PLISKA

STUDIA MATHEMATICA
BULGARICA

ПЛИСКА

БЪЛГАРСКИ
МАТЕМАТИЧЕСКИ
СТУДИИ

ON Y-LINKED GENES AND BISEXUAL BRANCHING PROCESSES[*]

M. González, C. Gutiérrez, R. Martínez, M. Mota

In this paper we survey the results concerning the extinction problem for a two-allele Y-linked gene in a two-sex monogamic population, with a preference of females for males carrying one of the two alleles of the gene. First we give the mathematical definition of the Y-linked bisexual branching process to model this situation and study some of its relevant properties. Then, we research the extinction of the population and also the survival of each genotype depending on the behaviour of the other genotype. Finally, we simulate the evolution of the population and conjecture its long term behaviour, for some critical situations.

1. Introduction

Recent investigations in genetics have shown the importance of some genes linked to the Y-chromosome in both, human (see for example the web page `www.nature.com/nature/focus/ychromosome/`, Quintana-Murci and Fellous (2001), Krausz et al. (2003, 2004) or Hughes et al. (2005)) and other animal populations (see for example Bernardo et al. (2001), Yamada et al. (2004) or

the review by Charles et al. (2005)). Since this chromosome is exclusive of male individuals, it is natural to think that the characters controlled by such genes can have some influence on the mating process of such species, with a preference of females for males carrying one of the alleles of the gene. On the basis of this fact, González et al. (2006) have developed a bidimensional bisexual branching process to describe the generation-by-generation evolution of the number of carriers of the two alleles, R and r, of a Y-linked gene in a two-sex monogamic population. In this scenario, females do not have the gene and then the couples (a female and a male) will be classified as R-type or r-type, depending on the genotype of the male. Following the inheritance rules, R-type mating units can generate females and R-type males while r-type mating units can produce females and r-type males. Moreover, a mating mechanism, where each individual mates with one individual of the opposite sex, if available (perfect fidelity mating), is considered. Furthermore, R-type males are preferred by females as mates, though a female will mate with an r-male if no R-male is available.

Using this model, the extinction problem for a Y-linked gene has been considered in González et al. (2006, 2008), providing conditions for the almost sure extinction of the whole population and also for the survival of each genotype with positive probability. Specifically, González et al. (2006) consider the problem of the extinction of the population, the survival of a genotype when the another one becomes extinct and begin the study of the simultaneous survival of both genotypes. This study was completed in González et al. (2008), along with the investigation of the growth rate of each genotype. In this work we present the results of both papers. For the proofs, the reader will be referred to González et al. (2006, 2008).

This paper contains three sections. In Section 2 we provide the definition of the Y-linked bisexual branching process to model perfect fidelity mating and preference. 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, in the latter case considering the behaviour of the other genotype. For some critical situations, we also simulate the evolution of the population and conjecture its long term behaviour.

2.   The probability model

The probability model we are concerned with is the Y-linked bisexual branching process, introduced by González et al. (2006), where perfect fidelity mating and preference of females for R-type males are assumed.

To provide a formal definition of this model, we consider two independent

sequences $\{(FR_{n,l}, MR_{n,l}) : l = 1, 2, \ldots; n = 0, 1, \ldots\}$ and $\{(Fr_{n,l}, Mr_{n,l}) : l = 1, 2, \ldots; n = 0, 1, \ldots\}$ of independent, identically distributed, non-negative and integer-valued bivariate random vectors on the same probability triple $(\Omega, \mathcal{F}, P)$, where (FR_{01}, MR_{01}) and (Fr_{01}, Mr_{01}) may have different distributions. Intuitively, the vector $(FR_{n,l}, MR_{n,l})$ (resp. $(Fr_{n,l}, Mr_{n,l})$) represents the number of females and males generated by the lth R-type (resp. r-type) mating unit in generation n.

With respect to the distribution of these vectors, we assume the reproduction scheme given in Daley (1968). That is, each R-type (resp. r-type) couple produces offspring, independently of one another, as specified by a given probability distribution $\{p_k^R\}_{k \geq 0}$ (resp. $\{p_k^r\}_{k \geq 0}$), with mean m_R (resp. m_r). An offspring will be female with probability α, $0 < \alpha < 1$, and male with probability $1 - \alpha$. These sex designations are made independently among the offspring of any couple. Then, if a R-type (resp. r-type) mating unit produces k offspring, i.e. $FR_{0,1} + MR_{0,1} = k$ (resp. $Fr_{0,1} + Mr_{0,1} = k$), the number of females among these, $FR_{0,1}$ (resp. $Fr_{0,1}$), follows a binomial distribution of size k and probability α. 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 $E[Mr_{0,1}] = (1 - \alpha)m_r$). Moreover, both distributions will be assumed with finite variances through the paper.

If for every $n \geq 0$, ZR_n and Zr_n represent the total number of R-type and r-type mating units, respectively, at generation n, then the Y-linked bisexual process $\{(ZR_n, Zr_n)\}_{n \geq 0}$ is defined, recursively as follows:

- The number of R-type and r-type mating units in the initial generation is fixed:
$$(ZR_0, Zr_0) = (i, j), \quad i, j \in \mathbb{N}_0.$$

- For every $n \geq 0$, the offspring generated by each genotype in the $(n + 1)$st generation is specified by the formulas:

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

  where (FR_{n+1}, MR_{n+1}) and (Fr_{n+1}, Mr_{n+1}) represent the total number of females and males given, respectively, by all the R-type and r-type couples in generation n. Moreover,

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

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

- Finally, from the vector $(F_{n+1}, MR_{n+1}, Mr_{n+1})$ and considering the perfect fidelity mating and the preference of females for R-type males, (ZR_{n+1}, Zr_{n+1}) is determined as follows:

  - Since R-type males are chosen first as mates and perfect fidelity mating is assumed, the number of R-type mating units is

    $$ZR_{n+1} = \min\{F_{n+1}, MR_{n+1}\}.$$

  - The number of females which do not mate with R-type males is $\max\{0, F_{n+1} - MR_{n+1}\}$. These females mate with r-type males and the assumption of perfect fidelity implies that the number of r-type mating units is

    $$Zr_{n+1} = \min\{\max\{0, F_{n+1} - MR_{n+1}\}, Mr_{n+1}\}.$$

Furthermore, the total number of mating units in the nth generation is given by $Z_n = ZR_n + Zr_n$ and it follows that $Z_n = \min\{F_n, MR_n + Mr_n\}$.

The process $\{(ZR_n, Zr_n)\}_{n\geq 0}$ is a homogeneous Markov chain while $\{Z_n\}_{n\geq 0}$ does not have the Markov's property. However, if $p_k^R = p_k^r = p_k$ for all $k \geq 0$, then both types of mating units have the same reproduction behaviour and therefore the process $\{Z_n\}_{n\geq 0}$ is a bisexual branching process with perfect fidelity mating and reproduction law $\{p_k\}_{k\geq 0}$ (see Daley (1968), for details of this process).

From now on, we focus our attention on the extinction and/or explosion of genotypes R and r. To this end, we call the events $\{ZR_n \to 0\}$ and $\{Zr_n \to 0\}$ *extinction of the R and r genotypes*, respectively, and the events $\{ZR_n \to \infty\}$ and $\{Zr_n \to \infty\}$ *explosion of the R and r genotype*, respectively. Since the processes $\{ZR_n\}_{n\geq 0}$ and $\{Zr_n\}_{n\geq 0}$ are integer valued and $ZR_{n+1} = 0$ if $ZR_n = 0$ and $Zr_{n+1} = 0$ if $Zr_n = 0$, it follows that

$$\{ZR_n \to 0\} = \{ZR_n = 0 \ \text{eventually}\} = \bigcup_{n=1}^{\infty} \{ZR_n = 0\}$$

and

$$\{Zr_n \to 0\} = \{Zr_n = 0 \ \text{eventually}\} = \bigcup_{n=1}^{\infty} \{Zr_n = 0\}.$$

Although $\{ZR_n\}_{n\geq 0}$ and $\{Zr_n\}_{n\geq 0}$ are not Markov chains, González et al. (2006) proved that both genotypes have the dual *extinction-explosion* asymptotic behaviour, typical in many homogeneous branching processes (see Haccou et al.

(2005)): either the total number of mating units of a genotype goes to zero or has an unlimited growth, that is,

$$P(ZR_n \to 0) + P(ZR_n \to \infty) = 1 \quad \text{and} \quad P(Zr_n \to 0) + P(Zr_n \to \infty) = 1.$$

Therefore, if we denote by $A_{0,0} = \{ZR_n \to 0, Zr_n \to 0\}$ *the extinction of the population*, by $A_{\infty,0} = \{ZR_n \to \infty, Zr_n \to 0\}$ *the fixation of R genotype*, by $A_{0,\infty} = \{ZR_n \to 0, Zr_n \to \infty\}$ *the fixation of r genotype* and by $A_{\infty,\infty} = \{ZR_n \to \infty, Zr_n \to \infty\}$ *the explosion of both genotypes*, we have that

$$P(A_{0,0}) + P(A_{\infty,0}) + P(A_{0,\infty}) + P(A_{\infty,\infty}) = 1.$$

In conclusion, to study the extinction problem of the gene in the population, we focus our attention on the events $A_{0,0}$, $A_{\infty,0}$, $A_{0,\infty}$ and $A_{\infty,\infty}$.

3. The extinction problem

From now on, in order to simplify the notation, we denote $P(\cdot|(ZR_0, Zr_0) = (i, j))$ by $P_{(i,j)}(\cdot)$, with $i, j \geq 0$. Moreover, we assume that $i, j > 0$. When $i = 0$ or $j = 0$, then there is only one surviving genotype at the initial generation which evolves like a bisexual process with its associated reproduction law (see González et al. (2006)), and therefore the extinction problem is deduced from Daley (1968).

A necessary and sufficient condition for the population to become extinct almost surely is given in the following result:

Theorem 1. *Let $i, j > 0$, then $P_{(i,j)}(A_{0,0}) = 1$ 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. Moreover, as it is shown in the following result, $\min\{\alpha m_r, (1 - \alpha)m_r\} > 1$, is a necessary and sufficient condition for a positive probability of fixation of the r genotype, and consequently the extinction probability of the whole population is less than one.

Theorem 2. *Let $i, j > 0$, then $P_{(i,j)}(A_{0,\infty}) > 0$ if and only if* $\min\{\alpha m_r, (1 - \alpha)m_r\} > 1$.

Therefore the almost sure extinction of r genotype implies that $\min\{\alpha m_r, (1 - \alpha)m_r\} \leq 1$. Moreover, this happens independently on behaviour of R genotype. Indeed, if $(1 - \alpha)m_r \leq 1$, which intuitively means that the average number of

males produced by a r-couple is less than or equal to one, then the r genotype becomes extinct almost surely because there are not enough r-type males. Also, if $\max\{\alpha m_R, \alpha m_r\} \leq 1$, that is, the average number of females generated by a mating unit is less than or equal to one, then there are not enough females in the population and it becomes extinct almost surely. The situation $(1 - \alpha)m_r > 1$, $\alpha m_r \leq 1$ and $\alpha m_R > 1$ is special. Since $\alpha m_R < (1 - \alpha)m_R$ then, possibly, some R-type males mate with females produced by r-couples (see left graphic in Figure 1) and the remaining females are not enough to keep the r genotype alive. To show this scenario we have simulated 10 generations for such a process with $(ZR_0, Zr_0) = (1, 10000)$, $\alpha = 0.2$ and the reproduction laws following Poisson distributions with $m_R = 6$ and $m_r = 5$. In Figure 1, we illustrate a path from such a process where we can observe the pattern previously described. In conclusion, we establish the following result:

Theorem 3. *Let $i, j > 0$, then $P_{(i,j)}(Zr_n \to 0) = 1$ if and only if $\min\{\alpha m_r, (1 - \alpha)m_r\} \leq 1$.*
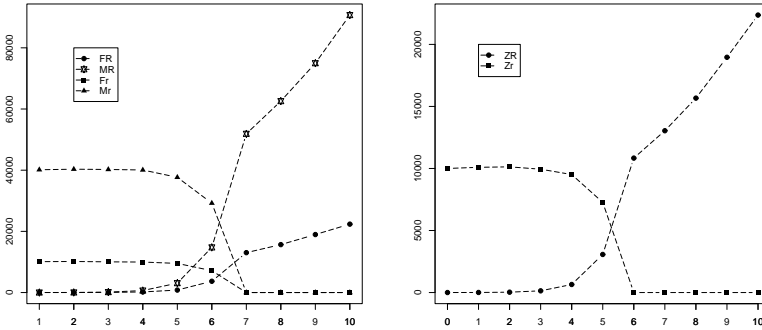

Figure 1: Path from a process where $(1 - \alpha)m_r > 1$, $\alpha m_r \leq 1$ and $\alpha m_R > 1$

On the other hand, the behaviour of R genotype does depend on the extinction or survival of r genotype, because of the preference. When the r genotype becomes extinct, we obtain the following result about fixation of R genotype, similar to Theorem 2:

Theorem 4. *Let $i, j > 0$, then $P_{(i,j)}(A_{\infty,0}) > 0$ if and only if $\min\{\alpha m_R, (1 - \alpha)m_R\} > 1$.*

From Theorem 3 and Theorem 4, we deduce that for the simultaneous survival of both genotypes with positive probability it is necessary that the average number of males generated by a mating unit of any genotype as well as the average number of females produced by a r-type couple are greater than one. When $\alpha > 0.5$, this condition is also sufficient because, provided that the number of mating units of both types is large enough, the number of females produced by each genotype is greater than the number of males generated by such a genotype. Therefore the number of mating units of both genotypes may grow indefinitely because the average number of males of both genotypes is greater than one and every male mates.

On the other hand, when $\alpha < 0.5$, R-type mating units do not produce enough females to mate all the R-type males and some of these mate with females produced by r-type mating units (see left graphic in Figure 1). Thus, the simultaneous survival of both genotypes depends on the relation between αm_r and $(1 - \alpha)m_R$, that is, the average number of females produced by an r-type couple and the average number of males produced by an R-type couple. Indeed, if $\alpha m_r < (1 - \alpha)m_R$, then, as long as the R genotype is alive, the number of R-type males is so big that, eventually, all the females choose these males and r-type males cannot mate, so the r genotype becomes extinct, and therefore the survival of both genotype has zero probability. Notice that r genotype can survive by itself, because $\min\{\alpha m_r, (1 - \alpha)m_r\} > 1$. Moreover, if $\alpha m_R > 1$, the R genotype can survive by itself, but not simultaneously with r genotype. This is an amazing fact, because although all parameters are greater than one, that is, $\min\{\alpha m_R, (1 - \alpha)m_R, \alpha m_r, (1 - \alpha)m_r\} > 1$, we obtain that $P_{(i,j)}(A_{\infty,\infty}) = 0$ and therefore $P_{(i,j)}(A_{0,0}) + P_{(i,j)}(A_{\infty,0}) + P_{(i,j)}(A_{0,\infty}) = 1$. To show this scenario we have simulated 20 generations of a Y-linked bisexual process with $(ZR_0, Zr_0) = (1,5)$, $\alpha = 0.3$ and reproduction laws following Poisson distributions with $m_R = 3.5$ and $m_r = 4$. In Figure 2 we illustrate three paths from such a process where we can observe the different patterns previously described.

Furthermore, when $\alpha < 0.5$ and $(1 - \alpha)m_R < \alpha m_r$, the number of females produced by r-type couples is much bigger than the number of males produced by R-type couples, so only a negligible amount of these females mates with R-type males. The rest of females mates with r-type males and they are enough for the survival of r genotype. Since $1 < (1 - \alpha)m_R$, then both genotypes survive simultaneously. In conclusion, we establish the following result concerning the event $A_{\infty,\infty}$:

Theorem 5. *Let $i, j > 0$,*

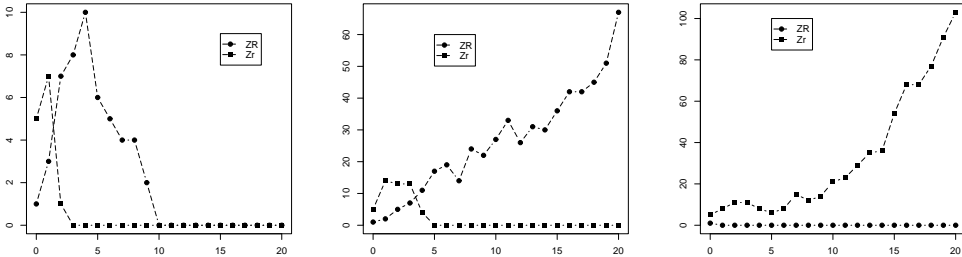*i) If $\min\{(1 - \alpha)m_R,\ (1 - \alpha)m_r,\ \alpha m_r\} \leq 1$, then $P_{(i,j)}(A_{\infty,\infty}) = 0$.*

Figure 2: Paths from a process where $1 < \alpha m_r < (1-\alpha)m_R$ and $1 < \alpha m_R$

ii) If $\alpha < 0.5$ and $1 < \alpha m_r < (1-\alpha)m_R$, then $P_{(i,j)}(A_{\infty,\infty}) = 0$.

iii) If $\alpha < 0.5$ and $1 < (1-\alpha)m_R < \alpha m_r$, then $P_{(i,j)}(A_{\infty,\infty}) > 0$.

iv) If $\alpha > 0.5$ and $\min\{(1-\alpha)m_R, (1-\alpha)m_r\} > 1$, then $P_{(i,j)}(A_{\infty,\infty}) > 0$.

The case $\alpha < 0.5$ and $1 < (1-\alpha)m_R = \alpha m_r$ is not included in the previous result. In order to conjecture the possible behaviour of the process under such conditions we have simulated five batches of 10000 processes with reproduction laws following Poisson distributions with means $m_R = 7/4$ and $m_r = 21/8$ and $\alpha = 0.4$. For these values of the parameters, $\alpha m_r = (1-\alpha)m_R = 1.05 > 1$. In all the simulated processes we set $ZR_0 = 10$ and $Zr_0 = 25$. The following table shows the number of processes in each batch that have not become extinct by generations 20, 40, 60, 80 and 100.

batch	1	2	3	4	5
generation 20	561	559	512	549	573
generation 40	67	79	65	72	87
generation 60	19	22	14	21	25
generation 80	8	10	5	9	12
generation 100	2	6	1	7	6

It is observed how the number of non-extinct processes decreases along generations with similar figures in the five batches. Therefore, from this simulation we conjecture for a Y-linked bisexual branching process that if $\alpha < 0.5$ and $\alpha m_r = (1-\alpha)m_R > 1$, then $P_{(i,j)}(A_{\infty,\infty}) = 0$.

Notice that the situation $\alpha = 0.5$ has not been considered in the previous theorem and still is an open problem. Finally, notice that, since the behaviour of

R genotype depends on r genotype, then it is not possible to provide a necessary and sufficient condition for the almost sure extinction of the R genotype.

REFERENCES

[1] A. Bernardo, B. A. Dobo, M. D. Vibranovski, A. G. Clark. Identification of five new genes on the Y chromosome of Drosophila melanogaster. *Proc. Natl. Acad. Sci. USA* **98**, No 23 (2001), 13225–13230.

[2] D. Charlesworth, B. Charlesworth, G. Marais. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95** (2005), 118–128.

[3] D. J. Daley. Extinction condition for bisexual Galton-Watson branching processes. *Z. Wahrscheinlichkeitsth* **9** (1968), 315–322.

[4] P. Haccou, P. Jagers, V. Vatutin. Branching Processes: Variation Growth and Extinction of Populations. Cambridge University, 2005.

[5] M. González, D. Hull, R. Martínez, M. Mota. Bisexual branching processes in a genetic context: The extinction problem for Y-linked genes. *Math. Biosci.* **202** (2006), 227–247.

[6] M. González, R. Martínez, M. Mota. Bisexual branching processes in a genetic context: Rates of growth for Y-linked genes. *Math. Biosci.* **215** (2008), 167–176.

[7] J. F. Hughes, H. Skaletsky, T. Pyntikova, P. J. Minx, T. Graves, S. Rozen, R. K. Wilson, D. C. Page. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **437** (2005), 100–103.

[8] C. Krausz, G. Forti, K. McElreavey. The Y chromosome and male fertility and infertility. *Int. J. Androl.* **26** (2003), 70–75.

[9] C. Krausz, L. Quintana-Murci, G. Forti. Y chromosome polymorphisms in medicine. *Ann Med.* **36**, No 8 (2004), 573–583.

[10] L. Quintana-Murci, M. Fellous. The human Y chromosome: the biological role of a "functional wasteland". *J. Biomed. Biotechnol.* **1** (2001), 18–24.

[11] D. YAMADA, Y. KOYAMA, M. KOMATSUBARA, M. URABE, M. MORI, Y. HASHIMOTO, R. NII, M. KOBAYASHI, A. NAKAMOTO, J. OGIHARA, J. KATO, S. MIZUNO. Comprehensive search for chicken W chromosome-linked genes expressed in early female embryos from the female-minus-male subtracted cDNA macroarray. *Chromosome Research* **12** (2004), 741–754.

Miguel González
Department of Mathematics
University of Extremadura
06071 Badajoz, Spain
e-mail: `mvelasco@unex.es`

Cristina Gutiérrez
Department of Mathematics
University of Extremadura
06071 Badajoz, Spain
e-mail: `cgutierrez@unex.es`

Rodrigo Martínez
Department of Mathematics
University of Extremadura
06071 Badajoz, Spain
e-mail: `rmartinez@unex.es`

Manuel Mota
Department of Mathematics
University of Extremadura
06071 Badajoz, Spain
e-mail: `mota@unex.es`