rate of growth of a genotype given its non-extinction could also be an interesting problem to investigate. Another question for future research is to determine the probabilities of fixation of each genotype or of the simultaneous survival of both. While difficult, the analytical development of these questions may be feasible given the versatility of branching models in population dynamics modeling, and the many probabilistic tools available in the branching processes literature.

5 Proofs

First, we provide a formal definition of the Y-linked bisexual branching process. To this end, we consider two independent sequences $\{(FR_{n,l}, MR_{n,l}) : l = 1, 2, \dots; n = 0, 1, \dots\}$ and $\{(Fr_{n,l}, Mr_{n,l}) : l = 1, 2, \dots; n = 0, 1, \dots\}$ of independent, identically distributed, non-negative, and integer-valued bivariate random vectors. Intuitively, $(FR_{n,l}, MR_{n,l})$ (resp. $(Fr_{n,l}, Mr_{n,l})$) represents the number of females and males generated by the l th FMR (resp. FMr) mating unit in generation n .

With respect to the distribution of these reproduction vectors, we assume Daley's scheme (see Daley (1968)) with probability α ($0 < \alpha < 1$) for a descendant to be female. That is, the number of descendants of a FMR (resp. FMr) couple, $FR_{0,1} + MR_{0,1}$ (resp. $Fr_{0,1} + Mr_{0,1}$), is distributed according to the reproduction law $\{p_k^R\}_{k \geq 0}$ (resp. $\{p_k^r\}_{k \geq 0}$) with mean m_R (resp. m_r) and probability generating function $f_R(\cdot)$ (resp. $f_r(\cdot)$). Then, since the probability of giving birth to a female is α , the conditional distribution of $FR_{0,1}$ (resp. $Fr_{0,1}$) given that $FR_{0,1} + MR_{0,1} = k$ (resp. $Fr_{0,1} + Mr_{0,1} = k$) is binomial of size k and probability α , and therefore $E[FR_{0,1}] = \alpha m_R$ and $E[MR_{0,1}] = (1 - \alpha)m_R$ (resp. $E[Fr_{0,1}] = \alpha m_r$ and $E[Mr_{0,1}] = (1 - \alpha)m_r$).

Given the total number of FMR and FMr mating units at generation n , ZR_n and Zr_n , respectively, the offspring generated by each genotype in generation $(n + 1)$ is specified by the formulas

$$(FR_{n+1}, MR_{n+1}) = \sum_{l=1}^{ZR_n} (FR_{n,l}, MR_{n,l}), \quad (Fr_{n+1}, Mr_{n+1}) = \sum_{l=1}^{Zr_n} (Fr_{n,l}, Mr_{n,l}),$$

where the empty sum is assumed to be null. Intuitively, (FR_{n+1}, MR_{n+1}) (resp.

(Fr_{n+1}, Mr_{n+1})) represents the total number of females and males, respectively, given by all the FMR (FMr) mating units in generation n . Moreover,

$$F_{n+1} = FR_{n+1} + Fr_{n+1}$$

denotes the total number of females in generation $(n + 1)$.

Given $(F_{n+1}, MR_{n+1}, Mr_{n+1})$, the total number of FMR and FMr couples in generation $n + 1$ is given by the formulas

$$ZR_{n+1} = MR_{n+1} \text{ and } Zr_{n+1} = Mr_{n+1}, \text{ if } MR_{n+1} + Mr_{n+1} \leq F_{n+1}$$

and

$$ZR_{n+1} = X_{n+1} \text{ and } Zr_{n+1} = F_{n+1} - X_{n+1}, \text{ if } MR_{n+1} + Mr_{n+1} > F_{n+1},$$

where X_{n+1} follows a hypergeometric distribution of parameters F_{n+1} , $MR_{n+1} + Mr_{n+1}$, and MR_{n+1} (see Hush and Scovel (2005) for details about the hypergeometric distribution).

Finally, we define the σ -algebras $\mathcal{F}_n = \sigma((ZR_k, Zr_k) : k = 0, \dots, n)$ and $\mathcal{G}_n = \sigma((ZR_k, Zr_k), (FR_{k+1}, MR_{k+1}), (Fr_{k+1}, Mr_{k+1}), k = 0, \dots, n), n = 0, 1, \dots$

Proof of Result 1

The proof of this result follows the same ideas as that in Theorem 4.1 of González et al. (2006).

Proof of Result 4

We need the following lemma:

Lemma 2 *Let $i, j > 0$. If there exists a constant $A > 0$ such that*

$$\begin{aligned} E[ZR_{n+1}|(ZR_n, Zr_n)] &\leq ZR_n \text{ a.s. on } \{ZR_n \geq A, Zr_n \geq A\} \\ \text{or } E[Zr_{n+1}|(ZR_n, Zr_n)] &\leq Zr_n \text{ a.s. on } \{ZR_n \geq A, Zr_n \geq A\}, \end{aligned} \quad (1)$$

then $P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) = 0$.

Proof of the Lemma

Given $A > 0$ satisfying (1), it is enough to prove that for every $N > 0$

$$P_{(i,j)} \left(\left\{ \inf_{n \geq N} ZR_n \geq A \right\} \cap \left\{ \inf_{n \geq N} Zr_n \geq A \right\} \cap \{ZR_n \rightarrow \infty, Zr_n \rightarrow \infty\} \right) = 0. \quad (2)$$

Fixed $N > 0$, define the *stopping time*, $T(A)$, as $T(A) = \infty$ if $\inf_{n \geq N} ZR_n \geq A$ and $\inf_{n \geq N} Zr_n \geq A$, and $T(A) = \min\{n \geq N : ZR_n < A \text{ or } Zr_n < A\}$ otherwise.

Define also the sequence of random variables $\{Y_n\}_{n \geq 0}$ as follows:

$$Y_n = \begin{cases} ZR_{N+n} & \text{if } N+n \leq T(A) \\ ZR_{T(A)} & \text{if } N+n > T(A). \end{cases}$$

To obtain (2), we show that $\{Y_n\}_{n \geq 0}$ is a non-negative supermartingale with respect to $\{\mathcal{F}_{N+n}\}_{n \geq 0}$. In fact, Y_n is \mathcal{F}_{N+n} -measurable for any $n \geq 0$.

If $n \geq 0$ and $ZR_{N+k}, Zr_{N+k} \geq A, k = 0, \dots, n$, then $T(A) \geq N+n+1$, and from (1) we obtain that

$$E[Y_{n+1} | \mathcal{F}_{N+n}] = E[ZR_{N+n+1} | \mathcal{F}_{N+n}] \leq ZR_{N+n} = Y_n \quad \text{a.s.}$$

on $\{ZR_{N+k}, Zr_{N+k} \geq A : k = 0, \dots, n\}$.

If $n \geq 1$ and for some $k \in \{1, \dots, n\}$ it is satisfied that $ZR_N, \dots, ZR_{N+k-1} \geq A$ and $ZR_{N+k} < A$, or $Zr_N, \dots, Zr_{N+k-1} \geq A$ and $Zr_{N+k} < A$, then $T(A) \leq N+k < N+n+1$ and also

$$E[Y_{n+1} | \mathcal{F}_{N+n}] = E[ZR_{T(A)} | \mathcal{F}_{N+n}] = Y_n \quad \text{a.s.}$$

on $\{ZR_N, \dots, ZR_{N+k-1} \geq A, ZR_{N+k} < A\} \cup \{Zr_N, \dots, Zr_{N+k-1} \geq A, Zr_{N+k} < A\}$. Finally, if $ZR_N < A$ or $Zr_N < A$ then $T(A) = N < N+n+1$ for all $n \geq 0$, and we get that

$$E[Y_{n+1} | \mathcal{F}_{N+n}] = E[ZR_N | \mathcal{F}_{N+n}] = Y_n \quad \text{a.s. on } \{ZR_N < A\} \cup \{Zr_N < A\}.$$

In short, since $A_n = \{ZR_{N+k}, Zr_{N+k} \geq A, k = 0, \dots, n\} \in \mathcal{F}_{N+n}$, one deduces that

$$\begin{aligned} E[Y_{n+1} | \mathcal{F}_{N+n}] &= E[Y_{n+1} I_{A_n} | \mathcal{F}_{N+n}] + E[Y_{n+1} I_{A_n^c} | \mathcal{F}_{N+n}] \\ &= E[Y_{n+1} | \mathcal{F}_{N+n}] I_{A_n} + E[Y_{n+1} | \mathcal{F}_{N+n}] I_{A_n^c} \\ &\leq ZR_{N+n} I_{A_n} + ZR_{T(A)} I_{A_n^c} = Y_n \quad \text{a.s.} \end{aligned}$$

Applying the Martingale Convergence Theorem, we obtain the almost sure convergence of the sequence $\{Y_n\}_{n \geq 0}$ to a non-negative and finite limit, Y_∞ ,

where

$$Y_\infty = \begin{cases} \lim_{n \rightarrow \infty} ZR_n & \text{if } \inf_{n \geq N} ZR_n \geq A \text{ and } \inf_{n \geq N} Zr_n \geq A \\ ZR_{T(A)} & \text{otherwise.} \end{cases}$$

Therefore we deduce (2) and the proof of the Lemma is complete.

Proof of the Result

(i) Assume $(1 - \alpha)m_R \leq 1$. The proof is analogous for $(1 - \alpha)m_r \leq 1$ because of the symmetry of the model with respect to the two genotypes. Using the definition of the process,

$$E[ZR_n | \mathcal{G}_{n-1}] = \begin{cases} MR_n & \text{if } F_n \geq MR_n + Mr_n \\ MR_n \frac{F_n}{MR_n + Mr_n} & \text{if } F_n < MR_n + Mr_n \end{cases} \text{ a.s.,}$$

so that $E[ZR_n | \mathcal{G}_{n-1}] \leq MR_n$ a.s. and, since $\mathcal{F}_n \subseteq \mathcal{G}_n$, $n = 0, 1, \dots$,

$$\begin{aligned} E[ZR_n | \mathcal{F}_{n-1}] &= E[E[ZR_n | \mathcal{G}_{n-1}] | \mathcal{F}_{n-1}] \leq E[MR_n | \mathcal{F}_{n-1}] \\ &= (1 - \alpha)m_R ZR_{n-1} \leq ZR_{n-1} \text{ a.s.} \end{aligned}$$

Applying the previous Lemma, we obtain that $P(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) = 0$.

(ii) Assume $\alpha m_R < 1$. The proof is analogous for $\alpha m_r < 1$. Assume also $\alpha < 0.5$. Otherwise the result is deduced from (i).

According to the previous Lemma we have to show (1) for some constant A . But, using the definition of the process,

$$\begin{aligned} E[ZR_n | \mathcal{G}_{n-1}] &= MR_n I_{\{F_n \geq MR_n + Mr_n\}} + \frac{MR_n F_n}{MR_n + Mr_n} I_{\{F_n < MR_n + Mr_n\}} \\ &\leq MR_n I_{\{FR_n \geq MR_n\}} + MR_n I_{\{Fr_n \geq Mr_n\}} + \frac{MR_n F_n}{MR_n + Mr_n} I_{\{F_n < MR_n + Mr_n\}} \text{ a.s.} \end{aligned} \tag{3}$$

Let us bound properly the conditional expectation given \mathcal{F}_{n-1} of each of these summands.

For the first summand of (3), since we are assuming Daley's scheme, we can apply the bounds for the tails of a binomial distribution given by Okamoto's inequality (see Johnson et al. (1993)), and we have, for all n and $k > 0$,

$$\begin{aligned} P(FR_n \geq MR_n | FR_n + MR_n = k) &= P(FR_n \geq k/2 | FR_n + MR_n = k) \\ &\leq e^{-k(0.5-\alpha)^2} = a_1^k, \end{aligned}$$

with $a_1 = e^{-(0.5-\alpha)^2} < 1$. Then, for all $i', j' > 0$

$$\begin{aligned} & P(FR_n \geq MR_n | ZR_{n-1} = i', Zr_{n-1} = j') \\ &= E[P(FR_n \geq MR_n | FR_n + MR_n) | ZR_{n-1} = i', Zr_{n-1} = j'] \\ &\leq E[a_1^{FR_n + MR_n} | ZR_{n-1} = i', Zr_{n-1} = j'] = f_R(a_1)^{i'}. \end{aligned} \quad (4)$$

Therefore, applying (4) and the Cauchy-Schwartz inequality,

$$\begin{aligned} E[MR_n I_{\{FR_n \geq MR_n\}} | \mathcal{F}_{n-1}] &\leq E[MR_n^2 | \mathcal{F}_{n-1}]^{1/2} P(FR_n \geq MR_n | \mathcal{F}_{n-1})^{1/2} \\ &\leq K_1 ZR_{n-1} f_R(a_1)^{ZR_{n-1}/2} \quad \text{a.s.} \end{aligned} \quad (5)$$

for some positive constant K_1 .

For the second summand of (3), we can proceed analogously and obtain

$$E[MR_n I_{\{Fr_n \geq Mr_n\}} | \mathcal{F}_{n-1}] \leq K_2 ZR_{n-1} f_r(a_1)^{Zr_{n-1}/2} \quad \text{a.s.} \quad (6)$$

for some positive constant K_2 .

To bound the third summand of (3), given $\varepsilon > 0$, define $\gamma_1 = \alpha(m_R - \varepsilon)$, $\gamma_2 = \alpha(m_R + \varepsilon)$, $\gamma_3 = (1 - \alpha)(m_R - \varepsilon)$, $\gamma_4 = (1 - \alpha)(m_R + \varepsilon)$, $\gamma_5 = \alpha(m_r - \varepsilon)$, $\gamma_6 = \alpha(m_r + \varepsilon)$, $\gamma_7 = (1 - \alpha)(m_r - \varepsilon)$, $\gamma_8 = (1 - \alpha)(m_r + \varepsilon)$, and $B_\varepsilon = 1 + 2\varepsilon/(\min\{m_R, m_r\} - \varepsilon)$. We take ε such that $\gamma_1, \gamma_5 > 0$ and $0 < \gamma_2 B_\varepsilon < 1$.

For each $n \geq 1$, define also

$$\begin{aligned} A_{FR,n} &= \{|FR_n - ZR_{n-1}\alpha m_R| \leq ZR_{n-1}\alpha\varepsilon\}, \\ A_{MR,n} &= \{|MR_n - ZR_{n-1}(1 - \alpha)m_R| \leq ZR_{n-1}(1 - \alpha)\varepsilon\}, \\ A_{Fr,n} &= \{|Fr_n - Zr_{n-1}\alpha m_r| \leq Zr_{n-1}\alpha\varepsilon\}, \\ A_{Mr,n} &= \{|Mr_n - Zr_{n-1}(1 - \alpha)m_r| \leq Zr_{n-1}(1 - \alpha)\varepsilon\}. \end{aligned}$$

From now on, n will be dropped in the notation if there is no ambiguity.

Since the reproduction laws are assumed to have finite variances, an immediate application of Chebyshev's inequality gives

$$P(A_{FR}^c \cup A_{MR}^c | ZR_{n-1} = i', Zr_{n-1} = j') \leq \frac{C_1}{i'} \quad (7)$$

and

$$P(A_{Fr}^c \cup A_{Mr}^c | ZR_{n-1} = i', Zr_{n-1} = j') \leq \frac{C_2}{j'} \quad (8)$$

for some positive constants C_1 and C_2 .

Now, if we denote $D = A_{FR} \cap A_{MR} \cap A_{Fr} \cap A_{Mr}$,

$$\begin{aligned}
& E \left[\frac{MR_n F_n}{MR_n + Mr_n} I_{\{F_n < MR_n + Mr_n\}} | \mathcal{F}_{n-1} \right] \\
& \leq E \left[\frac{MR_n F_n}{MR_n + Mr_n} I_{\{F_n < MR_n + Mr_n\}} I_D | \mathcal{F}_{n-1} \right] \\
& \quad + E \left[\frac{MR_n F_n}{MR_n + Mr_n} I_{\{F_n < MR_n + Mr_n\}} I_{A_{FR}^c \cup A_{MR}^c} | \mathcal{F}_{n-1} \right] \\
& \quad + E \left[\frac{MR_n F_n}{MR_n + Mr_n} I_{\{F_n < MR_n + Mr_n\}} I_{A_{Fr}^c \cup A_{Mr}^c} | \mathcal{F}_{n-1} \right] \quad \text{a.s.} \quad (9)
\end{aligned}$$

Applying (7) and the Cauchy-Schwartz inequality,

$$\begin{aligned}
& E \left[\frac{MR_n F_n}{MR_n + Mr_n} I_{\{F_n < MR_n + Mr_n\}} I_{A_{FR}^c \cup A_{MR}^c} | \mathcal{F}_{n-1} \right] \leq E \left[MR_n I_{A_{FR}^c \cup A_{MR}^c} | \mathcal{F}_{n-1} \right] \\
& \leq E [MR_n^2 | \mathcal{F}_{n-1}]^{1/2} P(A_{FR}^c \cup A_{MR}^c | \mathcal{F}_{n-1})^{1/2} \\
& \leq (K'_3 ZR_{n-1})(C_1 ZR_{n-1})^{-1/2} = K_3 ZR_{n-1}^{1/2} \quad \text{a.s.} \quad (10)
\end{aligned}$$

for some positive constants K'_3 and K_3 .

Analogously,

$$E \left[\frac{MR_n F_n}{MR_n + Mr_n} I_{\{F_n < MR_n + Mr_n\}} I_{A_{Fr}^c \cup A_{Mr}^c} | \mathcal{F}_{n-1} \right] \leq K_4 ZR_{n-1} Zr_{n-1}^{-1/2} \quad \text{a.s.} \quad (11)$$

for some positive constant K_4 .

Finally,

$$\begin{aligned}
& E \left[\frac{MR_n F_n}{MR_n + Mr_n} I_{\{F_n < MR_n + Mr_n\}} I_D | \mathcal{F}_{n-1} \right] \\
& \leq \gamma_4 ZR_{n-1} \frac{\gamma_2 ZR_{n-1} + \gamma_6 Zr_{n-1}}{\gamma_3 ZR_{n-1} + \gamma_7 Zr_{n-1}} \\
& = \alpha(m_R + \varepsilon) ZR_{n-1} \frac{(m_R + \varepsilon) ZR_{n-1} + (m_r + \varepsilon) Zr_{n-1}}{(m_R - \varepsilon) ZR_{n-1} + (m_r - \varepsilon) Zr_{n-1}} \\
& = \gamma_2 ZR_{n-1} \frac{(m_R - \varepsilon) ZR_{n-1} + 2\varepsilon ZR_{n-1} + (m_r - \varepsilon) Zr_{n-1} + 2\varepsilon Zr_{n-1}}{(m_R - \varepsilon) ZR_{n-1} + (m_r - \varepsilon) Zr_{n-1}} \\
& \leq \gamma_2 ZR_{n-1} \left(1 + \frac{2\varepsilon}{\min\{m_R - \varepsilon, m_r - \varepsilon\}} \right) = \gamma_2 B_\varepsilon ZR_{n-1} \quad \text{a.s.} \quad (12)
\end{aligned}$$

Summarizing, from (5), (6), (9), (10), (11), and (12), we deduce that

$$\begin{aligned}
E[ZR_n | \mathcal{F}_{n-1}] & \leq (K_1 f_R(a_1))^{ZR_{n-1}/2} + (K_2 f_r(a_1))^{Zr_{n-1}/2} \\
& \quad + K_3 ZR_{n-1}^{-1/2} + K_4 Zr_{n-1}^{-1/2} + \gamma_2 B_\varepsilon ZR_{n-1} \quad \text{a.s.}
\end{aligned}$$

Since $a_1 < 1$ and $\gamma_2 B_\varepsilon < 1$, we can take $A > 0$ such that, for $ZR_{n-1} > A$ and $Zr_{n-1} > A$, the term in parentheses is less than 1, so that (1) holds and therefore the result is proved.

Proof of Result 6

For each $\eta_1, \eta_2 > 1$, let $A_n = \{ZR_{n+1} > \eta_1 ZR_n, Zr_{n+1} > \eta_2 Zr_n\}$ for all $n \geq 0$

$$\begin{aligned} P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) &\geq P_{(i,j)}\left(\bigcap_{n=0}^{\infty} \{ZR_{n+1} > \eta_1 ZR_n, Zr_{n+1} > \eta_2 Zr_n\}\right) \\ &= \lim_{n \rightarrow \infty} P_{(i,j)}\left(\bigcap_{l=0}^n A_l\right) = \lim_{n \rightarrow \infty} P_{(i,j)}(A_0) \prod_{l=1}^n P_{(i,j)}\left(A_l \mid \bigcap_{k=0}^{l-1} A_k\right). \end{aligned} \quad (13)$$

Moreover, using the Markov property of $\{(ZR_n, Zr_n)\}_{n \geq 0}$, for all $n > 0$

$$\begin{aligned} P_{(i,j)}\left(A_n \mid \bigcap_{k=0}^{n-1} A_k\right) &= P_{(i,j)}\left(A_n \mid \bigcup_{i',j' > 0} \{(ZR_n, Zr_n) = (i', j')\} \cap \bigcap_{k=0}^{n-1} A_k\right) \\ &\geq \inf_{i' > \eta_1^n i, j' > \eta_2^n j} P_{(i',j')}\left(A_n \mid \{(ZR_n, Zr_n) = (i', j')\} \cap \bigcap_{k=0}^{n-1} A_k\right) \\ &= \inf_{i' > \eta_1^n i, j' > \eta_2^n j} P_{(i',j')}(A_0). \end{aligned} \quad (14)$$

Let us bound conveniently $P_{(i',j')}(A_0)$ for all $i', j' > 0$ or equivalently $P_{(i',j')}(A_0^c)$. First we assume that $\alpha > 0.5$, and both $(1 - \alpha)m_R$ and $(1 - \alpha)m_r$ are greater than 1. Take $\varepsilon > 0$, $\eta_1 = (1 - \alpha)m_R - \varepsilon$, and $\eta_2 = (1 - \alpha)m_r - \varepsilon$ such that $\eta_1, \eta_2 > 1$. Then

$$\begin{aligned} A_0^c &= \{ZR_1 \leq \eta_1 ZR_0\} \cup \{Zr_1 \leq \eta_2 Zr_0\} \\ &\subseteq \{ZR_1 \leq \eta_1 ZR_0, MR_1 > \eta_1 ZR_0, FR_1 > MR_1, Fr_1 > Mr_1\} \\ &\quad \cup \{MR_1 \leq \eta_1 ZR_0\} \cup \{FR_1 \leq MR_1\} \\ &\quad \cup \{Zr_1 \leq \eta_2 Zr_0, Mr_1 > \eta_2 Zr_0, FR_1 > MR_1, Fr_1 > Mr_1\} \\ &\quad \cup \{Mr_1 \leq \eta_2 Zr_0\} \cup \{Fr_1 \leq Mr_1\}. \end{aligned} \quad (15)$$

Since $ZR_1 = MR_1$ and $Zr_1 = Mr_1$ if $MR_1 + Mr_1 < F_1$, then we have that

$$P_{(i',j')}(ZR_1 \leq \eta_1 ZR_0, MR_1 > \eta_1 ZR_0, FR_1 > MR_1, Fr_1 > Mr_1) = 0 \quad (16)$$

and

$$P_{(i',j')}(Zr_1 \leq \eta_2 Zr_0, Mr_1 > \eta_2 Zr_0, FR_1 > MR_1, Fr_1 > Mr_1) = 0. \quad (17)$$

Moreover, since the reproduction laws are assumed to have finite variances, by Chebyshev's inequality, it follows that

$$P_{(i',j')}(MR_1 \leq \eta_1 ZR_0) = P\left(\sum_{k=1}^{i'}(MR_{k0} - (1-\alpha)m_R) \leq -\varepsilon i'\right) \leq \frac{C_1}{i'}, \quad (18)$$

and analogously

$$P_{(i',j')}(Mr_1 \leq \eta_2 Zr_0) \leq \frac{C_2}{j'}, \quad (19)$$

for some constants $C_1, C_2 > 0$.

In the same way as (4), we obtain, for some constant $a_1 < 1$, that

$$P_{(i',j')}(FR_1 \leq MR_1) \leq f_R(a_1)^{i'} \text{ and } P_{(i',j')}(Fr_1 \leq Mr_1) \leq f_r(a_1)^{j'}. \quad (20)$$

From (15), (16), (17), (18), (19), and (20), we have

$$P_{(i',j')}(A_0) = 1 - P_{(i',j')}(A_0^c) \geq 1 - \frac{C_1}{i'} - \frac{C_2}{j'} - f_R(a_1)^{i'} - f_r(a_1)^{j'},$$

and therefore, since $\eta_1, \eta_2 > 1$ and $f_R(a_1), f_r(a_1) < 1$, from (13) and (14) it follows that

$$\begin{aligned} & P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) \\ & \geq P_{(i,j)}(A_0) \lim_{n \rightarrow \infty} \prod_{l=1}^n \inf_{i' > \eta_1^l i, j' > \eta_2^l j} P_{(i',j')}(A_0) \\ & \geq P_{(i,j)}(A_0) \lim_{n \rightarrow \infty} \prod_{l=1}^n \left(1 - \frac{C_1}{\eta_1^l i} - \frac{C_2}{\eta_2^l j} - f_R(a_1)^{\eta_1^l i} - f_r(a_1)^{\eta_2^l j}\right) > 0, \end{aligned}$$

and the proof is complete for $\alpha > 0.5$.

Assume now that $\alpha < 0.5$, and both αm_R and αm_r are greater than 1. Given $\varepsilon > 0$, define the constants $\gamma_1, \dots, \gamma_8$ and the sets $A_{FR}, A_{MR}, A_{Fr}, A_{Mr}$ and D as in the proof of Result 4. Take ε small enough so that $\gamma_2 < \gamma_3, \gamma_6 < \gamma_7$, and

$$\gamma_1 \left(1 - \frac{2\varepsilon}{\min\{m_R + \varepsilon, m_r + \varepsilon\}}\right) > 1, \quad \gamma_5 \left(1 - \frac{2\varepsilon}{\min\{m_R + \varepsilon, m_r + \varepsilon\}}\right) > 1,$$

and choose

$$1 < \eta_1 < \gamma_1 \left(1 - \frac{2\varepsilon}{\min\{m_R + \varepsilon, m_r + \varepsilon\}}\right)$$

and

$$1 < \eta_2 < \gamma_5 \left(1 - \frac{2\varepsilon}{\min\{m_R + \varepsilon, m_r + \varepsilon\}} \right).$$

Specifically, we take $\eta_1 = \gamma_1(1 - 3\varepsilon/\min\{m_R + \varepsilon, m_r + \varepsilon\})$ and $\eta_2 = \gamma_5(1 - 3\varepsilon/\min\{m_R + \varepsilon, m_r + \varepsilon\})$.

With this notation

$$A_0^c \subseteq (\{ZR_1 \leq \eta_1 ZR_0\} \cap D) \cup (\{Zr_1 \leq \eta_2 Zr_0\} \cap D) \cup D^c. \quad (21)$$

Equations (7) and (8) provide a suitable bound for $P_{(i',j')}(D^c)$, so we focus our attention on the two first terms of (21).

Since $\gamma_2 < \gamma_3$ and $\gamma_6 < \gamma_7$, $F_1 < MR_1 + Mr_1$ on D , and, by the definition of the process, the conditional distribution of ZR_1 given $\mathcal{G}_0 = \sigma(F_1, MR_1, Mr_1)$ is hypergeometric. Hence

$$\begin{aligned} P_{(i',j')}(\{ZR_1 \leq \eta_1 ZR_0\} \cap D) &= E_{(i',j')}[P_{(i',j')}(ZR_1 \leq \eta_1 ZR_0 | \mathcal{G}_0) I_D] \\ &= E_{(i',j')} \left[P_{(i',j')} \left(ZR_1 - E_{(i',j')}[ZR_1 | \mathcal{G}_0] \leq \eta_1 i' - \frac{MR_1 F_1}{MR_1 + Mr_1} \middle| \mathcal{G}_0 \right) I_D \right] \end{aligned} \quad (22)$$

But on D

$$\begin{aligned} \eta_1 i' - \frac{MR_1 F_1}{MR_1 + Mr_1} &\leq \eta_1 i' - \gamma_3 i' \frac{\gamma_1 i' + \gamma_5 j'}{\gamma_4 i' + \gamma_8 j'} \\ &= \eta_1 i' - \gamma_1 i' \frac{(m_R - \varepsilon) i' + (m_r - \varepsilon) j'}{(m_R + \varepsilon) i' + (m_r + \varepsilon) j'} \\ &= \eta_1 i' - \gamma_1 i' \frac{(m_R + \varepsilon) i' + (m_r + \varepsilon) j' - 2\varepsilon(i' + j')}{(m_R + \varepsilon) i' + (m_r + \varepsilon) j'} \\ &\leq \left[\eta_1 - \gamma_1 \left(1 - \frac{2\varepsilon}{\min\{m_R + \varepsilon, m_r + \varepsilon\}} \right) \right] i' \\ &= - \frac{\varepsilon i'}{\min\{m_R + \varepsilon, m_r + \varepsilon\}} \gamma_1. \end{aligned}$$

Let us write $\delta = \gamma_1 \varepsilon / \min\{m_R + \varepsilon, m_r + \varepsilon\}$. Applying the previous inequality and the bounds for the tails of a hypergeometric distribution provided in Hush and Scovel (2005), for i' large enough, we can deduce from (22)

$$\begin{aligned} &P_{(i',j')}(\{ZR_1 \leq \eta_1 ZR_0\} \cap D) \\ &\leq E_{(i',j')} \left[P_{(i',j')} \left(ZR_1 - E_{(i',j')}[ZR_1 | \mathcal{G}_0] \leq -\delta i' \middle| \mathcal{G}_0 \right) I_D \right] \\ &\leq E_{(i',j')} \left[\exp \left\{ -2 \frac{\delta^2 i'^2 - 1}{MR_n + 1} \right\} I_D \right] \leq \exp \left\{ -2 \frac{\delta^2 i'^2 - 1}{\gamma_4 i' + 1} \right\} \leq K_1 e^{-B_1 i'} \end{aligned} \quad (23)$$

for some positive constants K_1 and B_1 . Notice that, without loss of generality, we can take i , and consequently i' , large enough because all the states with both coordinates non-null are communicating.

Analogously, it can be deduced that

$$P_{(i',j')}(\{Zr_1 \leq \eta_1 Zr_0\} \cap D) \leq K_2 e^{-B_2 j'} \quad (24)$$

for some positive constants K_2 and B_2 .

Then, taking into account the decomposition in (21), from (7), (8), (23), and (24), we deduce that

$$P_{(i',j')}(A_0) = 1 - P_{(i',j')}(A_0^c) \geq 1 - K_1 e^{-B_1 i'} - K_2 e^{-B_2 j'} - \frac{C_1}{i'} - \frac{C_2}{j'},$$

and therefore, since $\eta_1, \eta_2 > 1$ and $B_1, B_2 > 0$, from (13) and (14) it follows that

$$\begin{aligned} P_{(i,j)}(ZR_n \rightarrow \infty, Zr_n \rightarrow \infty) &\geq P_{(i,j)}(A_0) \lim_{n \rightarrow \infty} \prod_{l=1}^n \inf_{i' > \eta_1^l i, j' > \eta_2^l j} P_{(i',j')}(A_0) \\ &\geq P_{(i,j)}(A_0) \lim_{n \rightarrow \infty} \prod_{l=1}^n \left(1 - K_1 e^{-B_1 \eta_1^l i} - K_2 e^{-B_2 \eta_2^l j} - \frac{C_1}{\eta_1^l i} - \frac{C_2}{\eta_2^l j} \right) > 0, \end{aligned}$$

which completes the proof for $\alpha > 0.5$.

Acknowledgement

We would like to thank the anonymous referees for their constructive comments and interesting suggestions which have improved this paper.

References

- Abe, H., Fujii, T., Tanaka, N., Yokoyama, T., Kakehashi, H., Ajimura, M., Mita, K., Banno, Y., Yasukochi, Y., Oshiki, T., Neno, M., Ishikawa, T., Shimada, T., 2008. Identification of the female-determining region of the W chromosome in *Bombyx mori*. *Genetica* 133, 269–282.
- Alsmeyer, G., Rösler, U., 1996. The bisexual Galton-Watson branching process with promiscuous mating: extinction probabilities in the supercritical case. *Ann. Appl. Probab.* 6, 922–939.
- Alsmeyer, G., Rösler, U., 2002. Asexual versus promiscuous bisexual Galton-Watson branching processes: the extinction probability ratio. *Ann. Appl. Probab.* 12, 125–142.
- Bagley, J. H., 1986. On the asymptotic properties of a supercritical bisexual branching process. *J. Appl. Probab.* 23, 820–826.

- Ball, F., Britton, T., Lyne, O., 2004. Stochastic multitype epidemics in a community of households: estimation of threshold parameter R^* and secure vaccination coverage. *Biometrika* 91(2), 345–362.
- Bernardo, A., Dobo, B., Vibranovski, M., Clark, A., 2001. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 98 (23), 13225–13230.
- Bowden, G., Balaesque, P., King, T., Hansen, Z., Lee, A., Pergl-Wilson, G., Hurley, E., Roberts, S., Waite, P., Jesch, J., Jones, A., Thomas, M., Harding, S., Jobling, M., 2008. Excavating past population structures by surname-based sampling: the genetic legacy of the Vikings in northwest England. *Molecular Biology and Evolution* 25 (2), 301–309.
- Bruss, F. T., 1984. A note on extinction criteria for bisexual Galton-Watson processes. *J. Appl. Probab.* 21, 915–919.
- Calogero, A., Garofalo, M., Barone, N., Longo, G., De Palma, A., Fichera, M., Rappazzo, G., D’Agata, R., Vicari, E., 2002. Spontaneous transmission from a father to his son of a Y chromosome microdeletion involving the deleted in azoospermia(DAZ) gene. *J Endocrinol Invest* 25, 631–634.
- Campbell, R., 2003. A logistic branching process for population genetics. *Journal of Theoretical Biology* 225 (2), 195–203.
- Caron-Lormier, G., Masson, J., Mnard, N., Pierre, J., 2006. A branching process, its application in biology: Influence of demographic parameters on the social structure in mammal groups. *Journal of Theoretical Biology* 238 (3), 564–574.
- Charlesworth, D., Charlesworth, B., Marais, G., 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95, 118–128.
- Daley, D. J., Hull, D. M., Taylor, J. M., 1986. Bisexual Galton-Watson branching processes with superadditive mating functions. *J. Appl. Probab.* 23, 585–600.
- Daley, D. J., 1968. Extinction conditions for certain bisexual Galton-Watson branching processes. *Z. Wahrscheinlichkeitsth.* 9, 315–322.
- Fitch, N., Richer, C., Pinsky, L., Kahn, A., Opitz, J., Reynolds, J., 2005. Deletion of the long arm of the Y chromosome and review of Y chromosome abnormalities. *American Journal of Medical Genetics* 20 (1), 31–42.
- Galton, F., 1873. Problem 4001: On the extinction of surnames. *Educational Times* 26, 1717.
- Garske, T., Rhodes, C., 2008. The effect of superspreading on epidemic outbreak size distributions. *Journal of Theoretical Biology* 253 (2), 228–237.
- Geraldes, A., Rogel-Gaillard, C., Ferrand, N., 2005. High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) SRY gene. *Animal Genetics* 36 (4), 349–351.
- González-Fragoso, A., 1995. Ratio estimation for the offspring means of a bisexual Galton–Watson branching process. *J. Inter-American Stat. Inst.* 47, 17–36.
- González, M., Hull, D. M., Martínez, R., Mota, M., 2006. Bisexual branching processes in a genetic context: The extinction problem for Y-linked genes.

- Math. Biosci. 202, 227–247.
- González, M., Molina, M., Mota, M., 2001. Nonparametric estimation of the offspring distribution and the mean vector for a bisexual Galton-Watson process. *Comm. Statist. Theory Methods* 30, 497–516.
- González, M., Molina, M., 1996. On the limit behaviour of a superadditive bisexual Galton-Watson branching process. *J. Appl. Probab.* 33, 960–967.
- González, M., Molina, M., 1997. On the L^2 -convergence of a superadditive bisexual Galton-Watson branching process. *J. Appl. Probab.* 34, 575–582.
- Graves, J., 2006. Sex chromosome specialization and degeneration in mammals. *Cell* 124(5), 901–914.
- Gutiérrez, J., Teem, J., 2006. A model describing the effect of sex-reversed YY fish in an established wild population: The use of a trojan Y chromosome to cause extinction of an introduced exotic species. *Journal of Theoretical Biology* 241, 333–341.
- Haccou, P., Jagers, P., Vatutin, V., 2005. *Branching processes: variation, growth and extinction of populations*. Cambridge University Press.
- Hellborg, L., Gndz, I., Jaarola, M., 2005. Analysis of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology* 14 (7), 2025–2031.
- Hughes, J. F., Skaletsky, H., Pyntikova, T., Minx, P. J., Graves, T., Rozen, S., Wilson, R. K., Page, D. C., 2005. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* 437, 100–103.
- Hull, D. M., 1982. A necessary condition for extinction in those bisexual Galton-Watson branching processes governed by superadditive mating functions. *J. Appl. Probab.* 19, 847–850.
- Hull, D. M., 1984. Conditions for extinction in certain bisexual Galton-Watson branching processes. *J. Appl. Probab.* 21, 414–418.
- Hull, D. M., 2003. A survey of the literature associated with the bisexual Galton-Watson branching processes. *Extracta Math.* 13, 321–343.
- Hurles, M., Irven, C., Nicholson, J., Taylor, P., Santos, F., Loughlin, J., Jobling, M., Sykes, B., 1998. European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. *Am J Hum Genet.* 63 (6), 1793–1806.
- Hurles, M., Nicholson, J., Bosch, E., Renfrew, C., Sykes, B., Jobling, M., 2002. Y chromosomal evidence for the origins of Oceanic-speaking peoples. *Genetics* 160 (1), 289–303.
- Hush, D., Scovel, C., 2005. Concentration of the hypergeometric distribution. *Statist. Probab. Lett.* 75, 127–132.
- Iwasa, Y., Michor, F., Komarova, N., Nowak, M., 2005. Population genetics of tumor suppressor genes. *Journal of Theoretical Biology* 233 (1), 15–23.
- Johnson, N. L., Kemp, A. W., Kotz, S., 1993. *Univariate discrete distributions*. John Wiley and Sons, Inc.
- Kimmel, M., Axelrod, D., 2002. *Branching processes in biology*. Springer-Verlag.

- King, T., Ballereau, S., Schrer, K., Jobling, M., 2006. Genetic signatures of coancestry within surnames. *Curr. Biol.* 16 (4), 384–388.
- Klebaner, F., 1989. Geometric growth in near-supercritical population-size-dependent multitype Galton-Watson processes. *Ann. Prob.* 17, 1466–1477.
- Krausz, C., Forti, G., McElreavey, K., 2003. The Y chromosome and male fertility and infertility. *Int J Androl* 26, 70–75.
- Krausz, C., McElreavey, K., 2001. Y chromosome microdeletions in fertile males. *Hum Reprod.* 16 (6), 1306–1307.
- Krausz, C., Quintana-Murci, L., Forti, G., 2004. Y chromosome polymorphisms in medicine. *Ann Med.* 36 (8), 573–583.
- Kuhnert, B., Gromoll, J., Kostova, E., Tschanter, P., Luetjens, C., Simoni, M., Nieschlag, E., 2004. Case report: natural transmission of an AZFc Y-chromosomal microdeletion from father to his sons. *Hum Reprod.* 19, 886–888.
- Mode, C. J., Sleeman, C. K., 2000. *Stochastic processes in epidemiology.* World Scientific.
- Molina, M., González, M., Mota, M., 1998. Bayesian inference for bisexual Galton-Watson processes. *Comm. Statist. Theory Methods* 27, 1055–1070.
- Neves, A., Moreira, C., 2006. Applications of the Galton-Watson process to human DNA evolution and demography. *Physica A* 368, 132–146.
- O’Connell, N., 1995. The genealogy of branching processes and the age of our more recent common ancestor. *Adv. Appl. Probab.* 27, 418–442.
- Ogawa, A., Murata, K., Mizuno, S., 1998. The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. *Proc. Natl. Acad. Sci. USA* 95, 4415–4418.
- Pakes, A., 2003. Biological applications of branching processes. *Handbook of Statistic Vol. 21 Stochastic Processes: Modelling and Simulation* (Shanbhag, D.N. and Rao, C.R., eds.) Chapter 18, 693–773, Elsevier Science B.V.
- Patsalis, P., Sismani, C., Quintana-Murci, L., Taleb-Bekkouche, F., Krausz, C., McElreavey, K., 2002. Effects of transmission of Y chromosome AZFc deletions. *The Lancet* 360 (9341), 1222–1224.
- Quintana-Murci, L., Fellous, M., 2001. The human Y chromosome: the biological role of a “functional wasteland”. *J. Biomed. Biotechnol.* 1, 18–24.
- Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S. H., Hammer, M. F., Mehdi, S., Ayub, Q., Qamar, R., Mohyuddin, A., Radhakrishna, U., Jobling, M., Tyler-Smith, C., McElreavey, K., 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet.* 68 (2), 537–542.
- Rosa, A., Ornelas, C., Jobling, M., Brehm, A., Villems, R., 2007. Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. *BMC Evol Biol.* 27, 107–124.
- Saut, N., Terriou, P., Navarro, A., Lvy, N., Mitchell, M., 2002. The human Y chromosome genes BPY2, CDY1 and DAZ are not essential for sustained fertility. *Mol. Hum. Reprod.* 6, 789–793.
- Tosi, A., Morales, J., Melnick, D., 2002. Y-chromosome and mitochondrial

- markers in *Macaca fascicularis* indicate introgression with Indochinese *M. mulatta* and a biogeographic barrier in the Isthmus of Kra. *Int. J. Primatol.* 23 (1), 161–178.
- Toure, A., Szot, M., Mahadevaiah, S., Rattigan, A., Ojarikre, O., Burgoyne, P., 2004. A new deletion of the mouse Y chromosome long arm associated with the loss of Ssty expression, abnormal sperm development and sterility. *Genetics* 166, 901–912.
- Wallis, M., Waters, P., Graves, J., 2008. Sex determination in mammals - before and after the evolution of SRY. *Cell. Mol. Life Sci.* to appear. doi:10.1007/s00018-008-8109-z.
- Xing, Y., Wang, Y., 2005. On the extinction of a class of population-size-dependent bisexual branching processes. *J. Appl. Probab.* 42 (1), 175184.
- Yakovlev, A., Yanev, N., 2006. Branching stochastic processes with immigration in analysis of renewing cell populations. *Math. Biosci.* 203, 37–63.
- Yamada, D., Koyama, Y., Komatsubara, M., Urabe, M., Mori, M., Hashimoto, Y., Nii, R., Kobayashi, M., Nakamoto, A., Ogihara, J., Kato, J., Mizuno, S., 2004. Comprehensive search for chicken W chromosome-linked genes expressed in early female embryos from the female-minus-male subtracted cDNA macroarray. *Chromosome Research*, 12, 741–754.