A METHODOLOGY FOR THE STOCHASTIC MODELLING AND SIMULATION OF SYMPATRIC SPECIATION BY SEXUAL SELECTION

ROBERTO BARBUTI, ANDREA MAGGIOLO-SCHETTINI, PAOLO MILAZZO
Dipartimento di Informatica, Università di Pisa
Largo Pontecorvo 3, 56127 Pisa, Italy
{barbuti,maggiolo,milazzo}@di.unipi.it

ANGELO TROINA
Dipartimento di Informatica, Università di Torino
Corso Svizzera 185, 10149 Torino, Italy
troina@di.unito.it

Received (Day Month Year)
Revised (Day Month Year)

In the evolution literature, sympatric speciation is the origin of two, or more, species from a single local population. Many models have been developed to study the role of ecological competition and sexual selection in sympatric speciation.

In this paper we propose a methodology for systematically deriving efficient computational models to study speciation in populations evolving with overlapping generations. As a particular case, we consider sympatric speciation by sexual selection and we follow an individual based approach: a population is represented as a set of individuals that can mate and survive according to given probabilities.

We use our methodology to construct four different models for sympatric speciation, based on male traits and female preferences. These models differ in the genotypical representation of the individuals. Results of simulations in the different models are shown and discussed.

The study of the models show that sympatric speciation by sexual selection is unlikely, also with a favorable distribution of genotypes in the initial population.

Keywords: Evolution; Sympatric speciation; Sexual selection; Stochastic models; Simulation.

1. Introduction

In the last few years many research efforts have been devoted to the analysis of speciation by means of theoretical models [1,2,3,4]. Mathematical analysis of these models makes it possible to reason about empirical evidence and testing the plausibility of conjectures which try to explain such facts. An example of a question which has been a subject of mathematical analysis is the plausibility of reinforcement of premating isolation as a cause of speciation [5,6,7]. This has been suggested by empirical knowledge, for example on some species of *Drosophila* [8].

Particular relevance has recently been given to the study of the sympatric speciation process [9,10,11,12,13]. Sympatric speciation is the origin of two, or more, species occurring in a population of individuals living in the same geographic site. Empirical studies have suggested that speciation could have been sympatric, for example in cichlid fishes of Lake Malawi [14,15], crater lakes in Cameroon [16,17], and Nicaragua lakes [18,19]. Darwin believed that new species may emerge if intraspecific competition for different resources leads to disruptive selection which in turn favors diverging phenotypes. This causes both the elimination of intermediate phenotypes and the reproductive isolation of the diverging ones. Some examples of theoretical models which consider this kind of ecological competition are in [20,12,21,22,13,23,24,25,26,27].

Other studies show that sexual selection, either in conjunction with ecological selection or not, could be another possible cause of sympatric speciation [28,29,12,30,31,32,33,34,21,35,36,37,13]. Although some authors are in favor of sympatric speciation by sexual selection, the process remains controversial. Other authors consider it to be difficult that speciation driven only by mate preference and secondary sexual traits occurs, especially in the absence of a contemporary ecological selection.

Theoretical models address the evolution of female preference for male ornaments. They show how female preferences and male traits become genetically linked because of the non–random mate choice [28,38,37,33,32].

Due to the positive feedback of this linkage disequilibrium, male trait and female preference evolve in a runaway process if it is only weakly opposed by ecological forces. Runaway processes caused by sexual selection can provide both pre–zygotic isolation, and differently from evolution caused by natural selection their direction is arbitrary. Thus, sympatric speciation by sexual selection may occur if, in the same population, two different runaway processes proceed in different directions.

Some of the models proposed for the sympatric speciation process are computational and individual based [38,21,32,33,39]. That is, the population is represented as a set of individuals, and the evolution of the population is obtained by considering the mating and the survival of individuals according to given probabilities.

We believe that computational individual based models are better suited than deterministic non-individual based ones for describing the inherent random character of natural phenomena and can be used in a variety of cases. This is true especially when the size of the population is not large enough to ignore stochasticity. Individual based models can embody stochasticity in an easy and natural way.

Recently, many studies in Computer Science have been devoted to the definition of stochastic formalisms for the description and simulation of biological systems [40,41,42,43,44]. Such formalisms make it possible to verify some interesting properties of the modeled systems by means of automatic tools, e.g [45,46,47]

In [48,49] we have proposed an individual based stochastic modelling approach for simulating biochemical reactions. In this context a model consists of a set of replacement rules whose application depends on a probability. Each rule describes

a reaction, that is, how some reactants are replaced by products. The probability of the application of a rule is derived from the production rate suitably normalized.

In this paper we apply the same approach to derive models for studying speciation by sexual selection. Specific models can be derived by instantiating genotypes of individuals and production and cancellation rates relative to such genotypes. Therefore, we propose rather than a model, a systematic methodology for the construction of computational models for sympatric speciation. The purpose of the paper is twofold: to provide biologists with a methodology for specifying individual based speciation models, and to show how, by means of a simulator we have implemented, these models can be used to validate evolutionary hypotheses.

Our methodology is based on Gillespie's algorithm for the simulation of chemical reactions [50]. The algorithm, unlike most procedures for the numerical solution of differential equations describing chemical kinetics, never approximates infinitesimal time increments by finite time steps. Consequently, the time is increased exactly to the time in which the next event occurs. Thus, the resulting simulation algorithm is characterized by a higher efficiency with respect to other simulation methods.

The ease of specifying new models by using our methodology and the efficiency of our simulator make it possible to test hypotheses on a variety of different models. The models we construct are usually based on the same evolutionary concepts (for example speciation by mate choice based on male traits) and various representations of genotypes (haploid, diploid, with different number of loci, etc...).

We illustrate our methodology by deriving four models for sympatric speciation by sexual selection. These models differ in the genotype of the population. We consider a haploid model (the genotype is represented by a set of single chromosomes) in which male trait and female preference are coded each in one locus, a haploid model in which male trait is coded in one only locus and female preference is coded in two loci, and the two corresponding diploid models (the genotype is represented by a set of pairs of chromosomes). Haploid models are used, in computational and mathematical models, for their simplicity. Actually, in these model, gametes are represented in place of individuals and, usually, the population evolves in an approximated way because any pair of gametes generates only one gamete as an offspring [2].

We consider populations evolving with overlapping generations, namely individuals are born and die independently, such that at any time individuals of different generations may exist. The capacity of offspring generation and the viability selection are viewed as production rates. That is, the former is viewed as rate of production of new individuals and the latter as rate of cancellation of individuals. In the considered population rates may be different for different individuals. For instance, offspring production rates will depend on mating preferences. We have made these assumptions inspired by the population dynamics of polygynous cichlid fishes. These have been considered due to the large number of sister species which have evolved also by speciation driven by sexual selection.

The results of the simulation of the four models show that sympatric speciation

driven by sexual selection is possible only with strong initial assumptions. This result does not agree with the results of analogous individual based either deterministic or stochastic models [32,33] in which speciation seems to occur with weaker assumptions. Rather, our four models show that in the presence of a female preference, which is not sufficiently high, the population evolves towards a shift, that is towards a population with homogeneous ornamented males. Our results agree with those in [28].

Due to the generality of our methodology, we expect that other, more complex cases, such as the ones in which the mating preference depends on multiple traits, may be easily dealt with. Moreover, we expect that our methodology can be easily applied to describe speciation in different situations and due to different causes.

2. The Framework for Model Construction

In this section we present the formal definitions of the genetic model and the algorithmic technique we employ for the study of sympatric speciation by sexual selection.

We consider populations consisting of males and females represented by their genotypes. In particular, males exhibit external traits, each trait corresponding to one or more genotypes. With each trait a disadvantage is associated as regards survival of the individual (*trait cost*). On the other hand, females exhibit sexual preferences for male traits. These preferences are determined by female genotypes. As for males, female genotypes may influence survival.

We assume that populations evolve with overlapping generations. This implies that in our models we will not have, as for instance in [32], a cyclic occurrence of a mating process with offspring generation followed by a viability selection process in which a portion of the population die. In our models we will have a free alternation of birth and death events in which an offspring is born and an individual of the population dies, respectively. Examples of populations with overlapping generations are the ones of haplochromine cichlids fishes in african lakes [51,52,53,54].

With this assumption, births and deaths are almost independent events. Hence, the time between two occurrences of such events can be naturally modeled as an exponentially distributed random variable, and the whole process can be modeled as a Continuous Time Markov Chain (CTMC) [55]. Actually, birth and death events are not completely independent because, as an example, the birth of an offspring with a certain genotype increases the probability in the future of having new offspring with a similar genotype. But this simply means that the parameters of the exponential distributions in the Markov chains, namely the rates of the birth and death events, will not be constant, but functions of the state of the population. In other words, the rates of birth and death events of an individual with a certain genotype will depend on the number of males and females in the population with related genotypes.

We base the construction of models on the instantiation of general formulas describing the birth rate of offspring and the death rate of individuals with specific

genotypes. These rates will be used as parameters of the exponential distributions of a CTMC modeling the evolution of a population. The time unit assumed for these rates is the mean duration of a generation for the considered species.

2.1. *Representation of genotypes*

We describe how to represent genotypes of both haploid and diploid populations by considering an arbitrary number of loci. We model genotypes of a haploid population as tuples. In a tuple representing a genotype, positions correspond to loci and the values of such positions are the alleles composing the genotype. Given a tuple of n loci, with the set A_i for $i \in [1, n]$ we represent the allele domain for the locus i, namely locus i may assume any value in A_i. With $A = A_1 \times \ldots \times A_n$ we represent the domain of all the possible genotypes of an individual.

An individual of a population may therefore be defined as $x_{(a_1,\ldots,a_n)}$, where $x \in \{m, f\}$ represents the sex of the individual and $(a_1, \ldots, a_n)$, for $a_i \in A_i$, represents its genotype. With $\mathcal{I}_A$ we denote the set of all kinds of individual with a genotype in A. To each individual of a given sex and genotype a phenotype is associated by means of a function $pheno : \mathcal{I}_A \to Ph$, where Ph is the set of possible phenotypes.

Example 1 *Consider a simple three–locus haplotype, i.e. haploid genotype, where loci* 1 *and* 2 *determine the male secondary sexual character, with* $A_1 = A_2 = \{0, 1\}$*, and locus* 3 *determines the female preference, with* $A_3 = \{0, 1, 2\}$*. Consequently, in this case* $\mathcal{I}_A$*, where* $A = A_1 \times A_2 \times A_3$*, consists of 24 kinds of individuals and is defined as follows:*

$$\mathcal{I}_A = \left\{ x_{(a_1,a_2,a_3)} \mid x \in \{m, f\},\, a_1 \in \{0, 1\},\, a_2 \in \{0, 1\},\, a_3 \in \{0, 1, 2\} \right\} .$$

Assume that, as in [33]*, loci* 1 *and* 2 *give rise to three different phenotypes, denoted with* t_0*,* t_1 *and* t_2 *and corresponding to three different male secondary sexual characters (traits 0, 1 and 2, respectively). The value of the trait is obtained by summing up the values of the alleles in the two loci. Assume also that locus* 3 *gives rise to three different phenotypes, denoted with* p_0*,* p_1 *and* p_2*, corresponding to three different female preferences (for traits 0, 1 and 2, respectively). The set of phenotypes in this case is hence*

$$Ph = \{t_0, t_1, t_2, p_0, p_1, p_2\}$$

and the phenotype function is

$$pheno(x_{a_1,a_2,a_3}) = \begin{cases} ty & \text{if } x = m \text{ and } y = a_1 + a_2 \\ py & \text{if } x = f \text{ and } y = a_3 \,. \end{cases}$$

According to [33]*, females with phenotypes* p_0 *and* p_2 *prefer males with a phenotype* t_0 *and* t_2*, respectively, while females with phenotype* p_1 *have no preference. With our representation,* $pheno(m_{(0,0,1)}) = t_0$ *and* $pheno(m_{(1,1,0)}) = t_2$

are different phenotypes of males with different genotypes, while $pheno(m_{(0,1,1)}) = pheno\ (m_{(1,0,0)}) = t_1$ *is the same phenotype of males with different genotypes. Moreover,* $pheno(f_{(0,0,0)}) = p_0$ *and* $pheno(f_{(1,0,2)}) = p_2$ *are examples of phenotypes of females which prefer males having phenotypes* t_0 *and* t_2*, respectively, and* $pheno(f_{(0,1,1)}) = p_1$ *is an example of phenotype of a female with no preference.*

Genotypes of a diploid species can be modelled by using pairs of tuples. An individual of a diploid population may be denoted with $x_{\langle(a_1,\ldots,a_n),(a'_1,\ldots,a'_n)\rangle}$, where $x \in \{m, f\}$ represents the sex of the individual, $\langle(a_1, \ldots, a_n), (a'_1, \ldots, a'_n)\rangle$, where $a_i, a'_i \in A_i$, represents its genotype, and $A = (A_1 \times \ldots \times A_n) \times (A_1 \times \ldots \times A_n)$ represents the domain of all the possible genotypes.

Example 2 *Consider a diploid population with two–locus genotypes. Let locus* 1 *determine the male secondary sexual character with* $A_1 = \{0, 1\}$*, and let locus* 2 *determine the female preference with* $A_2 = \{0, 1\}$*. Moreover, let the set of phenotypes be* $Ph = \{t_0, t_1, p_0, p_1\}$*, where* t_0 *and* t_1 *denote the absence and the presence of the secondary sexual character in males, respectively, and* p_0 *and* p_1 *denote the absence and the presence of preference in females, respectively, for males with the secondary sexual character. Assume that male character and female preference are equal to* t_1 *and* p_1*, respectively, only if at least one allele* 1 *is present at the corresponding locus (i.e. 1 is a dominant allele). In this case, the phenotype function is defined as follows:*

$$pheno(x_{\langle(a_1,a_2),(a'_1,a'_2)\rangle}) = \begin{cases} ty & x = m \text{ and } y = max(a_1, a'_1) \\ py & x = f \text{ and } y = max(a_2, a'_2) \end{cases}$$

where $max(a, b)$ *gives the maximum between* a *and* b*.*

Individual $m_{\langle(0,0),(1,0)\rangle}$ *is an example of a male with phenotype* t_1*, namely in which the secondary sexual character is present, while* $m_{\langle(0,1),(0,1)\rangle}$ *is an example of a male with phenotype* t_0*, namely in which the character is absent. Similarly,* $f_{\langle(0,1),(0,1)\rangle}$ *is an example of a female with phenotype* p_1*, namely with a preference for males with the secondary sexual character, while* $f_{\langle(1,0),(0,0)\rangle}$ *is an example of a female with phenotype* p_0*, namely without preference.*

A population of individuals in $\mathcal{I}_A$ is a multiset $P = \{(I, k) \mid I \in \mathcal{I}_A, k \in \mathbb{N}\}$. Given a set of individuals $I_A \subseteq \mathcal{I}_A$ and a population P, we define the frequency of individuals in I_A within the population P as:

$$N_P(I_A) = \sum_{I \in I_A . (I,k) \in P} k$$

Viability selection and the cost of female mate choice, reduce the survival of ornamented males and choosy females. We consider two cost functions $cost_m : A \to [0, 1]$ and $cost_f : A \to [0, 1]$, from the set of all genotypes A to the interval $[0, 1]$,

that give a cost for the survival of males and females, respectively: a cost equal to 0 means no cost.

Moreover, we define a matrix of female preferences as a function $pref : A \times A \to \mathbb{R}^{\geq 0}$, which takes as input two genotypes $a, a' \in A$ and gives the preference parameter of a female with genotype a for males with genotype a'.

Note that, for the sake of simplicity, we define the cost and the preference as functions of male and female genotypes. Of course such a function could be defined on phenotypes, as phenotypes are function of genotypes.

Example 3 *Consider again the haplotype introduced in Example 1. We may assume that females with phenotype p_0, namely those with $a_3 = 0$, favor the ornamented males with phenotype t_0, namely those with $a_1 + a_2 = 0$, and disfavor males with phenotype t_2, namely those with $a_1 + a_2 = 2$. Symmetrically, we may assume that females with phenotype p_2, namely those with $a_3 = 2$, favor the ornamented males with phenotype t_2, namely those with $a_1 + a_2 = 2$, and disfavor males with phenotype t_0, namely those with $a_1 + a_2 = 0$. Females with phenotype p_1 have no preference. Therefore, we may have $pref((\_,\_,0),(a'_1,a'_2,\_)) = pref((\_,\_,1),(a'_1,a'_2,\_)) = pref((\_,\_,2),(a'_1,a'_2,\_)) = 1$ when $a'_1 + a'_2 = 1$ (males with phenotype 1 are neither favoured nor disfavoured), $pref((\_,\_,1),a) = 1$ for any a (females with $a_3 = 1$ have no preference), $pref((\_,\_,0),(0,0,\_)) = pref((\_,\_,2),(1,1,\_)) > 1$ and $pref((\_,\_,0),(1,1,\_)) = pref((\_,\_,2),(0,0,\_)) < 1$. Note that $\_$ means any allele.*

Finally, we incorporate a recombination probability in the model through the function $rec : A \times A \times A \to [0,1]$; namely $rec(a, a', a'')$, where $a, a', a'' \in A$ returns the probability that parents with genotypes a and a', respectively, give rise to an individual with genotype a''. We require that $\forall a, a' \in A \ \sum_{a'' \in A} rec(a, a', a'') = 1$.

In haploid populations, the recombination probability is used to determine from which parent an allele is inherited. In diploid populations, instead, this probability is used to determine from which of the two copies of a gene of a parent the allele is inherited.

Example 4 *Let us consider a haploid population with $A_1 = \ldots = A_n = \{0,1\}$, and probability $\frac{1}{2}$ of inheriting an allele from the father and the mother, respectively. We may define the rec function as follows:*

$$rec((a_1,\ldots,a_n),(a'_1\ldots,a'_n),(a''_1\ldots,a''_n)) = \prod_{i=1}^{n}(\frac{1}{2}(1-|a_i-a''_i|)+\frac{1}{2}(1-|a'_i-a''_i|))\,.$$

For example, in the case of $n = 2$, such a function gives:
$rec((0,0),(0,0),(0,0)) = 1$,
$rec((1,1),(0,1),(0,1)) = rec((1,1)(0,1),(1,1)) = \frac{1}{2}$,
and $rec((0,1),(1,0),(0,0)) = \frac{1}{4}$.

2.2. *Definition of birth and death rates*

The general formulas for offspring birth, in a population P, are the following:

$$\begin{aligned} birthrate_P(f_a, m_{a'}, m_{a''}) &= N_P(f_a) \cdot mateprob_P(f_a, m_{a'}) \cdot \frac{1}{2} \cdot rec(a, a', a'') \cdot \gamma \\ birthrate_P(f_a, m_{a'}, f_{a''}) &= N_P(f_a) \cdot mateprob_P(f_a, m_{a'}) \cdot \frac{1}{2} \cdot rec(a, a', a'') \cdot \gamma \end{aligned}$$

where

$$mateprob_P(f_a, m_{a'}) = \frac{N_P(m_{a'}) \cdot pref(a, a')}{\sum_{\tilde{a} \in A} N_P(m_{\tilde{a}}) \cdot pref(a, \tilde{a})} \ .$$

These equations define the rate of production of males (resp. females) of genotype a'', denoted $m_{a''}$ (resp. $f_{a''}$), from females and males of genotype a and a', respectively (in brief, called from now on a–females and a'–males), in a population P.

A birth rate is obtained by multiplying the number of a–females in the population P ($N_P(f_a)$) by the probability that an a–female mates an a'–male (given by the $mateprob_P$ formula), by the probability that the genetic recombination produces the genotype a'', and by γ, the *birth environmental constant*, which is the average number of offspring produced by a female in one generation, divided by 2 to take into account the possibility of producing a''–females instead of a''–males. The value of the $mateprob_P$ formula (namely the probability that an a–female chooses a a'–male) is proportional to the number of males $m_{a'}$ in the population and to the preference of females f_a for males $m_{a'}$.

The general formulas for death of individuals are the following:

$$\begin{aligned} deathrate_P(m_a) &= N_P(f) \cdot \delta \cdot deathprob_P(m_a) \\ deathrate_P(f_a) &= N_P(f) \cdot \delta \cdot deathprob_P(f_a) \end{aligned}$$

where

$$deathprob_P(x_a) = \frac{N_P(x_a) \cdot (1 - cost_x(a))^{-1}}{\sum_{\tilde{a} \in A} N_P(m_{\tilde{a}}) \cdot (1 - cost_m(\tilde{a}))^{-1} + \sum_{\tilde{a} \in A} N_P(f_{\tilde{a}}) \cdot (1 - cost_f(\tilde{a}))^{-1}} \ .$$

These equations define the death rate for a–males (resp a–females), in the population P. In the equation, $N_P(f) \cdot \delta$ is the reduction factor of the considered population, obtained by multiplying the total number of females $N_P(f)$ by a limiting factor δ related to the environment, called *death environmental constant*. The factor $N_P(f) \cdot \delta$, gives the number of individuals which die with respect to the produced offspring (which depends on the number of females). If we have $\delta = \gamma$, the number of individuals in the population remains constant with high probability. If $\delta < \gamma$, births prevail and the number of individuals increases geometrically (as in the model in [30]). When $\delta > \gamma$, the population eventually becomes extinct. The probability that the individual that dies in the population P is x_a is given by $deathprob_P(x_a)$. This value is proportional to the number of individuals x_a in the population, and inversely proportional to one minus the cost associated with the genotype of the individual.

As explained in [56], this kind of simple growth model in which both birth and death rates are directly proportional to the number of individuals in the population (in this case to the number of females) could be unrealistic because it does not consider the amount of available sustaining resources in the environment. In many cases, a more realistic growth model is the *logistic model* (see [56]), in which the carrying capacity of the environment is explicitly expressed as a constant factor K. A way for adopting the logistic growth model in our rate formulas is replacing the death environmental constant δ with $\gamma \cdot \frac{N_P(I)}{K}$, where $N_P(I)$ is the number of individuals in the population and the constant K is the carrying capacity of the environment. However, since we are not interested in studying how the number of individuals in the population changes over time, but only which species prevail, in this paper we choose to adopt the simpler directly proportional model instead of the logistic one. This approach is used also in [38].

A population is represented by a multiset of individuals. We represent the evolution of a population by a set of evolutionary rules $\mathcal{R} = \{R_1, R_2, \ldots, R_L\}$. An evolutionary rule can be either a *birth rule* or a *death rule*. A birth rule is a triple $(\{f_a, m_{a'}\}, \{f_a, m_{a'}, x_{a''}\}, birthrate_P(f_a, m_{a'}, x_{a''}))$ where x can be m or f, meaning that an a–female and an a'–male can produce, with rate $birthrate(f_a, m_{a'}, x_{a''})$, offspring with genotype a''. Given a population P, a birth rule $(\{f_a, m_{a'}\}, \{f_a, m_{a'}, f_{a''}\}, birthrate_P(f_a, m_{a'}, f_{a''}))$ is applicable if $\{f_a, m_{a'}\} \subseteq P$, and produces $(P \setminus \{f_a, m_{a'}\}) \cup \{f_a, m_{a'}, f_{a''}\}$, that is the offspring is added to the population. The rate gives the expected number of applications of the rule in the time unit. Analogously, a death rule is a triple $(\{x_a\}, \emptyset, deathrate_P(x_a))$ meaning that an individual x_a is cancelled from the population at a rate $deathrate_P(x_a)$. In general, an evolutionary rule $(P_\mu, P'_\mu, r^\mu_P) \in \mathcal{R}$ can be applied to a population P if $P_\mu \subseteq P$, and produces $(P \setminus P_\mu) \cup P'_\mu$ with rate r^μ_P. In the set $\mathcal{R}$ of evolutionary rules there are as many birth rules as the number of possible combinations of mother's and father's genotypes and offspring's genotype and sex. Death rules are as many as the possible kinds of individual, given by sex and genotype. Remark that evolutionary rules consider all possible genotypes which can be represented by using loci and alleles, not only the genotypes present in the population.

Example 5 *Consider a simple haploid population with a single locus and with the allele domain* $A_1 = \{0, 1\}$*. Individuals may have only two possible genotypes* 0 *and* 1*. The set of evolutionary rules for this population contains twenty rules, namely sixteen birth rules (eight for a male offspring obtained by considering all the possible combinations of* 0 *and* 1 *as genotypes of the mother, the father and the offspring, and eight for a female offspring obtained in the same way) and four death rules (one for* 0*–males, one for* 1*–males, one for* 0*–females and one for* 1*–females).*

Note that a set of evolutionary rules characterizes a CTMC. In fact, a state of the CTMC is a multiset of individuals representing (a state of) the modeled population. A transition from a state of the CTMC to another one, describing the occurrence of

an event, corresponds to the application of one of the evolutionary rules. The state of the CTMC reached after the occurrence of the event is the multiset of individuals obtained after the application of the evolutionary rule and the rate of the event in the CTMC is the rate of the evolutionary rule.

The CTMC represents a system whose most complete description is a *master equation*. This is a set of first order differential equations describing the probability of a system to be in a state at a given time. In general, it is very difficult to solve either analytically or numerically the master equation. An alternative to solving the master equation for the probability of obtaining a given state is to follow a standard Monte Carlo simulation procedure. Given an initial state of the system, such a simulation procedure consists in randomly choosing the time at which the next event will occur by following an exponential probability distribution function, and then applying one of the evolutionary rules with a probability that is proportional to its rate.

Assume a set of evolutionary rules $\mathcal{R} = \{R_1, \ldots, R_L\}$, where $R_\mu = (P_\mu, P'_\mu, r^\mu_P)$ for $\mu \in [1, L]$. A state of the simulation is a pair $(P, time)$, where P is a population and *time* represents the current time. Assuming an initial population P_0, the initial state of the simulation is $(P_0, 0)$.

The *probabilistic evolutionary algorithm* consists in the iteration of the following two steps:

(1) An evolutionary rule R_μ is randomly chosen with probability $\frac{r^\mu_P}{\sum_{\nu=1}^{L} r^\nu_P}$.
(2) The increment of time τ of this iteration is randomly chosen with an exponential distribution with parameter $\sum_{\nu=1}^{L} r^\nu_P$.

At each iteration the current time *time* has to be incremented by τ and the evolutionary rule R_μ has to be applied, namely the population P has to be updated by subtracting the population P_μ and adding the population P'_μ. Hence, the state of the simulation becomes $((P \setminus P_\mu) \cup P'_\mu, time + \tau)$, and the simulation continues its iteration until *time* reaches the expected amount of simulated time.

Note that this simulation procedure follows the biological process of evolution with overlapping generations. In fact, birth and death rules will be applied with a free (random) alternation as birth and death events occur in the population. The frequency of application of the rules will depend on an exponentially distributed random variable, that is the usual and natural way of modeling independent events such as births and deaths. Finally, the probability of application of an evolutionary rule will be proportional to the rate of the rule, namely to the *birthrate* and *deathrate* functions we introduced above, that will depend, as we shall see, on the characteristics of the involved individuals such as their mating preference, the cost of their phenotype, etc. . . .

The same simulation procedure has been applied to other kinds of biological systems in which different events may occur with a free alternation and with a notion of frequency. Relevant examples of such biological systems are epidemic

systems [57] and cellular pathways [58]. In particular, in the context of the simulation of cellular pathways (more generally, of chemical reactions) a simulation procedure similar to the one we adopt is known as Gillespie's stochastic simulation algorithm [50]. Such an algorithm is now a standard alternative to differential equations for the analysis of system dynamics.

2.3. *Analysis of the simulation framework*

The proposed simulation framework stems from Gillespie's framework for the simulation of chemical reactions [50]. The chemical analogous of our birth and death rules are reactions with two and one reactants, respectively. The rates of these reactions are computed (in Gillespie's approach) by multiplying a kinetic constant by the number of possible reactants combinations. In the case of our birth and death rules the use of a constant would be too simplistic as we have to take mating preferences, fitness, etc. into account. As a consequence we define the rate of each rule as expressed by $birthrate_P$ and $deathrate_P$.

Gillespie's approach computes the quantity of time that has to elapse between two occurrences of chemical reactions as an exponentially distributed random variable with the sum of all reaction rates as parameter. Analogously, in our case we shall define the sum of all birth rates (one for each pair of genotypes) and the sum of all death rates (one for each genotype and gender) and we use the sum of these two sums as the parameter of an exponential distribution.

Let us consider a simulation state $(P, time)$. Following Gillespie's approach we assume that the probability that an event (either birth of death) will happen at time $time + \tau$ follows an exponential distribution such that

$$Prob_P(\tau < t) = 1 - exp(-(Birthrate_P + Deathrate_P)t)$$

where

$$\begin{aligned} Birthrate_P &= \sum_{a,a',a'' \in A} (birthrate_P(f_a, m_{a'}, m_{a''}) + birthrate_P(f_a, m_{a'}, f_{a''})) \\ Deathrate_P &= \sum_{a \in A} (deathrate_P(f_a) + deathrate_P(f_a)) \,. \end{aligned}$$

This probability is used by the simulation algorithm to exactly compute, given $(P, time)$, the time $time+\tau$ at which the next event occurs. Once this time has been computed, Gillespie's algorithm chooses the occurring event. In our model, such an event can be either a birth event or a death event. The probability that the event is a birth is given by

$$Prob_P(birth) = \frac{Birthrate_P}{Birthrate_P + Deathrate_P} \,.$$

Similarly, the probability that the event is a death is given by

$$Prob_P(death) = \frac{Deathrate_P}{Birthrate_P + Deathrate_P} \,.$$

Note that, when the birth environmental constant γ and the death environmental constant δ are equal, it holds $Birthrate_P = Deathrate_P$, and hence $Prob_P(birth) = Prob_P(death) = \frac{1}{2}$.

The probability of selecting a given birth rule, $(\{f_a, m_{a'}\}, \{f_a, m_{a'}, f_{a''}\}, birthrate_P(f_a, m_{a'}, f_{a''}))$, which corresponds to the birth of a female with genotype a'' (a recombination of parents' genotypes a and a'), is given by

$$\frac{birthrate_P(f_a, m_{a'}, f_{a''})}{Birthrate_P + Deathrate_P}\,.$$

Since $birthrate_P(f_a, m_{a'}, f_{a''}) = N_P(f_a) \cdot \frac{N_P(m_{a'}) \cdot pref(a,a')}{\sum_{\tilde{a} \in A} N_P(m_{\tilde{a}}) \cdot pref(a,\tilde{a})} \cdot \frac{1}{2} \cdot rec(a, a', a'') \cdot \gamma$, the probability of a mate between f_a and $m_{a'}$ depends both on the density of f_a and $m_{a'}$ in the population, and on the intensity of the preference of a–females for a'–males, $pref(a, a')$. The probability of producing offspring of genotype a'' further depends on the recombination function on parents' genotypes (in the following we will consider free recombination).

The birth environmental constant γ does not influence the choice of a particular birth rule (because this factor is present in each birth rule). Constant γ influences the computation of the time between events. Greater values of γ correspond to smaller time intervals between rule applications. The result is that the number of offspring produced in a time unit is greater.

A similar analysis can be performed for death rules. The probability that a particular death rule $(\{x_a\}, \varnothing, deathrate_P(x_a))$ for the death of an individual of genotype a is selected, is given by

$$\frac{deathrate_P(x_a)}{Birthrate_P + Deathrate_P}\,.$$

Since
$deathrate_P(x_a) = N_P(f) \cdot \delta \cdot \frac{N_P(x_a) \cdot (1 - cost_x(a))^{-1}}{\sum_{\tilde{a} \in A} N_P(m_{\tilde{a}}) \cdot (1 - cost_m(\tilde{a}))^{-1} + \sum_{\tilde{a} \in A} N_P(f_{\tilde{a}}) \cdot (1 - cost_f(\tilde{a}))^{-1}}$,
the probability of a death of an individual of genotype a depends both on its frequency in the population and on the cost of its phenotype, $cost_x$. In our simulations we will consider different costs for males and females.

The biological interpretation of our computational model is the following. The model assumes that males and females in the population move freely and encounter each other. Some encounters result in matings. The frequency of matings depends on the composition of the population as described by the above given $Prob_P$. Moreover, the genotypes of the individuals involved in these matings are probabilistically determined by the population composition and mating parameters. Our model assumes that, on average, each female mates successfully once if there are available males. This is an assumption of several models, for instance those adopting a "best of n" mate rule. The only cases in which a–females cannot mate are either when there are no males or when the male population is composed only of a'–males and $pref(a, a') = 0$. The average number of offspring produced in a mating season by

the a–females with a' males is

$$\begin{aligned} birthrate_P(f_a, m_{a'}) &= \sum_{a'' \in A} (birthrate_P(f_a, m_{a'}, m_{a''}) + birthrate_P(f_a, m_{a'}, f_{a''})) \\ &= N_P(f_a) \cdot mateprob_P(f_a, m_{a'}) \cdot \gamma \,. \end{aligned}$$

This number depends both on the number of a'–males and on the preference of a–females for a'–males. Hence, an a–female has a probability to mate with an a'–male which increases when these two values increase. When the number of a'–males increases, the a–female has a greater chance of encountering one of them for mating. Moreover, the probability of mating increases together with the preference of the female for the phenotype of the male. Actually, the value of $birthrate_P(f_a, m_{a'})$ is not stable during a mating season, due to births and deaths that change dynamically the composition of the population. However, the above formula approximately shows how the population evolves.

It is important to remark that this method for determining the mates is adopted in other models. In [32] and [33] this method is used in a stochastic and a in deterministic model, respectively.

Analogously, by considering a stable population we have that the average number of deaths of a–individuals in a time unit is given by

$$deathrate_P(x_a) = deathrate_P(f_a) + deathrate_P(m_a) \,.$$

The value of $deathrate(f_a)$ depends both on the number of a–females and a–males, and on the cost of their phenotypes. In the following we will assign the same cost to females, independently from the genotype. As for males, we will assign the ornamented males a cost greater than the one assigned to the non–ornamented ones.

Remark that this simulation method, which stems from the one of Gillespie, is very efficient. In fact, at each step, the time of the next event is computed. Thus the simulation time is increased to the time of the next event without considering any intermediate time. Moreover, we could speed up the simulation, if necessary, by using a known method [59] which allows simulations to become very fast by slightly approximating the results. In our models, given the number of population individuals, it is not necessary to resort to such an approximated method.

3. Applications

In this section we present some applications of our methodology. In particular, we construct and compare models based on female preference and male fitness. We obtain four different models by varying both the genetic representation of individuals (haploid or diploid) and the number of loci coding the female preference. Models of these kinds are studied in [38,32,33]. In all the models we assume three phenotypical traits for males (t_0, t_1 and t_2), giving rise to two kinds of ornamented male (m_{t_0} and m_{t_2}) and to one kind of non–ornamented male (m_{t_1}). Analogously, we assume three phenotypical preferences for females (p_0, p_1 and p_2). Females f_{p_0} and f_{p_2} prefer

males m_{t_0} and m_{t_2}, respectively, whereas females f_{p_1} have no preferences. Note that in this section, with abuse of notation, we use as index for males and females the phenotype instead of the genotype. Females f_{p_0} (resp. f_{p_2}) can mate with males m_{t_1} but cannot mate with males m_{t_2} (resp. m_{t_0}). We assume that ornamented males have a disadvantage with respect to non–ornamented ones as concerns trait cost (+10%). The cost of 10% for male traits is used also in [33]; we use the same value to better compare the results. Female cost $f_f(a)$ is assumed to be equal to 0 for all phenotypes, namely there is no cost for female preferences. This is an assumption used in various models [37,60,38,32]. In the final discussion we motivate it biologically.

We assume the fixed–relative–preference for mating [61]. Female preferences $pref(a, a')$ and male costs $cost_m(a)$ are summarized in Table 1, where preference 0 represents the impossibility of mating. Parameter w (assumed to be greater than 1) will vary in simulations.

	m_{t_0}	m_{t_1}	m_{t_2}
f_{p_0}	w	1	0
f_{p_1}	1	1	1
f_{p_2}	0	1	w

	Cost Value
m_{t_0}	0.1
m_{t_1}	0
m_{t_2}	0.1

Table 1. Female preference (on the left) and male cost (on the right).

We assume initial populations with the same number of males and females, and we study the evolution of the initial populations after 5000 generations. In order to maintain the size of the population constant on average, we set to 1 both the birth and the death environmental constants. Moreover, we consider the recombination in the generation of offspring genotype to be free. We want to observe the distribution of males and females in the final populations and we are interested in the following final results:

- **All phenotypes**: The final population contains all kinds of male and female.
- **Ornamented males lost**: The final population contains only non–ornamented males.
- **Ornamented males of one kind lost**: The final population contains only one type of the ornamented males plus the non–ornamented ones.
- **Shift**: The final population contains only ornamented males of one type only.
- **Non–ornamented males lost**: The final population contains only ornamented males of both types (m_{t_0} and m_{t_2}), and all kinds of females.
- **Speciation**: The final population contains only ornamented males of both types (m_{t_0} and m_{t_2}), and only choosy females (f_{p_0} and f_{p_2}).

The cases listed are not exhaustive. In the simulation we obtained a few final results with the total extinction of the population. All these cases are collected under the label **Extinction**.

The four models we present differ in the genotype of the population. We shall consider a haploid model in which male trait and female preference are coded each in one locus, a haploid model in which male trait is coded in a single locus and female preference is coded in two loci, and the two corresponding diploid models. Coding the male trait in a single locus is a realistic choice: for example the color morph of various cichlid fishes is coded by alleles of a single locus [62].

In the four models we assume an initial population with the same number of individuals for each genotype (half males and half females). This assumption is not very realistic, but it is the one which makes speciation most probable and we chose it to show that, also with extremely favourable initial conditions, sympatric speciation based on sexual selection is not easy to reach.

In the final results of our simulation we group males with the same trait and females with the same preference. We consider groups the frequency of which in the final population is zero, to be extinct.

3.1. *Haploid model with a single locus for female preference*

We consider a two–locus haploid model: the first locus determines the male secondary sexual character and the second one gives the female sexual preference. In particular, genotypes of individuals are represented by pairs of loci and the allele domains are $A_1 = A_2 = \{0, 1, 2\}$. In males, alleles 0 and 2 for the first locus represent ornamented individuals (m_{t_0} and m_{t_2}, resp.), whereas allele 1 represents non–ornamented males (m_{t_1}). In females, alleles of locus 2 correspond to the preferred male character. In particular, females with genotypes $f_{(\_,0)}$ and $f_{(\_,0)}$ (denoted by f_{p_0} and f_{p_2}, respectively) prefer the ornamented males with genotypes $m_{(0,\_)}$ and $m_{(2,\_)}$, respectively, whereas females with genotypes $f_{(\_,1)}$ (f_{p_1}) have no preference.

We considered an initial population of 9000 individuals, 500 for each possible combination of genotype and sex. We varied the female preference parameter w between 1.2 and 32. The use of these values for w allows an easier comparison with the results in [33]. For each value of w we performed 100 simulations.

The results are summarized in Table 2. Each column of the table corresponds to a value of the parameter w. Each entry of a column shows the number of simulations whose final population belongs to the class on the row.

3.2. *Haploid model with two loci for female preference*

In this model individuals are represented by a three loci genotype. The first locus determines the male trait, the second and the third determine the female sexual preference. The allele domains are $A_1 = \{0, 1, 2\}$, as in the previous model, and $A_2 = A_3 = \{0, 1\}$. Here, female preference is given by the sum of the allele values at loci 2 and 3.

We considered again an initial population with 500 individuals for each possible combination, that is in this case a population of 12000 individuals. As in the previous

female preference w	1.2	1.5	2	4	8	16	32
All phenotypes	0	0	0	0	0	0	0
Ornam. males lost	95	94	86	0	0	0	0
Ornam. males of one kind lost	0	0	0	0	0	0	0
Shift	0	0	11	85	88	87	79
Non-ornam. males lost	0	0	0	6	5	9	16
Speciation	0	0	0	5	6	2	2
Extinction	5	6	3	4	1	2	3

Table 2. Simulation results of sympatric speciation by varying the female preference parameter w in the haploid model with a single locus for female preference.

case, we varied the female preference between 1.2 and 32, and we performed 100 simulations for each preference value.

The results are summarized in Table 3.

female preference w	1.2	1.5	2	4	8	16	32
All phenotypes	0	0	0	0	0	0	0
Ornam. males lost	100	99	90	0	0	0	0
Ornam. males of one kind lost	0	0	3	0	0	0	0
Shift	0	0	6	39	35	36	36
Non-ornam. males lost	0	0	0	59	65	64	62
Speciation	0	0	0	0	0	0	0
Extinction	0	1	1	2	0	0	2

Table 3. Simulation results of sympatric speciation by varying the female preference parameter w in the haploid model with two loci for female preference.

3.3. *Diploid model with a single locus for female preference*

We consider a diploid representation of individuals with two loci. Hence, each individual is represented by a 4–tuple $\langle (a_1, a_2), (a'_1, a'_2) \rangle$, where a_1, a'_1 are the alleles for locus 1 and a_2, a'_2 are the alleles for locus 2. The allele domains are $A_1 = A_2 = \{0, 1\}$. Male trait and female preference are given by the sum of the two alleles a_1, a'_1 and a_2, a'_2, respectively.

We considered an initial population of 9600 individuals, 300 for each possible combination of genotype and sex. The results of simulations are shown in Table 4.

female preference w	1.2	1.5	2	4	8	16	32
All phenotypes	12	8	23	63	0	0	0
Ornam. males lost	0	0	0	0	0	0	0
Ornam. males of one kind lost	27	29	38	11	0	0	0
Shift	54	59	35	22	57	56	53
Non-ornam. males lost	0	0	0	0	0	0	0
Speciation	0	0	0	0	43	42	47
Extinction	7	4	4	4	0	2	0

Table 4. Simulation results of sympatric speciation by varying the female preference parameter w in the diploid model with a single locus for female preference.

3.4. *Diploid model with two loci for female preference*

We consider a diploid representation of individuals with three loci . The allele domains are $A_1 = A_2 = A_3 = \{0, 1\}$. Each individual is represented by a 6–tuple $\langle (a_1, a_2, a_3), (a'_1, a'_2, a'_3)\rangle$. Male trait is given by the sum of alleles a_1 and a'_1. Female preference is derived from the sum of a_2, a'_2, a_3, a'_3. Strong preference corresponds to sums 0 or 1 and 3 or 4, respectively. Females with sum 0 or 1 (f_{p_0}) prefer males with trait 0, whereas females with sum 3 or 4 (f_{p_2}) prefer males with trait 2. Females with sum 2 (f_{p_1}) have no preference.

We considered an initial population of 9600 individuals, 75 for each possible combination of genotype and sex. The results of simulations are shown in Table 5.

female preference w	1.2	1.5	2	4	8	16	32
All phenotypes	19	16	31	78	0	0	0
Ornam. males lost	0	0	0	0	0	0	0
Ornam. males of one kind lost	57	58	42	13	0	0	0
Shift	22	25	21	9	36	49	52
Non-ornam. males lost	0	0	0	0	0	0	0
Speciation	0	0	0	0	62	51	48
Extinction	2	1	6	0	2	0	0

Table 5. Simulation results of sympatric speciation by varying the female preference parameter w in the diploid model with two loci for female preference.

3.5. *General remarks*

Recall that we assumed an initial population with the same number of individuals for each genotype. A consequence of this assumption is that the frequencies of phenotypes are slightly different in different models. In particular, in both the first

and the last model, the same frequency of different genotypes corresponds to the same frequency of different phenotypes. As for the second and the third one, the difference in the initial distribution of phenotypes is not significant for the simulation results. We performed 40 simulations of each of these two models: 20 with female preference set to 2 and 20 with female preference set to 16. The results we obtained are reported in Table 6, which agree with the results in Sections 3.2 and 3.3.

It appears that in both haploid and diploid models when the value of the preference increases, we observe final situations with shifts, non-ornamented males lost, and speciation. However, haploid and diploid models exhibit some differences. In particular, in the haploid models we observe the following situations:

(1) When the female preference is low, the ornamented males suffer the disadvantage of being ornamented without the advantage of being chosen, and thus in almost all simulations they disappear.
(2) In both the considered haploid models we observe that, when the female preference passes from the value 2 to the value 4, simulation results pass sharply from situations in which non–ornamented males prevail to situations in which all non–ornamented males disappear.
(3) With high values of female preference speciation seldom occurs. In fact, even if non–ornamented males disappear, specially in the two loci model, non–choosy females survive. It seems that haploid representation of individuals disfavor linkage between trait and preference.
(4) In no case do we have final situations with all male phenotypes.

As regards the diploid models we observe the following:

(1) Ornamented males survive also with low female preferences. This is a consequence of the genotypes of ornamented males which can be reintroduced by recombination. Moreover, we observe a high number of shifts. This tendency, which seems to be counterintuitive, is due to the fact that the disadvantage of the ornamented males is balanced by the fact that their genotypes are reintroduced by recombination. This causes strong oscillations in frequencies of phenotypes of the population. When an oscillation produces a shift no further oscillations are possible and the population remains stable. A typical simulation with low female preference is shown in Figure 1. Here we represent only males grouped by phenotype (female phenotypes exhibit the same evolution). Each step in the figure corresponds to 250 generations. We observe that after initial oscillations ornamented males of one kind prevail.
(2) When the female preference increases population becomes more stable. We can see that, in both diploid models, when the female preference is 4 a high number of simulations end with a population in which all phenotypes are present.
(3) When the female preference is greater than 4 we have only shifts and speciations. We remark that diploid representation of individuals favors the linkage between trait and preference. This is evident, in particular, in the model with two loci

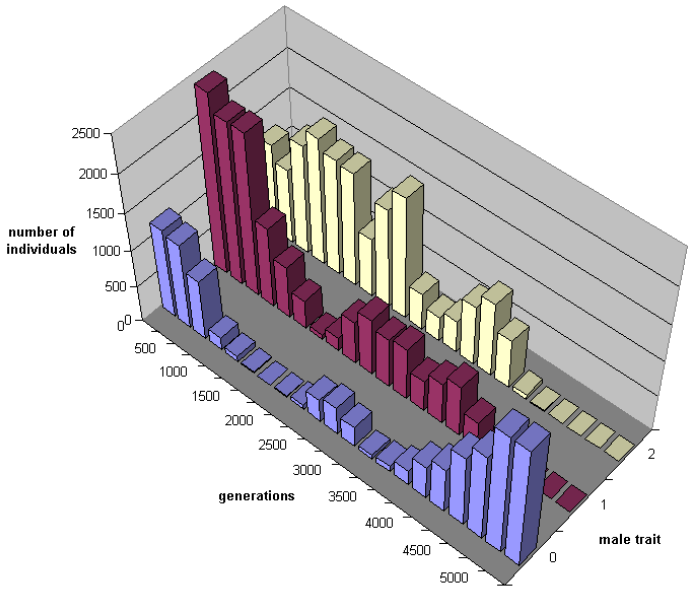

Fig. 1. Example of simulation in the diploid model with a single locus for female preference with $x = 1.5$.

for female preference.

	haploid		diploid	
female preference w	2	16	2	16
All phenotypes	0	0	8	0
Ornam. males lost	16	0	0	0
Ornam. males of one kind lost	0	9	5	0
Shift	4	10	7	13
Non-ornam. males lost	0	0	0	0
Speciation	0	0	0	7
Extinction	0	1	0	0

Table 6. Simulation results of sympatric speciation with the same initial frequency of each phenotype.

4. Discussion

Darwin introduced the idea that speciation can occur in sympatry due to natural selection. The general acceptance of the allopatric process of speciation proposed by Mayr [63] has made the possibility of sympatric speciation controversial.

In his seminal paper [9], Maynard Smith proposes a model in which stable polymorphism is possible. He argues that if alleles A and a have additive effects on the phenotype, a process of selection disadvantaging Aa genotypes would lead to a population consisting wholly of AA or wholly of aa individuals. However, if the population inhabits two contiguous environments, $E1$ and $E2$, such that AA is fitter in $E1$ and aa is fitter in $E2$, then a stable polymorphism is possible even if there is a genetic flow between the two subpopulations. In this situation if there are alleles at different loci, which cause assortative mating, two reproductive isolated populations can evolve. Maynard Smith concluded that the crucial step in his model of sympatric speciation is the establishment of a stable polymorphism in a heterogeneous environment.

The model of Maynard Smith, although posing strong conditions on the sympatric speciation hypothesis, renewed interest in such a topic. The Maynard Smith model is studied further in [10]. The approach to sympatric selection by considering both ecological and sexual selection is adopted also in [12], where the authors show that ecological forces may promote sympatric speciation even if assortative mating depends on an ecologically neutral trait.

Lande [37,60] proposed a model in which sexual selection can guide speciation. Although his model does not directly address the problem of sympatric speciation his arguments can suitably be used for supporting it. He considers a polygynous species with ornamented males. Males are promiscuous and females have many potential mates. Female mating preferences have no cost. Nevertheless, a selection of female preferences can occur as a product of the evolution of male traits. In this situation a male trait, which under natural selection would evolve towards an ecological optimal phenotype, may evolve to a suboptimal phenotype by sexual selection acting through mating success.

Two consequences of this model of evolution are: a positive genetic correlation between male characteristics and female sexual preferences, and a random genetic drift in female mating preferences which produce random selective forces on males.

The process is self–reinforcing since females choose more ornamented males and, through the genetic correlation, they are selected for more intense mating preferences. Runaway selection can thus proceed until halted by environmental selection on extreme male traits or extreme female preferences (due to the rarity of extreme males).

These considerations are the basis for the study of sympatric speciation by sexual selection: if sexual selection alone can produce an evolution in a random direction, such an evolution can have contemporary different directions in the same polymorphic population.

Many arguments can contrast this hypothesis. Mainly the fact that, at the beginning of the process, different runaway processes are prevented by recombination, and, to overcome this obstacle, a very strong mating preference is necessary. It seems to be improbable that strongly different mating preferences evolve in sympatry without any form of ecological diversification [36].

These arguments are supported by recent research in the field. In [53] Danley et al. conclude that species diversity among the "mbuna" cichlid fishes of Lake Malawi may depend on the fact that mbuna populations are isolated over extremely limited geographic scales and that such populations can diverge via drift or selection (as in the model of Maynard Smith). In [64] two species, *Pundamilia pundamilia* and *P. nyererei*, of cichlid fishes of Lake Victoria are considered. The two species live in sympatry and they are anatomically similar, but *P. pundamilia* males are metallic blue and *P. nyererei* males are bright red. The two species inhabit a continuous rocky slope from 0.5 to 7 m. water depth. *P. pundamilia* is most abundant at 0.5-2m, while *P. nyererei* is most abundant at 4-7m. The photic environment in the microhabitat of the two species differs, and it seems that the visual system of the two species have diverged in adaption to these different photic environments. Within a species, females prefer more conspicuous over less conspicuous males. The two species interbreed in locations where the water is very turbid. The authors conclude that the geographically sympatric and ecologically parapatric distribution suggests that the speciation process may have occurred without geographical isolation. This is a speciation theoretically described for cichlid fishes in [34].

Over the last few years many authors have argued in favor of sympatric speciation [65,11] by sexual selection either with or without weak ecological pressure. Models described in [38,32,33,28] follow this approach.

Turner and Burrows [38] propose an individual based computational model of sympatric speciation by sexual selection. The model assumes overlapping generations, male trait controlled by a number of unlinked loci, with additive phenotypical expression, and female preference determined by a single locus. Viability of males is reduced by their conspicuousness, while the mortality of females is random with respect to genotype. The absence of cost for female preference is inspired by real situations. As an example, consider species of the African cichlid genus *Cyprichromis* [54], that are torpedo–shaped open–water dwelling fishes which occur gregariously in the Tanganyika lake. Some species of this genus have polymorphic males in the same population. For example, yellow tailed and blue tailed males occur in the schools of *Cyprichromis leptosoma*. In such a situation there is no cost for choosy females which can easily find differently ornamented courting males in the same school. In [38], the population is considered spatially unstructured, and females, courted n times by randomly chosen males, mated with the preferred one among the males they encountered (the "best of n" rule).

Genotypes are diploid and female preference is determined by a single locus. Males range from white to black through a variety of shades of gray. The male's shade is controlled by a number of loci with additive effect. Black and white males

suffer very strong predation. Grey males are less conspicuous and suffer lower predation. Females prefer more conspicuous males. Initially, all the females prefer paler males. Despite female preference, if mortality of extreme males is high, the population can remain polymorphic. In this case the majority of males will be pale gray. The model shows that sympatric speciation can be induced by a dominant allele (arising from mutation) that produces a reversed female preference towards darker males.

The ease of speciation observed in this model may be due to the assumption that females choose mates using a "best of n" rule, which is optimal, although not common. It can approximate mate choice in lekking species, for example lekking cichlid fishes in which costs of sampling are low [51].

Arnegard and Kondrashov [28] propose individual based models which differ from each other by the adopted mating rule. The genotype representation is analogous to the one in [38], especially for male trait. The adopted mating rules are the following. The first is the sequential rule, in which a female tries to mate sequentially with a number of randomly selected males. Upon encounter, she accepts a male with a probability which depends on her preference on the male trait. Usually, there is a bound to the number of attempts. The second rule is the "best of n" rule with two forms of preference. The first form is the *disruptive form*, in which a female chooses the most extreme male she can find. The second one is the *affinity form*, in which a female chooses the male with the trait closest to her genotypical preference.

The model in [32] is still an individual based computation model with similar assumptions for female preference cost and population evolution. It differs in representing both male trait and female preference by multiple independent loci, and in coding the probability of mating according to the *psychophysical rule* [37]. Such a rule states that the mating probability is proportional to $e^{\alpha z}$, where α represents the efficiency of male discrimination by females, and z is the affinity between male trait and female preference. In [32] sympatric speciation occurs over a broad range of parameters. Kirkpatrick and Nuismer [36] suggested that this is due to the psychophysical rule which promotes speciation.

Kirkpatrick and Nuismer [36] present a deterministic model for sympatric speciation in which abiotic selection, intraspecific competition and sexual selection are based on the same trait. The trait is controlled by n loci that have equal and additive effects on the phenotype. The trait is expressed equally in males and females. In addition to ecological selection and assortative mating, the trait is responsible also for intraspecific competition. Two individuals with phenotypes x and y have a competitive effect on each other that depends on the distance of their phenotypes. A parameter, c_2, measures the specificity of the competition. Large values of c_2 mean that only phenotypically similar individuals compete. This kind of intraspecific competition is typically used in models based on adaptive dynamics. For a discussion about adaptive dynamics models see [66,67]. However, Kirkpatrick and Nuismer show the results of their model also for $c_2 = 0$, that is when intraspecific competition is independent from the phenotype of individuals (as assumed in the

models in [38,32,33] and in the present paper). Kirkpatrick and Nuismer assume that the trait is expressed in both males and females. Assortative mating is based on the similarity of phenotypes. These assumptions represent quite well the assortative mating reported on cichlid fishes of the genus *Tropheus* of Tanganyika lake [68]. The assumptions of our model, as well as the ones in [38,32,33], seem to better represent the assortative mating of other fishes, for instance haplochromine fishes of Lake Victoria [52].

Takimoto et al. in [33] propose a deterministic model with individuals represented by a haploid genotype with three loci. Two loci are used to designate three different male secondary sexual characters, and the third locus, with three alleles, determines female preferences. The model used the fixed–relative–preference for mating [61], as defined in Table 1. The analysis of results shows that the intensity of female preference has a strong positive effect in initiating speciation. Also the male cost for conspicuous ornaments influences speciation. The deterministic model results in speciation from any initial condition if there is no male cost. As expected, the introduction of female preference costs makes the speciation less possible. The surprising result of the model is that, under the female preferences and male fitness of Table 1 with $x = 4$, speciation occurs if the choosy females are more than 20% of the total population, independently from male frequency. A deeper discussion of these results is presented in the following.

We use our methodology to derive four stochastic models which differ in representations of individuals. We use essentially the approach in [33] for constructing female preferences and male traits in our models. The results of simulations are discussed in Section 3.5.

In the results of simulations we do not explicitly show the female genotypes distribution. Actually, we remark that the distribution of female preference genotypes is always wider than the one of the male trait genotypes. The reason is that non–choosy females are not limited by the availability of suitable males and consequently the selection pressure on them is weak. Similar results are found in [13].

The assumption on assortative mating on which our models are based are similar to the ones in the model by Takimoto et al. [33]. Therefore, it is important to summarize the similarities and the differences with respect to this model. Takimoto et al. consider a three loci haploid model, in which two loci determine the male trait and one locus determines the female preference. They construct a set of differential equations for the evolution of populations and perform intensive computer simulations for different values of initial phenotypical frequencies.

In the models, ornamented males are represented by phenotypes t_0 and t_2, respectively. Non–ornamented males are represented by the phenotype t_1. Analogously, females preferring t_0 (resp. t_2) are represented by phenotype p_0 (resp. p_2). The phenotype p_1 represents non–choosy females.

To simplify the model, Takimoto et al. focus on a population with symmetric initial distribution, and they assume that all the parameters relative to ornamented

males (t_0 and t_2) are the same, as well as the ones for choosy females (p_0 and p_2). Thus, the evolutionary process of the part of population characterized by t_0 and p_0 is symmetric with respect to the one characterized by t_2 and p_2. Within this model it is not possible to study shift of the population towards only one kind of ornamented males and choosy females. This happens because an evolution towards this kind of composition of the population is always interpreted as having a symmetric evolution of the other half of the population.

The representation of genotypes in [33] is very similar to the one of Section 3.1, where we consider a haploid model with only one locus for female preference. The results in [33], Figure 3 (a) and (c) where the preference intensity is set to 4 and 10, respectively, agree with our simulation results, Table 2, $x = 4$ and $x = 8$. The difference is that in [33] they are considered as speciations, while our simulations show that they correspond to shifts of the population.

Moreover, Figure 3 (b) in [33] shows that when female preference is set to 2.5, the equilibrium of the frequencies of choosy and non–choosy females is a separatrix between speciation and non–speciation. When the frequency of choosy females is greater than the frequency of non–choosy ones, the system evolves towards speciation. In the other case the system evolves towards the extinction of ornamented males. In all our simulations we have an initial population in which the frequencies of the different genotypes are the same. Thus, the different kinds of evolutions are obtained by varying the female preference. Accordingly to the haploid model in [33], a sharp change in the kind of evolution happens when the female preference passes from 2 to 4. Note that in [33] some simulations are performed by considering p_0, p_2, t_0 and t_2 as independent. However, given the deterministic definition of the model, when the initial frequencies of p_0 and p_2 (t_0 and t_2) are equal (as in the situations we consider) p_0 and p_2 (t_0 and t_2) evolve symmetrically.

Remark that we analyze also models formed by diploid populations by retaining all the assumptions on which assortative mating is based. This is easily accomplished by using our methodology of model construction.

We can conclude that sympatric speciation based only on sexual selection, especially in diploid populations, seldom occurs. Simulation results show that speciation is possible only with very strong female preference, even if there is a symmetric density of genotypes in the initial population (which is a favorable situation for speciation). It seems unlikely that such a strong preference may arise at the beginning of the speciation process, and that a low female preference produces shifts. This result can be considered robust with respect to genotype representation. Namely, the results of the simulations do not change considerably when the genotype representation varies.

Our results agree with the ones in [28] in considering sympatric speciation based on sexual selection unlikely, even though they are based on different mating rules. However, the first mating rule they adopt can be seen as analogous to ours: a female evaluates randomly chosen males and she mates according to given probabilities.

Because of the bound on the number of evaluated males, a female may not mate in a reproduction season. Our mating rule is similar, with the difference that, by using our simulation method, all females mate on average once during the reproduction season, as in [32,33]

5. Conclusions

We have proposed a methodology for the systematic derivation of computational models. We have shown the use of the methodology and the simulator by studying the problem of sympatric speciation by sexual selection with four different models, which produce essentially the same results. We believe that our methodology can be profitably adopted by biologists. Hypotheses can be validated by showing their independence from individual representation. This can be achieved by constructing a variety of models for different representations and by simulating them.

We have developed a computer program written in C for the simulation of the evolution of populations described by models built using our methodology. Such a software tool (called *SPECsim simulator* version 0.1) is available to be downloaded at the web page `http://www.di.unipi.it/~milazzo/biosims/`.

The simulation tool is fairly efficient. We have performed our simulations on a workstation powered by an Intel Pentium IV 2.0 Ghz processor. A single simulation of 5000 generations using the haploid model with a single locus for female preference (the simplest model) has taken on average about 150 seconds. A single simulation of 5000 generations using the diploid model with two loci for female preference (the most complex model) has taken on average about 500 seconds.

A comparison of the efficiency of our simulation method with respect to other individual based models (for example [28,32,38]) is not easy. In these papers a presentation of the algorithm used in the simulator and of its implementation is missing. It is important to remark that our method is based on a widely recognized efficient algorithm. In order to deal with populations with a number of individuals significantly greater than the one used in the present paper we could exploit known techniques [59].

1. Turelli M, Barton NH, Coyne JA, Theory and speciation, *Trends Ecol Evol* **16**:330–343, 2001.
2. Gavrilets S, *Fitness landscapes and the origin of species*, Monographs in Population Biology, Princeton University Press, 2004.
3. Kirkpatrick M, Ravigné V, Speciation by natural and sexual selection: models and experiments, *Am Nat* **159**:S22–S25, 2002.
4. Kirkpatrick M, Johnson T, Barton N, General models of multilocus evolution, *Genetics* **161**:1727–1750, 2002.
5. Liu LW, Price TD, Speciation by reinforcement of premating isolation, *Evolution* **48**:1451–1459, 1994.
6. Servedio MR, Reinforcement and the genetics of nonrandom mating, *Evolution* **54**:21–29, 2000.
7. Kelly JK, Noor MAF, Speciation by reinforcement: A model derived from studies of *Drosophila*, *Genetics* **143**:1485–1497, 1996.

8. Coyne JA, Orr HA, Patterns of speciation in *Drosophila*, *Evolution* **43**:362–381, 1989.
9. Maynard Smith J, Sympatric speciation. *Am Nat* **100**:637–650, 1966.
10. Gavrilets S, The Maynard Smith model of sympatric speciation, *J Theor Biol* **239**:172–182, 2006.
11. Via S, Sympatric speciation in animals: The ugly duckling grows up, *Trends Ecol Evol* **16**:381–390, 2001.
12. Dieckmann U, Doebeli M, On the origin of species by sympatric speciation, *Nature* **400**:354–357, 1999.
13. Van Doorn GS, Weissing FJ, Ecological versus sexual selection models of sympatric speciation: A synthesis, *Selection* **2**:17–40, 2001.
14. Shaw PW, Turner GF, Idid MR, Robinson RL, Carvalho GR, Generic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids, *P Roy Soc B-Biol Sci* **267**:2273–2280, 2000.
15. Turner GF, Speciation mechanism in Lake Malawi cichlids: A critical review, *Arch Hydrobiol* 44:139–160, 1994.
16. Schliewen UK, Tautz D, Pääbo S, Sympatric speciation suggested by monophyly of crater lake cichlids, *Nature* **368**:629–632, 1994.
17. Schliewen U, Rassmann K, Markmann M, Markert J, Kocher T, Tautz D, Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon, *Mol Ecol*, **10**:1471–1488, 2001.
18. Wilson AB, Noack–Kunnmann K, Meyer A, Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: Sexual selection versus ecological diversification, *P Roy Soc B-Biol Sci* **267**:2133–2141, 2000.
19. Barluenga M, Meyer A, The Midas cichlid species complex: incipient sympatric speciation in Nicaraguan cichlid fishes? *Mol Ecol* **13**:2061–2076, 2004.
20. Nagel L, Schluter D, Body size, natural selection, and speciation in sticklebacks, *Evolution* **52**:209–218, 1998.
21. Kondrashov AS, Kondrashov, FA, Interactions among quantitative traits in the course of sympatric speciation, *Nature* **400**:351–354, 1999.
22. Drossel B, McKane A, Competitive speciation in quantitative genetic models, *J Theor Biol* **204**:467–478, 2000.
23. McKinnon JS, Mori S, Blackman BK, David L, Kingsley DM, Jamieson L, Chou J, Schluter D, Evidence for ecology's role in speciation, *Nature* **429**:294–298, 2004.
24. Boughman JW, Rundle HD, Schluter D, Parallel evolution of sexual isolation in sticklebacks, *Evolution* **59**:361–373, 2005.
25. Bolnick DI, Multi-species outcomes in a common model of sympatric speciation, *J Theor Biol* **241**:730–744, 2006.
26. Rolan-Alvarez E, Sympatric speciation as a by-product of ecological adaptation in the Galician *Littorina saxatilis* hybrid zone, *J Mollus Stud* **73**:1–10, 2007.
27. Quesada H, Posada D, Caballero A, Moran P, Rolan-Alvarez E, Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail, *Evolution* **61**:1600–1612, 2007.
28. Arnegard E, Kondrashov AS, Sympatric speciation by sexual selection alone is unlikely, *Evolution* **58**:222–237, 2004.
29. Wu C-I, A stochastic simulation study on speciation by sexual selection, *Evolution* **39**:66–82, 1985.
30. Gavrilets S, Boake CRB, On the evolution of premating isolation after a founder event, *Am Nat* **152**:706–716, 1998.
31. Geritz SAH, Kisdi E, Adaptive dynamic in diploid sexual populations and the

evolution of reproductive isolation, *P Roy Soc B-Biol Sci* **267**:1671–1678, 2000.
32. Higashi M, Takimoto G, Yamamura N, Sympatric speciation by sexual selection, *Nature* **402**:523–526, 1999.
33. Takimoto G, Higashi M, Yamamura N, A deterministic genetic model for sympatric speciation by sexual selection, *Evolution* **54**:1870–1881, 2000.
34. Van Doorn GS, Noest AJ, Hogeweg P, Sympatric speciation and extinction driven by environment dependent sexual selection, *P Roy Soc B-Biol Sci* **265**:1915–1919, 1998.
35. Koeslag JH, On the engine of speciation, *J Theor Biol* **177**:401–409, 1995.
36. Kirkpatrick M, Nuismer SL, Sexual selection can costrain sympatric speciation, *P Roy Soc B-Biol Sci* **271**:687–693, 2004.
37. Lande R, Models of speciation by sexual selection on polygenic traits, *P Natl Acad Sci USA* **76**:3721–3725, 1981.
38. Turner GF, Burrows MT, A model of sympatric speciation by sexual selection, *P Roy Soc B-Biol Sci* **260**:287–299, 1995.
39. Pérez-Figueroa A, Cruz F, Carvajal-Rodríiguez A, Rolán-Alvarez E, Caballero A, The evolutionary forces maintaining a wild polymorphism of *Littorina saxatilis*: model selection by computer simulations, *J Evolution Biol* **18**:191–202, 2005.
40. Barbuti R, Caravagna G, Maggiolo-Schettini A, Milazzo P, An intermediate language for the stochastic simulation of biological systems, *Theor Comput Sci* (in press).
41. Barbuti R, Maggiolo-Schettini A, Milazzo P, Troina A, Bisimulations in calculi modelling membranes, *Form Asp Comput* **20**:351–377, 2007.
42. Barbuti R, Maggiolo-Schettini A, Milazzo P, Tiberi P, Troina A, Stochastic CLS for the modeling and simulation of biological systems, *Transac Comput Syst Biol IX, Lect Notes Comput Sc* **5121**:86–113, 2008.
43. Phillips A, Cardelli L, A correct abstract machine for the Stochastic Pi–Calculus, *Proc Concurrent Models in Molecular Biology (Bioconcur'04),* , 2004.
44. Priami C, Regev A, Shapiro E, Silverman W, Application of a stochastic name-passing calculus to representation and simulation of molecular processes, *Inform Process Lett* **80**:25–31, 2005.
45. Danos V, Desharnais J, Panangaden P, Labelled Markov Processes: Stronger and faster approximations, *Electronic Notes in Theor Comput Sci* **87**:157–203, 2004.
46. Heath J, Kwiatkowska M, Norman G, Parker D, Tymchyshyn O, Probabilistic model checking of complex biological pathways, *Theor Comp Sci* `391`:239–257, 2008.
47. Lanotte R, Tini S, Probabilistic bisimulation as a congruence, *ACM Trans Comput Log* **10**(2):9.1–9.48, 2009.
48. Barbuti R, Cataudella S, Maggiolo-Schettini A, Milazzo P, Troina A, A probabilistic model for molecular systems, *Fund Inform* **67**:13–27, 2005.
49. Barbuti R, Maggiolo-Schettini A, Milazzo P, Troina A, A Calculus of Looping Sequence for modelling microbiological systems, *Fund Inform* **72**:21–35, 2006.
50. Gillespie D, Exact stochastic simulation of coupled chemical reactions, *J Phys Chem* **81**:2340–2361, 1977.
51. McKaye KR, Sexual selection and the evolution of the cichlid fishes of Lake Malawi in Africa, in Keenleyside MHA (ed.), *Cichlid fishes: Behaviour, ecology and evolution*, Chapman and Hall, New York, 1991.
52. Seehausen O, van Alphen JJM, Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecol Lett* **2**:262–271, 1999.

53. Danley PD, Markert JA, Arnegard ME, Kocher TD, Divergence with gene flow in the rock–dwelling cichlids of Lake Malawi, *Evolution* **54**:1725–1737, 2000.
54. Takahashi T, Hori M, Description of a new Lake Tanganyikan cichild fish of the genus Cyprichromis (Perciformes: Cichlidae) with a note on sexual dimorphism, *J Fish Biol* **68**:174–192, 2006.
55. Ross S, *Stochastic processes*, John–Wiley, New York, 1983.
56. Murray JD, *Mathematical Biology*, Springer, Berlin, 1989.
57. Chen WY, Bokka S, Stochastic modeling of nonlinear epidemiology, *J Theor Biol* **234**:455–470, 2005.
58. Arkin A, Ross J, McAdams HH, Stochastic kinetic analysis of developmental pathway bifurcation in phage λ–infected *Escherichia coli* cells, *Genetics* **149**:1633–1648, 1998.
59. Gillespie D, Approximate accelerated stochastic simulation of chemically reacting systems, *J Chem Phys* **115**:1716–1733, 2001.
60. Lande R, Rapid origin of sexual isolation and character divergence in a cline, *Evolution* **36**:213–223, 1982.
61. Kirkpatrick M, Sexual selection and the evolution of female choice, *Evolution* **36**:1–12, 1982.
62. Majumdar KC, Nasaruddin K, Ravinder K, Pink body colour in *Tilapia* shows single gene inheritance, *Aquac Res* **28**:581–589, 1997.
63. Mayr E, *Animal species and evolution*, Belknap Press, 1963.
64. Maan HE, Hofker KD, van Alphen JJM, Seehausen O, Sensory drive in cichlid speciation, *Am Nat* **167**:947–954, 2006.
65. Panhuis TM, Butlin R, Zuk M, Tegenza T, Sexual selection and speciation, *Trends Ecol Evol* **16**:364–371, 2001.
66. Waxman D, Gavrilets S, 20 questions on adaptive dynamics, *J Evol Biol* **18**:1139–1154, 2005.
67. Doebeli M, Dieckmann U, Adaptive dynamics as a mathematical tool for studying the ecology of speciation process, *J Evolution Biol* **18**:1194–1200, 2005.
68. Salzburger W, Niederstätter H, Brandstätter A, Berger B, Parson W, Snoeks J, Sturmbauer C, Colour–assortative mating among populations of *Tropheus moorii*, a cichlid fish from Lake Tanganyika, East Africa, *P Roy Soc B-Biol Sci* **273**:257–266, 2006.