Impact of population size at a rate of morphological and molecular evolution - the use of an object-oriented model

Miłosława Sokół

Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-098 Warsaw, Poland, email: milka@biol.uw.edu.pl

Abstract

A generalization of Moran model of evolution is created using object-oriented method of modelling. A population consists of individuals which have a genotype and a phenotype. The genotype is inherited by descendants and it can mutate. The phenotype is dependent on the genotype. Moreover, the phenotype causes changes in the fitness of the individuals (natural selection which four kinds are defined and analysed). Evolution of the population appears spontaneously. This model is used to analyse how population size influence the rate of evolution.

Evolution is manifested by two processes: the increase of the phenotype size (morphological evolution) and number of mutations accumulated on genes (molecular evolution). The rate of evolution increases if population size increases. An adaptive natural selection causes nonlinear changes in the phenotype size and number of mutations accumulated on genes. A competitive natural selection causes linear evolution. A survive natural selection causes the faster evolution than a reproductive natural selection.

Keywords: morphological evolution, molecular evolution, genotype-phenotype interaction, adaptive natural selection, competitive natural selection, survive natural selection, reproductive natural selection, object-oriented model

1. Introduction

In various branches of biology "evolution of a population" is defined differently. In palaeontology and ecology the evolution is defined as an increase,

rarely a decrease of phenotype size of individuals from generation to generation [23], [21], [42]. This process is named morphological or phenotypic evolution. Molecular biologists and geneticists define evolution as an accumulation of mutations (substitutions) on genes, a formation of new alleles and often new RNAs or proteins in cells. For this phenomenon the terms molecular evolution or genetic evolution are used [12], [5]. But the molecular evolution influences the morphological one and rates of both processes can be compared [2], [3], [32].

The relationship between morphological and molecular evolution has not been well explained by mathematical models so far. In accordance with the principle of modelling only one numeric variable, mathematical models correspond to the morphological evolution ([20], [27], [22], [7]) or to the molecular one ([28], [14], [48])separately. The basic difference between these groups of the models depends on the assumptions about a single mutation of a gene.

1. A mutation changes the size of the phenotype by a random positive or negative number in the models of morphological evolution. Beneficial and adverse mutations have the same probability.
2. A mutation changes the sequence of nucleotides, creates a new allele of a gene in the models of molecular evolution. Back mutations are impossible or they have very little probability.

We know that the genotype influences the phenotype of an individual. This is a dependence at an individual's level.

$$\text{GENOTYPE} \rightarrow \text{PHENOTYPE}$$

Genotypes and phenotypes of all individuals in each generation of the population form the genetic and phenotypic structures and the dependence between them runs in the opposite direction:

$$\text{PHENOTYPIC STRUCTURE} \overset{SELECTION}{\rightarrow} \text{GENETIC STRUCTURE}$$

This is a dependence at a population level. So far only some generalizations of Wright - Fisher and Moran genetic drift models with natural selection take into account this pattern of dependencies ([43], [17], [33], [6], [44]). The selection in these models are introduced by assigning to each allele a specific numerical value that is called phenotypic fitness. This number may be identified with a probability of death, a probability of reproduction or

mean number of descendants of individuals of a given genotype. In these models heredity and selection are only assumed. Evolution (genetic drift to the best fitness) appears spontaneously.

Including mutations to the Moran model creates the biggest problems. So far it is realised on a population level ([43]). But the important mutations appeared when copying parental genes to a descendant. They create a new allele but we can never predict their effect on the phenotype. The mutations should be modelled on an individual's level. Such assumptions are difficult to make in Moran models created as numeric stochastic processes or differential equations. But they are possible in Moran model created as an object-oriented model of a population.

The object-oriented model forms a virtual population after start-up. The virtual population is a set of individuals and the individuals are formed as classes, structures or otherwise named objects. Nowadays simulations of such models are very fast. This model has many advantages (for instance: it allows us to analyse many variables, all parameters of the model have biological interpretation and can be estimated by experiments, the time can be expressed with real units and many others ([15]) and only has one disadvantage. The simulation is possible if all parameters are replaced by numbers. So the interpretations of results of such modelling are not so general as within mathematical models are done. According to mathematical rules it can only be used to show contrasamples to some hypotheses. But it can also be used to correctly formulate a mathematical model of such phenomena, which are nowadays modelled in very different ways generating inconsistent results. Evolution of a population is such a phenomenon.

The primary aim of this paper is to present the basic form of genotype-phenotype (gen-phen) model. It is a model of the virtual population with heredity, mutations and selection. Morphological and molecular evolution of this population occurs spontaneously. Such evolution can be analysed in the same way as it is done by biologists (calculation of the number of mutations accumulated on genes and the changes of phenotype size from generation to generation). Another aim of this article is to use the presented model to analyse the rate of evolution due to population size to check if some assumptions are true in some mathematical models of evolution.

2. Methods

2.1. The model

The population was programmed as a finite set of virtual objects (individuals) that they appear as a result of "breeding" of existing individuals, they exist during some time-period, they sometimes reproduce (a random event called "reproduction") and they disappear (a random event called "death"). Each individual has two genes. Each descendant inherits one gene from its mother individual and the other gene from a randomly selected individual (father). The inheritance involves copying these genes. Sometimes the parental gene mutates with probability p_{mut} during this copy. In this model each mutation gives a new allele of the gene and change the descendant's phenotype size.

The individual's gene is a sequence of three numbers: (x,y,z), where x is a real number (after a mutation of this gene this number can be decreased or increased by a random number drawn from the normal distribution $N(0, \sigma_g)$), y is an integral number - allele's name (after the mutation this number is changed to a new one, which has not been used yet), z is an integral number - the number of mutations accumulated on the gene relative to the ancestor from the initial population (this number is increased by 1 after the mutation of this gene).

Each individual has phenotype size - a real number which can be assigned to some numerical variable (individual mass, length of legs, rate of escape, etc.). The phenotype size is formed basing on the inherited genotype and different random features. In the model it is a random number Φ drawn from the normal distribution $N((x_1 + x_2)/2, \sigma_e)$ where x_1 is a first number of the mother gene, x_2 is first number of the father gene. So, the number $x_1 = (x_1 + x_1)/2$ is mean phenotype size of all homozygous individuals AA with gene A=(x_1, y_1, z_1).

The parameter σ_g describes the size of the effect of one mutation to the phenotype size. The parameter σ_e describes the size of the impact of non-genetic (among others environmental) factors on the phenotype size. Values of them are dependent on a unit that measures of the phenotype.

The probability of reproduction of an individual during some time-step (generation) depends on its phenotype size and population density (classical regulatory mechanisms). It is calculated by the function $p_r(N/V, \Phi)$ where N is a population size (number of individuals) and V is an area of the territory. The offspring size (number of descendants of one individuals produced at the

same time) is a random variable drawn from some discrete distribution $P_\lambda(L)$, whose expected number λ is calculated with the function $L_{off}(N/V, \Phi)$. The probability of death of an individual during some time step depends on its phenotype size and population density and it is calculated by the function $p_s(N/V, \Phi)$.

These assumptions allow to construct an algorithm of virtual population changing in time by random events (Fig.1). The program simulating such population can be written in many computer languages, but the simulations are faster for non-interpreted languages. The program written using C++ was shown in the appendix 1, using Python - in appendix 2.

The model is characterized by three functions: $L_{off}(N/V, \Phi)$, $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$, the probability of mutation p_{mut} and parameters σ_g and σ_e. The effects of simulation dependent on used functions and values of parameters but also on size and structure of an initial population.

2.2. Conditions for a long simulation

For many functions $L_{off}(N/V, \Phi)$, $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$ the simulated population becomes extinct very fast or its density runs to infinity. The conditions of long simulation are only well known for populations in which these functions are not dependent on individual's phenotype or the phenotype does not change over time ([37], [41]). Then the dynamics of the population size is the same as realizations of simple stochastic process of a population dynamics in discrete time generalized by the random offspring sizes.

If Φ is constant and:

1. $L_{off}(0, \Phi)p_r(0, \Phi) > p_s(0, \Phi)$,
2. $L_{off}(N/V, \Phi)p_r(N/V, \Phi) < p_s(N/V, \Phi)$ for all great N,

then the population stabilizes its own size as numbers fluctuating around the N_E (equilibrium number), such that:

$$L_{off}(N_E/V, \Phi)p_r(N_E/V, \Phi) = p_s(N_E/V, \Phi). \quad (1)$$

Ecologists say the functions $L_{off}(N/V, \Phi)$, $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$ create a significant regulatory mechanism in the population. Populations with significant regulatory mechanism can subsist for very long time ([37]).

If functions $L_{off}(N/V, \Phi)$, $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$ significantly depend on Φ, the equilibrium number N_E does not exist. But the low probability of mutation does not cause major changes in the average phenotype

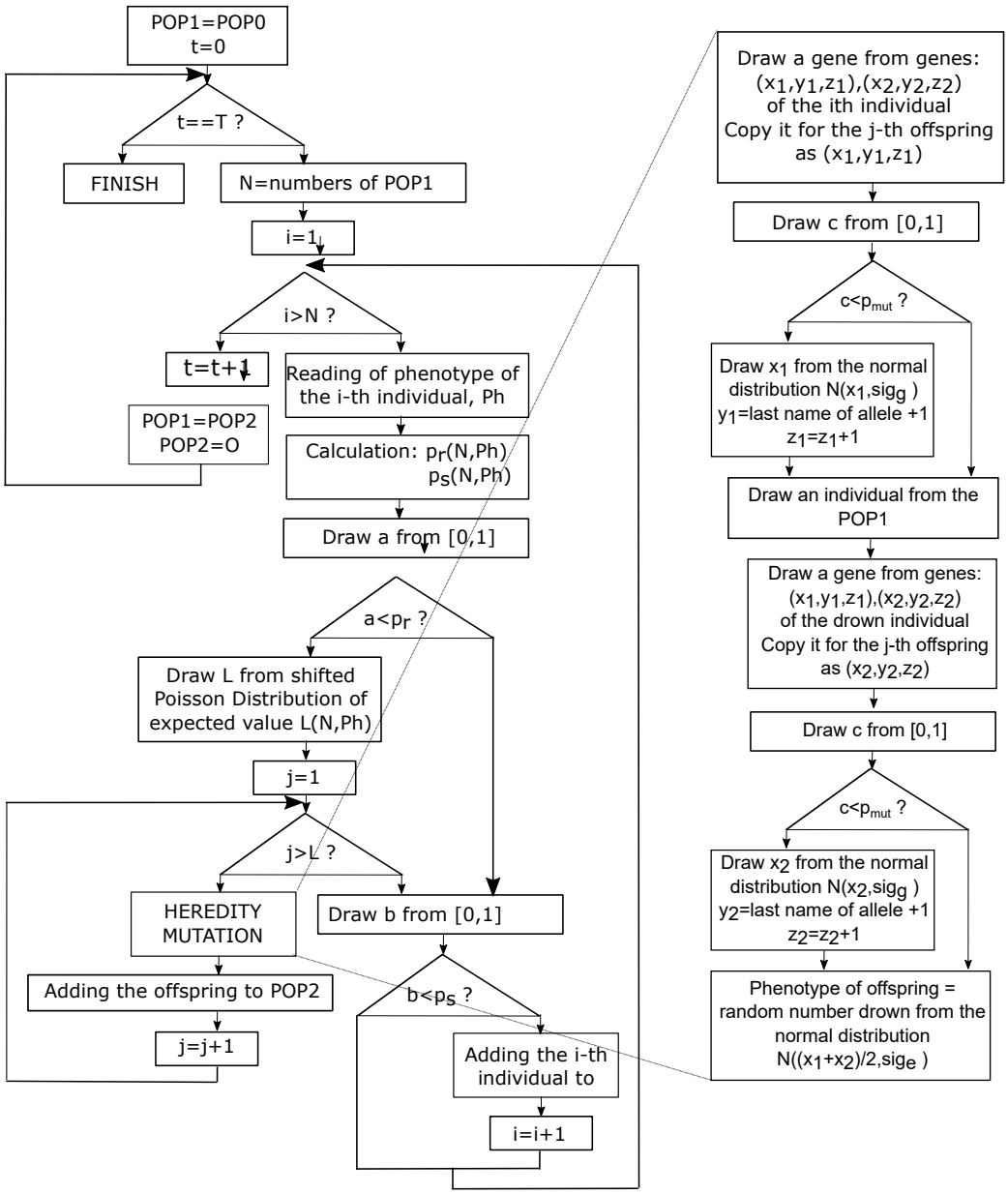

Figure 1: The algorithm of the programme simulating a virtual population consisting individuals having genotypes and phenotypes. The methods of programming the heredity, mutations of the genes and selections among the phenotypes are shown.

size from generation to generation, so a number N_E being a solution of the equation:

$$L_{off}(N_E/V, \bar{\Phi}) p_r(N_E/V, \bar{\Phi}) = p_s(N_E/V, \bar{\Phi}). \quad (2)$$

well characterizes population size (where $\bar{\Phi}$ is mean value of the sizes of the phenotypes in given generation). The population fastly stabilizes its size as fluctuations around such N_E.

2.3. Non-linear functions

Probabilities have values between 0 and 1. This obvious fact is often omitted when probabilities are a part of some model's parameter that can take any value. In the virtual population the functions: $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$ are only upper limits for numbers drawn from the interval $[0, 1]$. Even though $p_r(N/V, \Phi) = a_r N/V + b_r \Phi + c_r$ this function is not linear. If $a_r N/V + b_r \Phi + c_r < 0$ then the reproduction is impossible, so $p_r(N/V, \Phi) = 0$. If $a_r N/V + b_r \Phi + c_r > 1$ then the reproduction always runs, so $p_r(N/V, \Phi) = 1$. The probability of death also has these features.

Distribution of offspring size usually has its support consisting positive or non-negative integers. Its expected value is positive. The draw an integer from a non-existed distribution (of negative expected value) cannot takes place for many algorithms. In my program if $L_{off}(N/V, \Phi) = a_L N/V + b_L \Phi + c_L$ and $a_L N/V + b_L \Phi + c_L \leq 0$, a number of descendants is equal 1. So, the function $L_{off}(N/V, \Phi)$ is not linear even though $L_{off}(N/V, \Phi) = a_L N/V + b_L \Phi + c_L$.

If population density and phenotype size changes in time, the linear functions will take values greater than 1 or less than 0 after some generations. This usually results in a rapid extinction of the population or its growth to infinity. To avoid these situations nonlinear formulas for functions $L_{off}(N/V, \Phi)$, $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$ should be used. The probabilities should include on interval $[0, 1]$ and expected value of the offspring size should be positive. The nonlinearity of dependence of the probabilities of reproduction and death (also offspring size) on various environmental factors applies to all object-oriented models of populations and to all real populations.

2.4. Adaptive and competitive natural selection

Only special phenotypes influence all: offspring size, probability of reproduction and probability of death. This article regards the phenotypes that affect only one of these features. Moreover, the assumption was done that

all functions $L_{off}(N/V,\Phi)$, $p_r(N/V,\Phi)$ and $p_s(N/V,\Phi)$ are monotone. The functions $L_{off}(N/V,\Phi)$, $p_r(N/V,\Phi)$ increase if Φ increases and they decrease if N/V increases. The function $p_s(N/V,\Phi)$ decreases if Φ increases and it increases if N/V increase. This means a directional natural selection and typical density-dependent regulatory mechanisms.

All dependences:

$$L_{off}(N_E/V,\bar{\Phi})p_r(N_E/V) = p_s(N_E/V), \tag{3}$$

$$L_{off}(N_E/V)p_r(N_E/V,\bar{\Phi}) = p_s(N_E/V), \tag{4}$$

$$L_{off}(N_E/V)p_r(N_E/V) = p_s(N_E/V,\bar{\Phi}), \tag{5}$$

show that increase of the phenotype size causes increase of N_E/V. Such natural selection is typical for populations that adapt to an environment. It is named adaptive natural selection ([34], [29]).

Many evolutionists imagine natural selection in another way. Between individuals of any generations occurs the struggle for existence. The individuals have different probabilities of death or reproduction, however the total probability does not change in time (Fig.2).

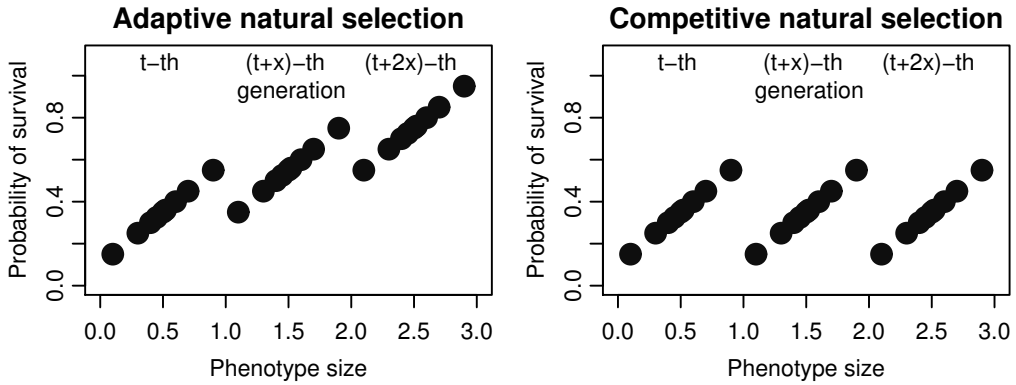

Figure 2: Dependence between phenotype size and probability of survival in successive generations of the populations with adaptive and competitive natural selection.

This natural selection will be called competitive. It can be programmed using functions:
$L_{off}(N/V,\Phi-\bar{\Phi})$ for expected value of offspring size,
$p_r(N/V,\Phi-\bar{\Phi})$ for probability of reproduction,

$p_s(N/V, \Phi - \bar{\Phi})$ for probability of death.
where $\bar{\Phi}$ is mean of the size of the individuals' phenotypes in given generation. Such formulas do not force population growth during simulation.

2.5. Survive and reproductive natural selection

In reality a given phenotype very rarely influences both: the probability of death and the probability of reproduction. So, two kinds of natural selection will be considered: for $b_r = 0$ and $b_s < 0$ (phenotype size influencing mortality), and for $b_r > 0$ and $b_s = 0$ (phenotype size influencing reproduction). The population size fluctuates around the N_E but in populations with adaptive natural selection this number N_E increases during evolution (Fig.3).

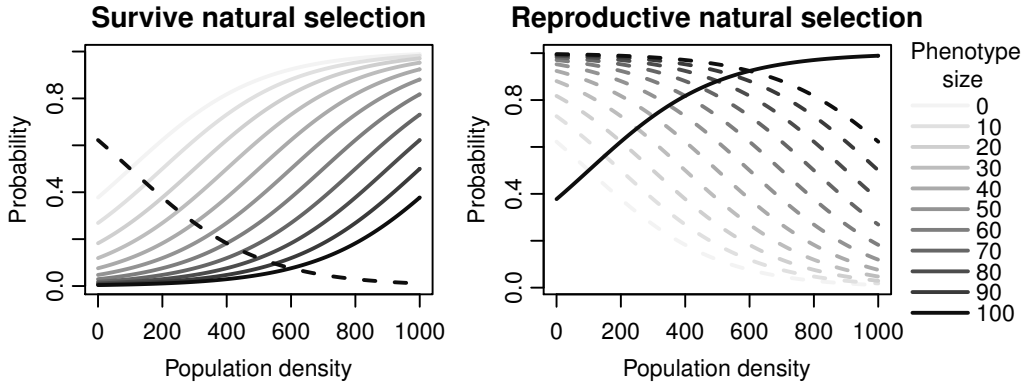

Figure 3: Regulatory mechanisms of modelled populations with adaptive natural selection. Population density fluctuates over the intersection point of the curves: $N \to p_r(N/V, \Phi)$ (dashed line) and $N \to ps(N/V, \Phi)$ (solid line). If the phenotype size Φ increases, then this intersection N_E increases too.

The decrease of probability of death if phenotype size is increased are named survive natural selection in life science ([1]). So, the increase of the probability of reproduction due to increase of the phenotype size will be named reproductive natural selection.

2.6. Used functions and parameters

In order to run the simulation all functions $L_{off}(N/V, \Phi)$, $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$ must be changed to adequate formulas. All non-linear functions of values in interval $[0, 1]$ can be approximated by functions $1/(1+exp(-w(x)))$ where $w(x)$ is some polynomial. All functions of positive values only can

be approximated by functions $log(1 + exp(w(x)))$. Such transformations of polynomials keep their type of monotony on the same intervals and they have the same places of local extremes.

In this article the constant value of 1 of offspring size was assumed because evolution of an offspring size is important issue in biology worthy of a separate article. For modelling adaptive natural selections the following formulas were to be used:

$$p_r(N/V, \Phi) = \frac{1}{1 + \exp(-(a_r N/V + b_r \Phi + c_r))}, \tag{6}$$

$$p_s(N/V, \Phi) = \frac{1}{1 + \exp(-(a_s N/V + b_s \Phi + c_s))}. \tag{7}$$

In models with competitive natural selections following formulas were used:

$$p_r(N/V, \Phi) = \frac{1}{1 + \exp(-(a_r N/V + b_r(\Phi - \bar{\Phi}) + c_r))}, \tag{8}$$

$$p_s(N/V, \Phi) = \frac{1}{1 + \exp(-(a_s N/V + b_s(\Phi - \bar{\Phi}) + c_s))}. \tag{9}$$

The values of parameters a_r, b_r, c_r, a_s, b_s and c_s were selected in such a way that the equilibrium number $N_E(\Phi)$ is positive, where:

$$N_E = -\frac{b_r \bar{\Phi} - b_s \bar{\Phi} + c_r - c_s}{a_r - a_s} V \tag{10}$$

for adaptive natural selection and

$$N_E = -\frac{c_r - c_s}{a_r - a_s} V \tag{11}$$

for competitive natural selection.

Parameters σ_g and σ_e estimate the size of the impact of genotype and non-genotype factors on the phenotype size. The values of these parameters depend on the unit of phenotype size measurement. The ratio $\sigma_g/(\sigma_g+\sigma_e)100\%$ determines the degree to which the variability of genotypes explains the variability of phenotype sizes. The relationship between the rate of evolution and this ratio is quite unexpected and deserves a separate article. In this paper it was assumed that $\sigma_g = 1$ and $\sigma_e = 1$.

Probability of mutation of any part of DNA is very small. It depends on the type of mutation (duplication or reduction of a large part of DNA,

transition, transversion, insertion or deletion) and probably - on the location of this part in a nucleus. But first of all it depends on the length of it. The rates of human genomic mutation have been estimated at $p_{m,nuc} = 2.5 * 10^{-8}$ per nucleotide per generation [25]. Nowadays, using data available from the whole genome sequencing, the human genome mutation rate is estimated to be $p_{m,nuc} = 1.1 * 10^{-8}$ per nucleotide per generation [30]. Average length of the gene in eukaryotes is from $L = 5 * 10^3$ to $L = 50 * 10^3$ nucleotydes [47]. So, probability of mutation of the gene, equal $p_m = p_{n,nuc} * L$, is a number of order 10^{-4}. In this paper the probability of mutation is constant and equals $p_m = 0.0001$.

2.7. Analysis of results of the simulations

Object-oriented computer models allow calculating and analysing many numerical variables. The program simulating population with heredity and natural selection (Appendixes 1 and 2) writes to a text file for each generation: population size, number of mutation that have happened from initial generation, mean phenotype size (with standard deviation) and minimal number of mutations on genes in each generation. In the other text file the genetic structure in each generation can be recorded. The data in text files can be analysed in the same way as data collected in population ecology research.

To illustrate the relationship between molecular and morphological evolution, the frequencies of all alleles in successive generations were recorded. Such detail analyse was done for 10 thousand generations only and for one exemplary simulation.

The morphological evolution was illustrated as the changes of mean phenotype size for successive generations. The molecular evolution was illustrated as minimal number of mutations accumulated on genes at each generation. The minimal number of the mutations on a gene corresponds to a number of accepted alleles by a population during evolution. These alleles arose as a mutation happened on a gene of one distant ancestor of all individuals and replaced previous such alleles.

Each simulation was done for 1000000 generations. For each set of parameters the simulations was repeated 1000 times. The results was elaborated using primary statistical methods (calculation the means and standard deviations at each generation for all repetitions) without testing due to an arbitrarily large number of repetitions of the simulations.

3. Results

3.1. Molecular and morphological evolution

An initial population consisted of 500 individuals living on territory of an area $V = 5$ units. Genotypes of all initial individuals were equal AA where $A = (0, 0, 0)$. The parameters of probability of reproduction and the probability of death had following values: $a_r = -0.005, b_r = 0, c_r = 0.5, a_s = 0.005, b_s = -0.05, c_s = -0.5$. The probability of mutation of one gene was equal to $p_{mut} = 0.0001$. Both parameters σ_g and σ_e were equal 1. One simulation was done. It was carried out for 10 000 generations.

During the simulation the 528 mutations happened, but most of them (137) appeared on one individual, which died at the next generation without descendants. Very often some allele appeared in a few individuals and disappeared. Number of alleles fluctuated from 1 to 4 and for few time periods population was monomorphic. Only two new alleles dominated the whole population and were accepted (Fig.4). In the last generation the genes accumulated two mutations. It was only 0.4% of all mutations.

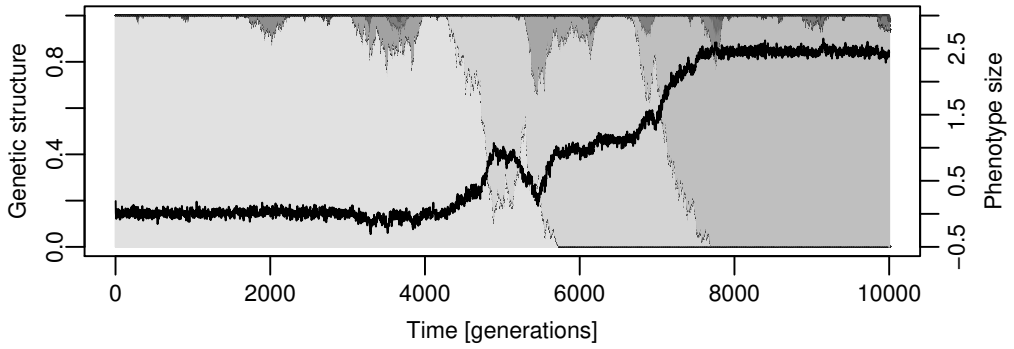

Figure 4: The molecular evolution was illustrated as changes in time the genetic structure of each generation (10000 bars for genetic structures of successive generations where different alleles was shown by different shades of grey colour). Morphological evolution was illustrated as mean phenotype size of individuals of each generation by black curve.

Both accepted alleles caused the increase of the phenotype size of individuals, so they decreased the probability of their death. As result the chart of changes of the phenotype size looked like a stair function, and particular steps were associated with the domination of a specific allele on the genes of individuals. Only in periods of big genetic diversity, their shape was

disturbed. Such images of the molecular and morphological evolution were similar when the simulations were repeated.

3.2. Population size and rate of evolution

Systematic analysis of the impact of population size on the rate of evolution was conducted for 4 variants of natural selection: adaptive and reproductive, adaptive and survive, competitive and reproductive, competitive and survive. The density of all initial populations were equal 100 individuals per area unit. All their individuals had genotype AA where $A = (0, 0, 0)$. The initial probability of reproduction and the probability of death had the same scale of values (around 0.5).

At the beginning all populations were the same. They were different in size only due to the different sizes of their area: from $V = 1$ to $V = 10$. In each case the simulations were conducted for $1,000,000$ generations and repeated 1000 times. Then the mean characteristics of morphological evolution and molecular evolution were calculated.

The morphological evolution was faster in great populations than in small ones for all variants of natural selections (Fig. 5). But dependence $V \rightarrow \bar{\Phi}(t+1) - \bar{\Phi}(t)$ where $\bar{\Phi}(t)$ is the mean phenotype size at t-th generation was nonlinear, convex upward function.

For the initial 50 thousand generations the fastest evolution appeared in populations with survive and adaptive natural selection. But after this time it was suddenly broken and in the last generation these populations had the smalest mean phenotype size. The inhibition of the evolution was less severe in the population with adaptive and reproductive natural selection. In the population of competitive natural selection the rate of evolution was constant (Fig. 5).

The molecular evolution was faster in great populations than in small ones for all variants of natural selections (Fig. 6). The dependencies between evolution rate and kinds of the natural selections were the same as for morphological evolution.

The population density increased during evolution for all variants of populations of adaptive natural selection. For instance, for the variant of bigger initial population ($V = 10$) it grew from $100N/V$ to over $1300N/V$ (reproductive natural selection) and from $100N/V$ to over $1100N/V$ (survive natural selection). But this increase was non-linear: fast at the beginning and slow at the end. Only during initial 50 thousand generations this increase had almost an exponential course.

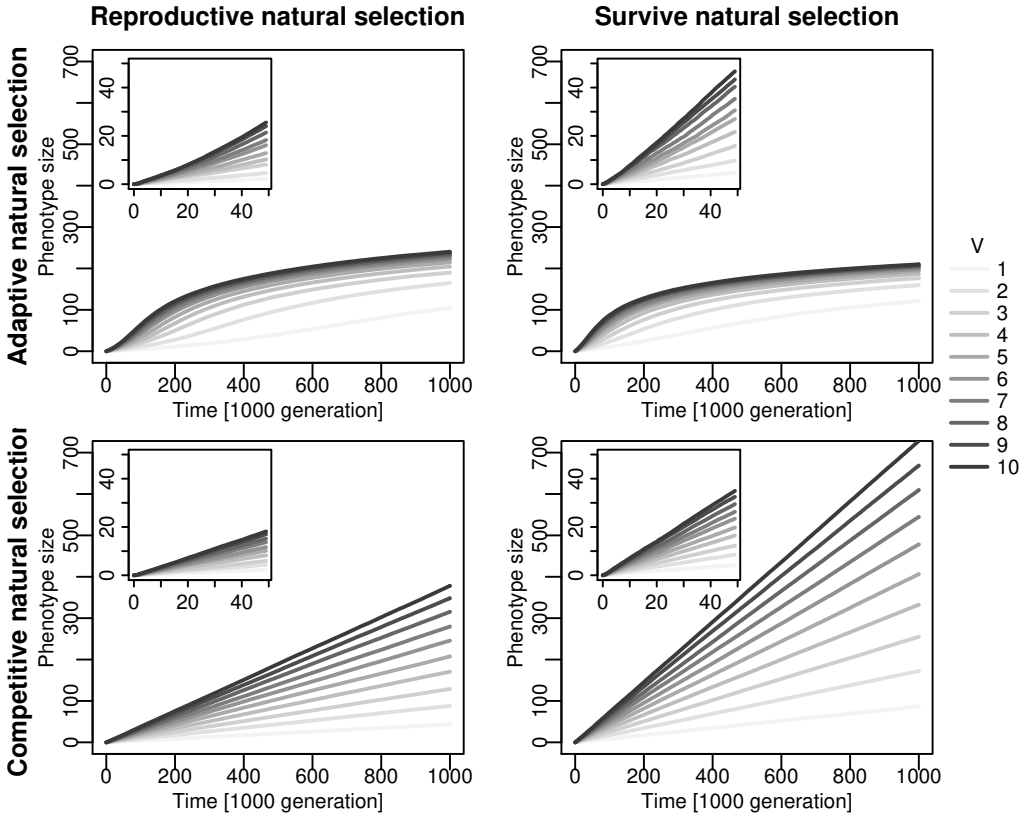

Figure 5: Morphological evolution (changes of the mean phenotype size of individuals in time) for populations living on different area V. On little charts the changes in initial 50 thousand generations are shown. The results for one thousand repetitions of the simulations of the gen-phen model.

4. Discussion

4.1. Gen-phen model and mathematical models of evolution

The object-oriented modeling is not only using an object-oriented approach to software development. It is the reproduction in the program of these assumptions of the model, which are hidden behind a usually short description of the real phenomena that are modelled ([15], [18], [8]). In my model individuals are programmed as simple object-oriented structures having identify number, pair of genes and phenotype. Additionally the genes are programmed as object-oriented structures, although it is not necessary due

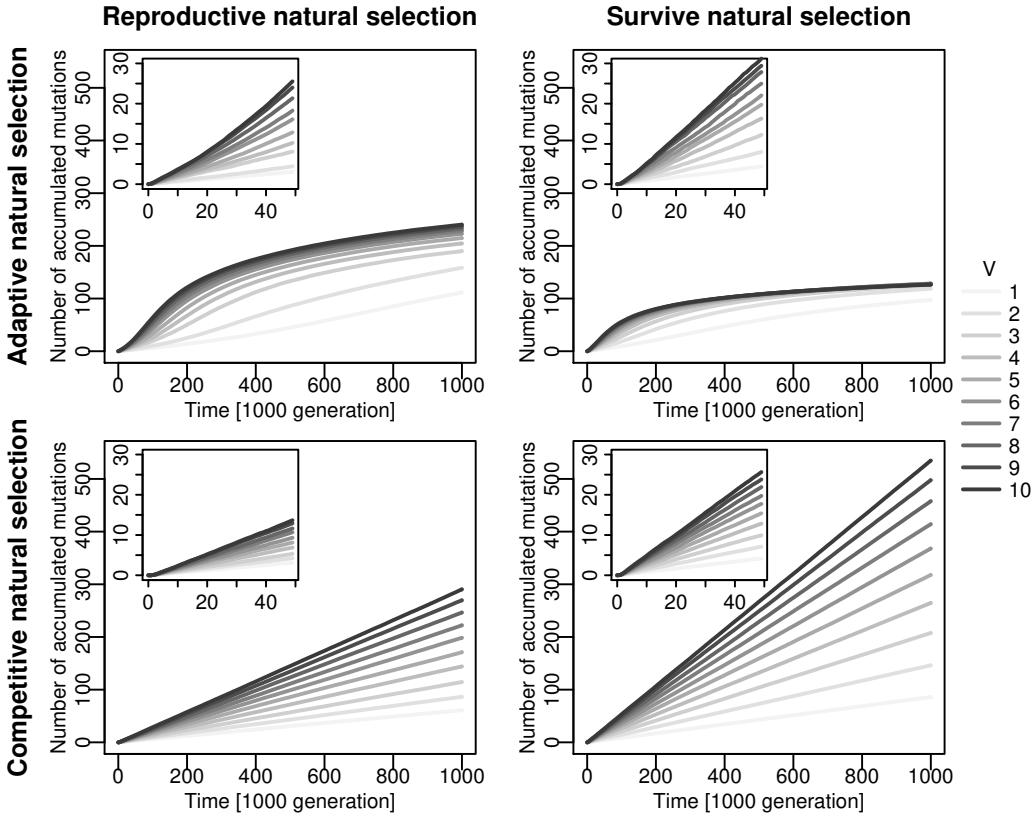

Figure 6: Molecular evolution (changes in minimal number of mutations accumulated on genes in time) for populations living on different area V. On little charts the changes in initial 50 thousand generations are shown. The result of 1000 simulations of the gen-phen model.

to the simple form of the gene (three numbers). But it simplifies to make individuals with more complex genes.

Each individual is born, gives birth to new individuals and dies. These are random events of the probability depending on the individual's phenotype size and state of a population at given time. If $L_{off}(N/V, \Phi) = 1$ and the functions $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$ are dependent on population size only (they are: $p_r(N/V)$ and $p_s(N/V)$), the dynamics of population size formed by the gen-phen model are exactly the same as in Markov chains $(N_t)_{t \in N}$ named as discrete time stochastic model of the population ([41]).

Heredity in gen-phen model is modelled according to Mendel's rules, so changes in genotypes of subsequent generations correspond to the assump-

tions of the Wright-Fisher ([46]) and Moran ([24]) models. If in gen-phen model a probability of mutation $p_{mut} = 0$ and initial population includes a few alleles of the genes then the dynamics of the frequency of each allele is compatible with the genetic drift model. Write-Fisher model with selection arises when $p_s = 1$ and $L_{off}(N/V, \Phi)p_r(N/V, \Phi) = 1$. Moran model with selection arises for $p_s(N/V, \Phi) < 1$. So, the gen-phen model is a generalization of Write-Fisher and Moran models by the use of the new assumptions: the impact of the phenotype size on individuals' fitness, a dependence between genotype and phenotype and possibility of the mutations of the gene on the individual level. Probably this is the simplest object-oriented model of the populations with heredity, mutations, phenotypes and natural selection.

Nowadays classical population models are described using modern language of mathematics ([13]). The gen-phen model should also be developed in a similar way: as multivariable numerical Markov chain because only for numeric Markov chains the expected values and other statistical characteristics can be calculated. But after specify the values of parameters such analysis should give the same results as they did in this article.

4.2. Population size and the rate of evolution

The rate of evolution has been interesting for scientists for many years. For some of them this term means an increased number of species living in a given territory [38], [26]. But in palaeontology it means an increasing or decreasing rate of a selected phenotype size inside the one species [35], [11], [38], [45], [31]. Geneticists define a rate of evolution as a number of mutations (mainly substitutions) in genes in comparison to a common ancestor of related species ([9]). For non-neutral genes (phenes) genetic and palaeontological definitions are mutually dependent. It has been shown using my model.

Haldane [11] has proposed a measure of the rate of evolution for logarithmically transformed data. This rate would be measured in units called "darwins". If the phenotype size increases, then its variation increases and according to Haldane the possibility to receive a larger phenotype by mutation also increases. In the gen-phen model there is a possibility to make an assumption that the variability of the phenotype size increases if the mean value of the phenotype size increases (σ_g or/and σ_e can be functions of Φ). But then the growth of the phenotype size would be inhibited very quickly. Such inhibition of the evolution has been shown without this assumption. Only at the beginning the rate of evolution increases in accordance with

Haldane's hypothesis, but for another reason (the increase of the population size).

Because the genetic drift in small populations is faster than in the big ones [16], [39], [40] many scientists have come to a conclusion that evolution in small populations is faster than in the big ones ([10]), though there are no experimental or palaeontological publications confirming this view to be found. For evolution the vanishing rate of certain alleles does not matter but what is important is the accumulation rate of new alleles on the genes. If the drift rate is different in two populations of the same size, then evolution will be faster in populations with a faster genetic drift. But in a big population the mutations happen more frequently than in small ones and this fact accelerates evolution more than slower genetic drift slows down it.

There are very few mathematical papers analysing the impact of population size on the rate of evolution. I found only one such paper ([36]) from which it also follows that evolution is faster in larger populations and this relationship is not linear. Some experimental research shows, that evolution in bigger populations is faster ([19]). In fields of science in which the term evolution has been extended to the appearance and consolidation of non-genetics-related phenomena (evolution of human behaviour or culture), the phenomenon of an increase in the rate of evolution along with the increase in population size is better researched ([4]). The described relationships in this work confirm this thesis, although they are not proof that for all parameters and types of natural selection, the evolution of a larger population is faster than the smaller one.

In this article only for a few values of parameters has it been shown that evolution is faster in a larger population than in a smaller one. But simulations of the gen-phen model had been done for many different values of the parameters and this regularity was repeated. Only for $b_s = 0$ and $b_r = 0$, e.g. for neutral genes accumulating random mutations, the rate of evolution did not depend on the size of the population.

The increase in the rate of evolution with the increase in the population size is a fact of fundamental importance for the theory of evolution. It can explain why local populations living on islands lose in competition with related species originating from continents although the insular populations have adapted better to the local environment.

4.3. Types of natural selections and a rate of evolution

The sample simulations were done for exactly symmetrical functions for probability of reproduction $1/(1 + exp(-(-0.005N/V + 0.5\Phi + 0.5)))$ and probability of death $1/(1 + exp(-(0.005N/V - 0.5\Phi - 0.5)))$. The probability of survival $1 - p_s(N/V, \Phi)$ is exactly the same as probability of birth new individual $p_r(N/V, \Phi)$. It would seem that the rate of evolution caused by reproductive and survive selection with used parameters will be the same. Meanwhile, the phenotype reducing the probability of death caused twice faster evolution than the phenotype increasing the probability of reproduction.

Responsible for this is the fact that it's omitted in some of the articles on the Moran model with natural selection, namely that the directed genetic drift caused by decrease in mortality of individuals with a given genotype is much faster than the genetic drift caused by increase of the reproduction of individuals with a given genotype (Fig. 7).

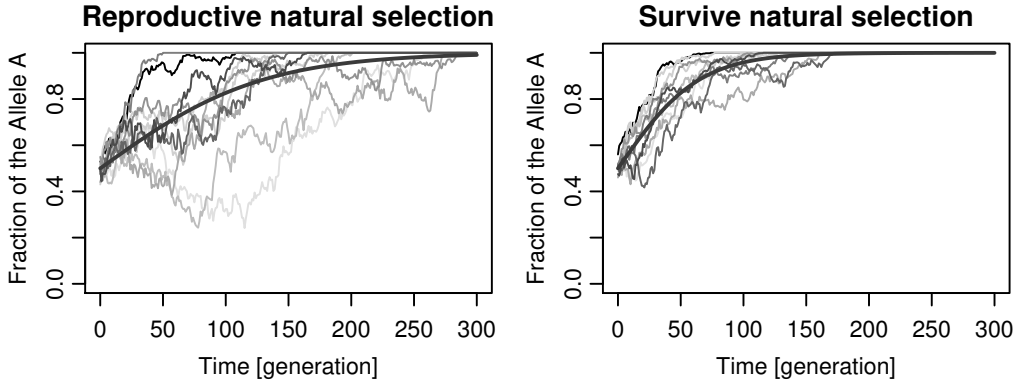

Figure 7: Ten realizations of the directional genetic drift formed by gen-phen model with probability of mutation $p_{mut} = 0$ and initial population consisted $N_0 = 100$ individuals with genes of alleles A and B such that $\Phi(AA) = 5$, $\Phi(AB) = 2.5$ and $\Phi(BB) = 0$. The natural selection caused by $p_r = 1/(1 + exp(-(-0.005N/V + 0.05\Phi + 0.5)))$ (left chart) or $p_s = 1/(1 + exp(-(0.005N/V - 0.05\Phi - 0.5)))$ (right chart). The black, thick curve shows the predicted course of changes in the allele A fraction calculated using the formulas (12) and (13).

Realizations of the Markov chains fluctuate over the conditional expected values forming recursive sequences. For the first population the expected numbers of individuals at time t+1 on condition that at time t it was equal

$N_{AA,t}$, $N_{AB,t}$ and $N_{BB,t}$) are equal:

$$\begin{cases} N_{AA,t+1} = N_{AA,t}(1-p_s) + p_{r,AA}N_{AA,t}x_{A,t} + 0.5p_{r,AB}N_{AB,t}x_{A,t} \\ N_{AB,t+1} = N_{AB,t}(1-p_s) + p_{r,AA}N_{AA,t}x_{B,t} + 0.5p_{r,AB}N_{AB,t} + p_{r,BB}N_{BB,t}x_{A,t} \\ N_{BB,t+1} = N_{BB,t}(1-p_s) + 0.5p_{r,AB}N_{AB,t} + p_{r,BB}N_{BB,t}x_{B,t} \\ x_{A,t+1} = \frac{N_{AA,t+1}+0.5N_{AB,t+1}}{N_{AA,t}+N_{AB,t}+N_{BB,t}} \\ x_{B,t+1} = 1 - x_{A,t+1} \end{cases} \tag{12}$$

where $p_s = p_s(N_t/V)$ is a probability of death of individuals, $p_{r,AA} = p_r(N_t/V, \Phi(AA))$, $p_{r,AB} = p_r(N_t/V, \Phi(AB))$, $p_{r,BB} = p_r(N_t/V, \Phi(BB))$ are probabilities of reproduction of individuals with genotypes AA, AB and BB and with phenotypes $\Phi(AA)$, $\Phi(AB)$, $\Phi(BB)$ respectively.

In populations with the impact of the phenotype size on probability of death the conditional expected numbers of individuals at time t+1 are equal:

$$\begin{cases} N_{AA,t+1} = N_{AA,t}(1-p_{s,AA}) + p_r N_t x_{A,t}^2 \\ N_{AB,t+1} = N_{AB,t}(1-p_{s,AB}) + 2p_r N_t x_{A,t}x_{B,t} \\ N_{BB,t+1} = N_{BB,t}(1-p_{s,BB}) + p_r N_t x_{B,t}^2 \\ N_{t+1} = N_{AA,t+1} + N_{AB,t+1} + N_{BB,t+1} \\ x_{A,t+1} = \frac{N_{AA,t+1}+0.5N_{AB,t+1}}{N_{t+1}} \\ x_{B,t+1} = 1 - x_{A,t+1} \end{cases} \tag{13}$$

where $p_{s,AA} = p_s(N_t/V, \Phi(AA))$, $p_{s,AB} = p_s(N_t/V, \Phi(AB))$, $p_{s,BB} = p_r(N_t/V, \Phi(BB))$ are probabilities of death of individuals with genotypes AA, AB and BB and with phenotypes $\Phi(AA)$, $\Phi(AB)$, $\Phi(BB)$ respectively, $p_r = p_r(N/V)$ is a probability of reproduction.

The impact of phenotype size on the probability of reproduction causes that among the descendants there are many individuals with unfavourable genes, because the choice of father for a descendant's gene has the probability independent of his phenotype size. In virtual populations, just like in real ones, the events: "to become a mother" and "to become a father" are independent of each other and they are usually dependent on other phenotypes. Selection by phenotype size influencing the probability of death is more effective.

The dependence between the rate of evolution and types of natural selection (adaptive or competitive) is an effect of nonlinear formulas of the probability of death and probability of reproduction. In models with adaptive natural selection if the increase of probability of reproduction or death

are almost linear, then the population size linearly increases. But nonlinear functions for probabilities bend when they reach values close to 1 or 0. An intensity of natural selection (that can be calculated as $\frac{\partial p(N/V.\Phi)}{\partial \Phi}(\bar{\Phi})$) decreases to 0, and this is a factor which affects evolution rate the most. As a result the evolution of the population with the adaptive natural selection is faster at the beginning and slower after some generation then evolution of the population with competitive natural selection. Biologists explain this fact as an adaptation of the population to the environment and an increase in the importance of other factors regulating the population than the size of the analysed phenotype.

Competitive natural selection does not have such restrictions. Theoretically, if the phenotype increases the chances of survival or reproduction in comparison to individuals having this phenotype of smaller-size, it can grow to infinity. In practice, the energetic costs of having enlarged parts of the body part may affect the condition of the individuals. But described effect can explain why phenotypes related with intraspecific competition are usually bigger and better visible than phenotypes related with adaptation to the environmental.

5. Conclussions

1. Regardless of the type of directional natural selection, the increase in the size of the population has always caused the increase in the rate of the evolution.
2. An adaptive natural selection caused a non-linear changes in the rate of the evolution. At the beginning of the simulation the evolution were increasing according to the Haldane's hypothesis. After long time this increase was inhibited.
3. A competitive natural selection caused a linear evolution.
4. A survive natural selection caused faster rate of the evolution than a reproductive natural selection.
5. If natural selection did not directly or indirectly affected a gene (neutral gene) then the population size didn't influence the rate of accumulation of mutations on the gene. But this observation done for object-oriented models should be proved with mathematical precision.

6. Author contribution

All the work was done by the author himself.

7. Acknowledgements

The author thanks the reviewers for valuable comments.

8. Data accessibility statement

The data used in the article comes from the simulation of the programs, which are attached as supplementary files.

9. References

References

[1] Battail, G. (2014) Information and life. Springer Science+Business Media Dordrecht, Berlin/Heidelberg, 260 pp.

[2] Bradshaw, W. E., Holzapfel, C. M. (2001) Phenotypic evolution and the genetic architecture underlying photoperiodic time measurement. Journal of Insect Physiology 47 809–820.

[3] Carroll, S. B. (2008) Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell 134: 25–36.

[4] Collard, M.; Buchanan, B.; O'Brien, M. J. (2013) Population Size as an Explanation for Patterns in the Paleolithic Archaeological Record: More Caution Is Needed. Current Anthropology 54 (Supp. 8): S388–S396.

[5] Cooper, D. N.; Kehrer-Sawatzki, H. (Eds.). (2008) Handbook of human molecular evolution. Wiley, Chichester, 2 vol., 1717 pp.

[6] Cordero, F. (2017) Common ancestor type distribution: A Moran model and its deterministic limit. Stochastic Processes and their Applications 127: 590–621

[7] Engen, S.; Lande, R.; Sæther, B.-E. A Quantitative Genetic Model of r- and K-Selection in a Fluctuating Population. The American Naturalist , Vol. 181: 725–736.

[8] Fatmi, M. R.; Habib, M. A. (2018) Microsimulation of life-stage transitions and residential location transitionswithin a life-oriented integrated urban modeling system. Computers, Environment and Urban Systems 69: 87-–103.

[9] Gillespie, J. H.; Langley, C.H. (1979). Are evolutionary rates really variable? Journal of Molecular Evolution 13, 27–34.

[10] Gillespie, J. H. (2001) Is the Population Size of a Species Relevant to Its Evolution? Evolution 55(11): 2161–2169.

[11] Haldane, J. B. S. (1949) Suggestions as to quantitative measurement of rates of evolution. Evolution 3: 51–56.

[12] Harris, L. F.; Sullivan, M. R.; Hatfield, D. L. (1999) Directed Molecular Evolution. Origins of Life and Evolution of Biospheres 29 (4): 425–435.

[13] Hössjer, O.; Tyvand, P. A. (2016) A monoecious and diploid Moran model of random mating. Journal of Theoretical Biology 394: 182–196.

[14] Ichinosea, M.; Iizukab, M.; Kusumic, J.; Takefu, M. (2013) Models of compensatory molecular evolution: Effects of back mutation. Journal of Theoretical Biology 323: 1–10.

[15] Johnson, C. G.; Goldman, J. P.; Gullick, W. J. (2004) Simulating complex intracellular processes using object-oriented computational modelling. Progress in Biophysics & Molecular Biology 86: 379–406.

[16] Kimura, M. (1968) Evolutionary rate at the molecular level. Nature 217 (5129): 624–626.

[17] Kluth, S.; Baake, E. (2013) The Moran model with selection: Fixation probabilities, ancestral lines, and an alternative particle representation. Theoretical Population Biology 90: 104–112.

[18] Kool, J. T. (2009) An object-oriented, individual-based approach for simulating the dynamics of genes in subdivided populations. Ecological Informatics 4: 136–146.

[19] Lachapelle, J.; Reid, J.; Colegrave, N. (2015) Repeatability of adaptation in experimental populations of different sizes. Proceeding B: Biological Sciences 282 (1805): 1-8.

[20] Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. Evolution 30: 314–334.

[21] Larkin, J. C. (2009) Morphological Evolution: By Any Means Necessary? Current Biology 19 (20): R953–R954.

[22] Magal, P.; Webb, G. F. (2000) Mutation, selection, and recombination in a model of phenotype evolution. Discrete and Continuous Dynamical Systems 6: 221–236.

[23] Marugán-Lobón, J.; Buscalioni, A. D. (2006) Avian skull morphological evolution: exploring exo- and endocranial covariation with two-block partial least squares. Zoology 109 (2006) 217–230.

[24] Moran, P. A. P. (1958). A general theory of the distribution of gene frequencies. Overlapping generations. Proceedings Of The Royal Society B, Biological Sciences 149: 102–112.

[25] Nachman, M. W.; Crowell, S. L. (2000) Estimate of the mutation rate per nucleotide in humans. Genetics 156: 297–304.

[26] Pawar, S. S. (2005) Geographical Variation in the Rate of Evolution: Effect of Available Energy or Fluctuating Environment? Evolution 59 (1): 234–237.

[27] Pekalski A. (1999) Mutations and changes of the environment in a model of biological evolution. Physica A 265: 255–263.

[28] Penny, D.; McComish, B. J.; Charleston, M. A. ;Hendy, M. D. (2001) Mathematical Elegance with Biochemical Realism: The Covarion Model of Molecular Evolution. Journal of Molecular Evolution (2001) 53: 711–723.

[29] Rao, V.; Nanjundiah, V. (2017) Haldane's view of natural selection. Journal of Genetics 96(5): 765–772.

[30] Roach, J. C.; Glusman, G.; Smit, A. F., et al. (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328: 636–639.

[31] Roopnarine, P. D. (2003) Analysis of Rates of Morphologic Evolution. Annual Review of Ecology, Evolution, and Systematics, 34: 605–632

[32] Seligmann, H. (2010) Positive correlations between molecular and morphological rates of evolution. Journal of Theoretical Biology, 264: 799–807.

[33] Schraiber, J. G. (2014) A path integral formulation of the Wright–Fisher process with genic selection. Theoretical Population Biology 92: 30–35.

[34] Schrider, D. R.; Hourmozdi, J. N.; Hahn, M. W. (2011) Pervasive Multinucleotide Mutational Events in Eukaryotes. Current Biology 21: 1051–1054.

[35] Simpson, G. G. (1953) The major features of evolution. Columbia Univ. Press, New York. 434 pp.

[36] Smith, M. J. (1976) What determines the rate of evolution? The American Naturalist, 110: 331–338.

[37] Sokół, M. (2013) Metody modelowania populacji [PL, Methods of modelling of the population]. Wydawnictwo Naukowe PWN. 534 pp.

[38] Stanley, S. M. (1985) Rates of Evolution. Paleobiology, 11 (1): 13–26.

[39] Tomiuk, J.; Guldbrandtsen, B.; Loeschcke, V. (1998) Population differentiation through mutation and drift – a comparison of genetic identity measures. Genetica 102/103: 545–558.

[40] Travis, J. M. J.; Ezard, T. H. G. (2006) Habitat geometry, population viscosity and the rate of genetic drift. Ecological Informatics 1: 153–161.

[41] Tuljapurkar, S.; Steinsaltz, D. (2019) Stochastic Models for Structured Populations. Handbook of Statistics 40: 133–155.

[42] Versieux, L. M., Barbará, T., Wanderley, M. G. L., Calvente, A., Fayb, M. F., Lexer, C. (2012) Molecular phylogenetics of the Brazilian giant bromeliads (Alcantarea, Bromeliaceae): implications for morphological evolution and biogeography. Molecular Phylogenetics and Evolution 64: 177–189.

[43] Vogl, C.; Clemente, F. (2012) The allele-frequency spectrum in a decoupled Moran model with mutation, drift, and directional selection, assuming small mutation rates. Theoretical Population Biology 81: 197–209.

[44] Wang, X.-J.; Gu C.-L.; Quan, J. (2019) Evolutionary game dynamics of the Wright-Fisher process with different selection intensities. Journal of Theoretical Biology 465: 17–26.

[45] Wray, G. A. (1992) Rates of Evolution in Developmental Processes. American Zoologist 32: 123–134.

[46] Wright, S. (1931) Evolution in Mendelian populations. Genetics 16: 97–159.

[47] Xu, L.; Chen, H.; Hu, X.; Zhang, R.; Zhang, Z.; Luo, Z. W. (2006) Average Gene Length Is Highly Conserved in Prokaryotes and Eukaryotes and Diverges Only Between the Two Kingdoms. Mol. Biol. Evol. 23 (6): 1107–1108.

[48] Zamudio G. S.; Prosdocimi, F.; de Farias. S. T.; Joséa, M. V. (2019) A neutral evolution test derived from a theoretical amino acid substitution model. Journal of Theoretical Biology 467: 31–38.