Computer Simulation of Sympatric Speciation with Penna Ageing Model

K. Luz-Burgoa[1], S. Moss de Oliveira[1], J.S. Sá Martins[1], D. Stauffer[1,2], and A.O. Sousa[2]

[1] *Instituto de Física, Universidade Federal Fluminense, Av. Litorânea s/n, Boa Viagem, Niterói 24210-340, RJ, Brazil*

[2] *Institute for Theoretical Physics, Cologne University, D-50923 Köln, Euroland*

One species is simulated to split into two separate species via random mutations, even if both populations live together in the same environment. This speciation is achieved in the Penna bitstring model of biological ageing, with modified Verhulst factors, and in part by additional bitstrings regulating phenotype and mate selection.

I   Introduction

The common ancestors of today's humans and today's chimpanzees presumably lived several million years ago. Then, due to genetic mutations and/or changes in the environment, the population split into the ancestors of humans and the ancestors of chimpanzees. Such a separation of one species into two is called "speciation". It is easily explained if the two populations live in separate environments, like one on an island and the other on a continent, making the mating of males from one population with females from the other population impossible. This effect is called allopatric speciation. More difficult to explain is sympatric speciation, where the two populations continue to live in the same environment but nevertheless cease to mate each other [1]. A recent computer model [2] in the physics literature also cites biological examples, serving as a background also for our paper. We concentrate here on models with age-structured populations [3, 4, 5] and in particular use the Penna bitstring model [6, 7], which is the presently most widespread model to simulate biological ageing. We deal only with sexual reproduction where two populations are defined as being different species if the mating from different populations produces no viable offspring.

The next section shortly explains this Penna model and then presents the minimal modifications which we found necessary to get sympatric speciation. A more realistic model involving three pairs of bitstrings instead of only one is discussed in the following section. We end with a short summary.

II   Simple Model

II.1 Penna Model

The Penna bitstring model for biological ageing [6, 7] simulates the mutation accumulation by storing bad mutations (= hereditary diseases) in a string of (usually) 32 bits. The position of the bit corresponds to the age of the individual; its value corresponds to health (zero) or sickness (one). Sick bits act from the age to which their positions belong up to the death of the individual. Three active sicknesses kill the animal; in addition all animals are killed at each time step $t = 1, 2, \ldots$ with the Verhulst probability $V(t) = N(t)/N_{\max}$ where $N(t)$ is the current population and $N_{\max}$ is often called the carrying capacity describing the limitation of food and space. After reaching an age of eight "years" (= bit positions or iterations), each individual gives birth to one child per year until its death; the child inherits the same bitstring as the mother except for one possible mutation at a randomly selected position where the mutated bit is set to one irrespective of its previous value. Initially, all bit strings are zero.

For the sexual Penna model used here, each individual has two bitstrings inherited from mother and father, respectively, and produces gametes (single bitstrings) by random crossover between these two bitstrings, followed again by one random mutation. Each female of age eight or above tries 20 times to find randomly a male aged eight or above for mating, and if she succeeds she gets one child having one of the father's gametes and one of the mother's gametes as its two bitstrings. The child's sex is fixed randomly. If at a specific bit position, one of two bitstrings has zero and the other has one, it affects the health of that individual if and only if this position is one for which the harmful allele is dominant and not recessive. Six out of the 32 possible positions are randomly selected as dominant, the other 26 as recessive. The complete Fortran program is listed in [7].

II.2 Speciation Model

Speciation is now attempted by defining one bit position, which we take as position 11, as influencing the mating. Each individual has $n = 0, 1,$ or 2 bits set at this position. A female with n such bits at position 11 selects only males with the same number n of such speciation bits. Due to the

randomness of mutations and crossover, their child does not necessarily have n speciation bits set to one, and this randomness allows the emergence of a new species out of the original one where all n were zero. At every time step t we thus have three populations N_n depending on the number $n = 0, 1, 2$ of speciation bits set to one, and each of these three sub-populations is half male and half female.

To get speciation in this model, it is not sufficient to start with one population and to change it due to random mutations into another population. Instead, we want to start with one population and at the end have two populations coexisting with each other in stable equilibrium but without crossmating. With the above Verhulst factor applying to both populations together this seems impossible: After some time (proportional to the population size), all survivors have one certain "Eve" as mother [7] as a result of the random mutations due to which one individual happens to become the fittest in the population. Thus one population finally wins against the other, just as in a standard Ising model at low temperatures, with equally many up spins and down spins at the beginning, one of the two spin orientations will win at the end.

Coexistence is achieved by replacing the Verhulst factor $N/N_{\max}$ into three separate Verhulst factors $N_n/N_{\max}$ for the separate populations $n = 0, 1, 2$ with $N = N_0 + N_1 + N_2$. We may imagine, for example, that the original population $n = 0$ is vegetarian, and that the second population $n = 2$ emerging out of it consists of carnivores. Both populations are limited by the amount of food, but their food sources are completely different; thus, there is no competition between the two different populations, but the meateating females will not select any herbivore males for mating, and vice versa. (The small population with $n = 1$ is added half to $n = 0$ and half to $n = 2$ for the evaluation of the two Verhulst factors $V_0 = (N_0 + N_1/2)/N_{\max}$ and $V_2 = (N_2 + N_1/2)/N_{\max}$ and has the arithmetic average of these two Verhulst factors as its own food-limiting Verhulst factor.)
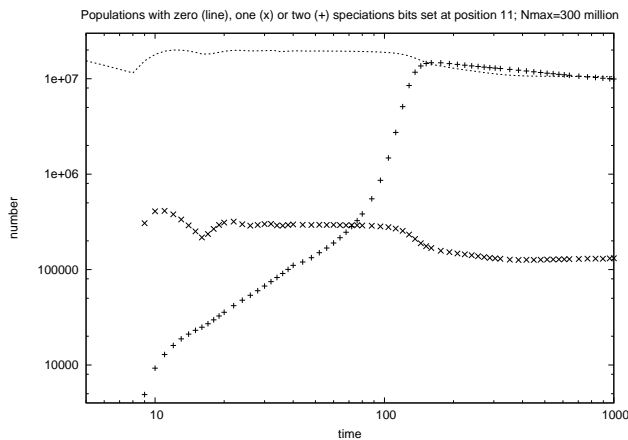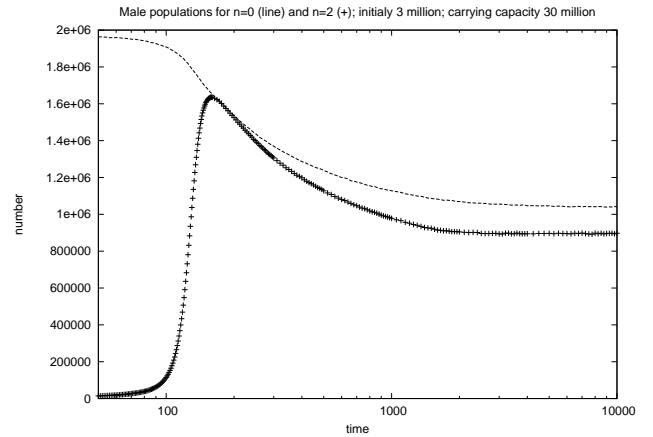

Figure 2. Same as in Fig. 1, but with populations ten times smaller simulated over ten times more iterations, showing equilibration after several thousand iterations. Note linear, not logarithmic, vertical scale. The dots refer to the original population, the + to the new species (N_1 is not shown).

II.3 Speciation Results

Figure 1 shows, for nearly 10^8 individuals, how the new species N_2 emerges, within about a hundred iterations, from the old species N_0. The intermediate population N_1 is only about one percent of the total. Thus, we see two separate populations clearly emerging: Sympatric speciation. Fig. 1 shows the males; for the females the results are about the same. Making ten times more iterations for a ten times smaller population gives a better equilibrium, Fig. 2. Shifting the speciation position from 11 to 21 or to 1 does not change much the results. If we change the birth rate from 1 to $1 + n$ where n refers to the female, then the new species ends up with a larger population than the original one but still may both coexist (not shown).

Similar pseudo-speciation was observed in the asexual case (not shown), or when the speciation bit was mutated with a special probability different from that of the other bits (not shown). Also on a square lattice without Verhulst factor, similar results were found [9].

III   The Penna model with phenotype selection

In this section we will briefly describe the features we added to the standard Penna model to represent phenotypic selection [3].

The first modification deals with the Verhulst factor, which has already been criticized in the literature (see, for instance, Ref. [10]). Rather than random, one should expect that deaths caused either by predation or because of intraspecific competition should also have selective value. The probability $V(t)$ of their occurrence should be dependent on the fitness of the individual to the environment, expressed as a function of the match between the individual's phenotypic expression of genetically acquired traits and a phenotype ideally adapted to its habitat.

Figure 1. Variation in time of N_0 (line, original species), N_1 (x, mixed genomes) and N_2 (+, new species), with $N_{\max} = 300$ million. We start with 30 million males and as many females of the original species.

To address this issue in the context of the Penna model we add, to the usual double strand of genes with age-dependent expression, extra pairs for each individual in the population. Each of these pairs will stand for a particular multi-locus phenotypic trait, such as size or colour, which may have selective value. The dynamics of reproduction and mutations are the same for both the age-structured and the new strings. Gamete production, phenotypic expression (phenotype), and the introduction of random mutations at birth affect equally all loci.

A final addition refers to mating selectiveness. We introduce into the genome a locus that codes for this selectiveness, also obeying the general rules of the Penna model for genetic heritage and mutation. If it is set to 0, the individual will not be selective in mating (panmictic mating), and it will be selective (assortative mating) if this locus is set to 1. The mutation probability for this locus is set to 0.001 in all simulations. Females that are selective will choose mating partners that satisfy some criterium related to the sexual selection trait. In the first version described below, we add a single phenotypic trait to the model, and this trait acts both on the individual's fitness and on its sexual selectiveness. In the second version, two traits are added, one that is related to fitness and another to sexual selection. In the road to speciation, as one population splits into two, the degraded fitness of the offspring of inter-populations parents will lead nature to choose for sexual selectiveness and give rise to two non-interbreeding populations.

Assortative mating following the establishment of a stable polymorphism is essentially equivalent to speciation in this context, and one of the purposes of these simulations is to follow the rising of the fraction of the population that becomes sexually selective as a result of the evolutionary conflict between selection and mutation.

III.1 Version with a single phenotypic trait

In the version with a single trait, one pair of bitstrings is added to the genome of each individual (see Fig. 3). Its selective value is given by a fitness function $F(k)$, chosen to depend on a single variable k. Here, k results from the pair of non-structured strings representing the trait by counting the number of recessive bit positions where both bits are set to 1, plus the number of dominant positions with at least one of the two bits set. It will therefore be a number between 0 and 32. For a given value of the phenotype k, $F(k)$ quantifies the availability of resources for individuals with that particular phenotype. The fitness function encapsulates the selective value of a particular phenotype, and can be thought of as describing the profile of the ecology in which the population evolves, as related to the size of the individuals, for example.

Our simulations were done with two different functional forms for the function $F(k)$. Our intention was to mimic the seasonal effect of rainfall on the availability of seeds of different sizes in the Galapagos islands and its impact on the morphology of beak sizes in the population of ground finches that feed on these seeds [8]. In this application, k is a measure of the beak size. At the beginning of the simulations, $F(k)$ is a single-peaked function with a maximum at $k = 16$, representing large availability of medium-sized

seeds. The whole population of finches will compete for the same resources. After 12 000 iterations, function $F(k)$ changes to a two-peaked shape, with maxima at $k = 0$ and $k = 32$, and now the food resources concentrate on either small or large seeds, with a vanishing number of medium-sized ones.

In this discretized mathematical model, the probability of death by intra-specific competition at each time step is now given by $V(t) = N(t)/(N_{max} * F(k))$, where $N(t)$ will account for the population that competes for resources available to individuals of phenotype (beak size, say) k. From step 0 up to step 12 000, the whole population is competing for the same general food resource, and $N(t)$ is the total population. After step 12 000, only small(large)-beaked individuals - those with $k < (>)16$ - can compete for the small(large) seeds. For that reason, the death probability $V(t)$ of an individual with $k < (>)16$ is computed by assigning to $N(t)$ the number of individuals with $k < (>)16$ plus half of the population that has $k = 16$. An individual with $k = 16$ will compete either for small or large seeds, and this choice is made at random.

With this strategy, the interaction with other individuals is still mediated by a mean-field $V(t)$, but its effect is no longer uniform. It depends on the particular genetically acquired configuration of each individual and, although stochastic, escapes from the biologically unmotivated randomness mentioned above.

Sexual selection also depends on this single trait. When the fitness function $F(k)$ is single-peaked, there is no selective pressure for mating selectiveness, and the population is panmictic. After $F(k)$ becomes double-peaked, females that mutate into selectiveness will choose mating partners that have beak sizes similar to their own: if a female has $k < (>)16$ and is selective, she will only mate with a partner that also has $k < (>)16$.
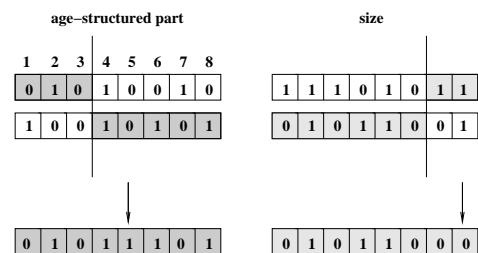

Figure 3. Schematic representation of crossing, recombination, and mutation in the formation of the gamete from one of the parents. The other gamete is generated in the same way, and the union of the two gametes concludes the formation of the offspring's genome. We show these processes occurring both in the age-structured and the non-structured portions. The arrows indicate where the random mutations occurred.

III.2 Simulation results: version with a single phenotypic trait

Simulations were run for population sizes of the order of 10^5, with minimum reproduction age set to 10 for both males and females. Each female gives birth to 2 offspring after finding a suitable partner to mate with. The age-structured part of the genome has the harmful allele as dominant in 3 loci, and is initially set to blank in the whole population. The allele 1 is dominant in 16 out of the 32 loci of the non-structured part; an allele will mutate in the gamete of the parents with probability 0.1 per birth. At the start of the run, this part of the genome is randomly generated, and all females are set as non-selective. The fitness function is, up to step 12 000,

$$F(k) = 1 - \frac{\mid 16 - k \mid}{128}$$

The denominator 128 ensures a mild selective pressure for middle-sized beaks. Only a negligible fraction of the females end up becoming sexually selective, showing that there is no evolutionary advantage for selectiveness in this ecology. At step 12 000, and up to the end of the run at step 50 000, the fitness function suddenly changes to
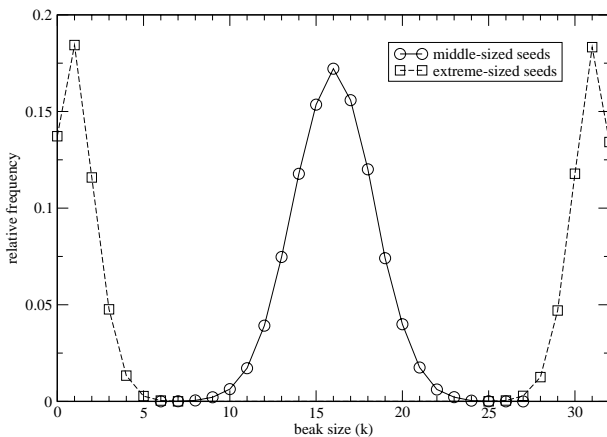
$$F'(k) = \frac{\mid 16 - k \mid}{16}$$

Figure 4. Morphology of the population when only middle-sized seeds are abundant (circles) and when only small and large-sized seeds are left (squares). We show the fraction of the population with each beak size.

In Fig. 4 we show the distribution of phenotypes - in this case, the beak size - of the populations at time step 12 000, up to which the fitness function was single-peaked, and at time step 50 000, after it has been double-peaked for 30 000 time steps. A stable polymorphism has clearly been established as a result of the duplicity of food resources. The fraction of selective females in the population has also increased to nearly 1.0, and we now have two distinct populations each of which does not mate with a partner from the other, since

hybrids are poorly fit to the environment: evolutionary dynamics made it advantageous to develop assortative mating in this bi-modal ecology, and as a consequence of reproductive isolation one single species has split into two.

III.3 Version with two traits

The version with two traits has two additional pairs of non-structured bitstrings in the genome of each individual. One of them is related to fitness, like in the previous version, and the second introduces a representation of a trait that drives sexual selection. The purpose here is to make the model more realistic by assigning different traits to different functions and to study the interactions between these traits [11], as well as to address issues raised by recent observations of speciation in fish [12]. The dynamics of death and birth follow the rules already stated, and we map the phenotype space for this new sexual trait onto an integer between 0 and 32, just as done for the fitness trait described above. For mating, a female will choose, among a random selection of a fixed number of males, chosen to be 6 in our simulations, a suitable mating partner for whom the phenotype for this second trait (colour, say) matches her own, under some strategy. For the results we show in this paper the mating strategy is defined as follows: Call f the phenotype of the female and m the one for the male, for the sexual selection trait. Then, the female follows the rules, if mutated into selectiveness,
- If $f < 16$ then it selects the male with the smallest m;
- If $f > 16$ then it selects the male with the largest m;
- If $f = 16$ then the female chooses randomly to act as one of the above.

These rules amount to choosing a mating partner that further enhances the phenotype into which the trait is mapped. If we think about this mating trait as colour, for instance, and assign $f < (>)16$ to a blue (red) character, a blue (red) female will choose the male that lies deepest into the blue (red) region.

III.4 Simulation results: version with two traits

Simulations were run with the same parameters as the ones for the single trait version, and the equilibrium population size is 6×10^4. The mutation probability for new sexual selection trait is set to 1. The fitness function is set to a constant up to step 12 000. The distribution of the fitness trait is single-peaked at $k = 16$ at step 12 000, as a consequence of the number of loci (16) where the 1 allele is dominant, and moves into a polymorphism after the ecology becomes bi-modal. The sexual selection trait also shows a single peak at step 12 000, and splits the population in two groups afterwards. But now a strong correlation develops between these traits, and the individuals with sexual selection phenotype $< (>)16$ have their fitness trait $< (>)16$. In other words, a female chooses a mating partner because of his colour, and its correlation with size allow them to generate viable offspring. Fig. 5 shows the distribution of both traits at the end of the simulation.
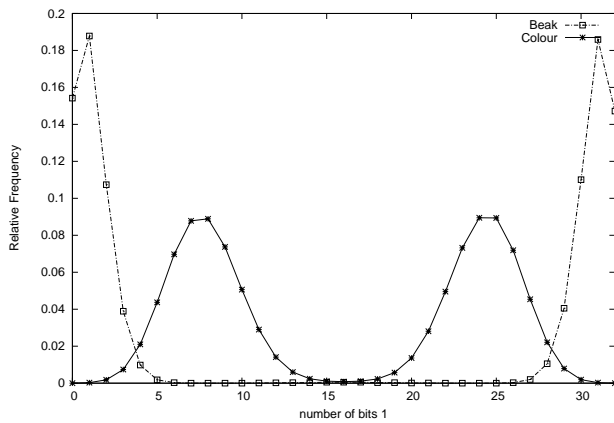
Figure 5. Distribution of fitness (squares) and sexual (*) traits at the end of the simulation. These traits are correlated, and the population with fitness trait to the left has its sexual trait also to the left of the plot.

Sexual selectiveness also develops as a result of the evolutionary dynamics. At the end of the simulation all females are selective, and again assortative mating and reproductive isolation are the proxies in this model to the development of two separate species out of the single one that existed at the beginning.

IV   Summary

We presented three distinct versions of bit-string models developed to study speciation. They are modifications of the original Penna model for ageing, and were discussed in an order of increasing complexity and realism. In the first, a simple interpretation of one locus in the genome as identifier of species, the imposition of assortative mating from the start, and a bi-modal ecology, lead to the flourishing and co-existence of two separate populations. The second version introduces a non-structured part in the genome and, by mapping the genetic information onto a number, relates it to the fitness of the individual. Variability in the ecology from a single- to a double-peaked supply of resources generates a

stable polymorphism. By introducing sexual selectiveness in the genetic information and using the fitness trait as the one that also codes for mating choice, assortative mating and reproductive isolation develops. The last version deals with two separate traits, one for fitness and another for sexual selection. Evolutionary dynamics again develops a splitting into two separate species by generating a strong correlation between the traits.

References

[1] U. Dieckmann and M. Doebeli, Nature **400**, 354 (1999).

[2] S.V. Semovski, Y.S. Bukin, and D.Yu. Sherbakov, Int. J. Mod. Phys. C **14**, issue 8 (2003); see also M. Rost and M. Lässig, Diapatric speciation, preprint.

[3] J.S. Sá Martins, S. Moss de Oliveira, G.A. de Medeiros, Phys. Rev. E **64**, 021906 (2001).

[4] A. Nowicka, A. Duda and M.R. Dudek, cond-mat/0207198; A. Łaszkiewicz, Sz. Szymczak, S. Cebrat, Int. J. Mod. Phys. C 14, issue 6 (2003).

[5] D. Chowdhury, D. Stauffer, and A. Kunwar, Phys. Rev. Lett. **90**, 068101 (2003).

[6] T.J.P. Penna, J. Stat. Phys. **78**, 1629 (1995).

[7] S. Moss de Oliveira, P.M.C. de Oliveira and D. Stauffer, *Evolution, Money, War and Computers*, Teubner, Leipzig and Stuttgart 1999, ISBN 3-519-00279-5.

[8] P.T. Boag and P.R. Grant, Nature **274**, 793 (1978); P.T. Boag and P.R. Grant, Science **214**, 82 (1981).

[9] A.O. Sousa, Stuttgart University, priv. comm. (April 2003)

[10] J.S. Sá Martins and S. Cebrat, Theory Biosc. **119**, 156 (2000).

[11] A.S. Kondrashov and F.A. Kondrashov, Nature **400**, 351 (1999)

[12] A.B. Wilson, K. Noack-Kunnmann, and A. Meyer, Proc. R. Soc. Lond. B **267**, 2133 (2000).