A phenotype-genotype model of a population and observation of the morphological and molecular evolution due to intensity of natural selection

Mirosława Sokol
Faculty of Biology, University of Warsaw, Poland.
milka@biol.uw.edu.pl

Abstract
A phenotype-genotype model of a population is a natural generalization of a simple model of the population (a birth-death model in discrete time) with Mendelian model of heredity including:
- the possibility of arising of new alleles by mutation,
- the influence of genotype on phenotype,
- the influence of phenotype on fitness (probability of birth, probability of survival, the number of offspring of one individual born at the same time).

These assumptions define the populations with various kinds of natural selection. The evolution of such a population appears alone.

In this paper a phenotype-genotype model of a population was used to analyse the rate of the evolution: the phenotype variation amounts 100 generations. This rate depends on the probability of mutation, the size of the influence of genotype on phenotype, the intensity of natural selection, the population numbers and the type of regulatory mechanisms in the population.

The most interesting result of these observations is a non-linear course of phenotypic changes during evolution. The alleles increasing the fitness displace worse alleles and the population size increases. If the population size increases, then the rate of evolution increases but at the same time the size of the influence of phenotype on fitness decreases. If the intensity of natural selection decreases, the rate of evolution decreases. The rate of evolution can increase to some optimum and then decrease. As a result the curves showing the size of phenotype and number of mutation on genes change during the evolution resemble the logistic or/and logarithmic curves.

Keywords: phen-gen model, morphological evolution, molecular evolution, intensity of natural selection, regulatory mechanisms.
1 Introduction

In paleobiology, ecology and evolutionism morphological or phenotypic evolution is defined as an increase, rarely a decrease of the phenotype of individuals of a given population from generation to generation (Marugán-Lobón & Buscalioni 2006, Larkin 2009, Versieux et al. 2012). This is not a speciation or the increase of complexity of a system. Nowadays evolution also means the accumulation of mutations on individuals’ genes, the formation of new alleles and new proteins in cells. For this phenomenon the term “molecular evolution” or “genetic evolution” was used (Harris et al. 1999, Cooper & Kehrer-Sawatzki 2008). Both processes were difficult to compare due to entirely different methods of measuring a phenotype and genetic changes of individuals in time. But for scientists molecular evolution influences the morphological one and the rates of both evolutions can be compared (Bradshaw & Holzapfel 2001, Carrol 2008, Seligmann 2010).

Morphological and molecular evolutions are modelled separately according to the idea that each numerical variable needs to have its own model. But for models of evolution such an idea does not work too well. The driving force of evolution is natural selection, and it can be defined only for the phenotypes of individuals. Phenotypes depend primarily on genotypes. Genotypes can only undergo mutations affecting the phenotype changes. In this paper a phenotype-genotype model of population was presented. It takes into account these relationships in the simplest way possible.

The primary aim of this paper is to present the basic form of phenotype-genotype model (phen-gen model) of a population. Morphological and molecular evolution of this population occurs spontaneously. Another aim of this study is to use the presented model to analyse the rate of evolution. Among many factors influencing the rate of evolution only the intensity of natural selection were analysed. At the end the new and unexpected results were discussed.

2 Methods

2.1 The model

The population was programmed on computer as a finite set of virtual objects (individuals) that appeared as a result of “breeding” of the existing individuals, survived some time-steps, sometimes reproduced (a random event called “reproduction”) and disappeared (a random event called “death”). Each individual has two genes and each descendant inherits one gene from a mother individual and the other gene from a randomly selected individual.
Sokol, M. 2014. A phenotype-genotype model of the population and observation of the morphological and molecular evolution due to intensity of natural selection. Faculty of Biology, University of Warsaw. On-line: www.biol.uw.edu.pl/informatyka/phengenmodel

(father). Inheritance involves copying genes and sometimes during copying a parental gene mutation with probability $p_{\text{mut}}$ happens. In this model each mutation gives a new allele of the gene.

Such assumptions were presented in all models describing evolution. However, they do not allow modelling natural selection. Natural selection can be defined only for phenotypes. So in the phen-gen model, each individual has a phenotype – a real number which can be identified with one of different values of measurement (individual mass, length of its legs, the rate of escape, etc.). It is not the whole shape or all physiological features of individuals. The phenotype is formed as a result of the inherited genotype and different random features. In the model it is a random number $\Phi$ drawn from the normal distribution $\mathcal{N}\left(\frac{x_1+x_2}{2}, \sigma_e\right)$ where $x_1$ is a phenotypic value of the mother gene, $x_2$ is phenotypic value of the father gene. These numbers describe the functional mode of the genotype. The number $x_1 = \frac{x_1+x_1}{2}$ is a mean phenotype of all individuals of the same homozygous genotype.

The parameter $\sigma_e$ describes the size of the influence of non-genetic (i.e. environmental) factors on the phenotype. The genes of virtual individual are three numbers $(x,y,z)$, where $x$ is a real number (after mutation this number can be decreased or increased by a random number drawn from the normal distribution $\mathcal{N}(0, \sigma_y)$); $y$ is an integral number – allele’s name (after mutation this number is changed to a new one, which has not been used yet); $z$ is an integral number – the number of mutations on the gene relative to the ancestor from the initial population (after mutation this number was increased by 1). The parameter $\sigma_g$ describes the size of the effect of one mutation on the functional mode of a genotype.

The probability of reproduction of an individual in a time-step depends on its phenotype and population density (classical regulatory mechanisms). For $N$ – population numbers and $V$ – area of population territory it was calculated with the function $p_r(N/V, \Phi)$ of values in interval $[0, 1]$. The offspring size (litter size, clutch size, number of descendants of one individuals produced at the same time) was a random variable drawn from some discrete distribution $P(\eta = L)$, whose expected number $\lambda$ was calculated with the function $L_{\text{off}}(N/V, \Phi)$ of values in interval $[1, \infty)$. The probability of death of an individual in a time step depends on its phenotype and population density and is calculated with the function $p_d(N/V, \Phi)$ of values in interval $[0, 1]$.

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

Fig. 1. The algorithm of the programme simulating the virtual population with mutations of the genotype, affecting of the genotype on the phenotype and influencing of the phenotype on fitness (probability of death or reproduction).

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 \( \sigma_g \) and \( \sigma_e \). If \( \Phi \) is constant and \( L_{off}(0, \Phi)p_r(0, \Phi) > p_s(0, \Phi) \), and \( L_{off}(N/V, \Phi)p_r(N/V, \Phi) < p_s(N/V, \Phi) \) for all great
N, then the population stabilizes its own numbers as a fluctuation around the number $N_E$ (equilibrium numbers), such that:

$$L_{\text{off}}\left(\frac{N_E}{V}, \Phi \right) p_r\left(\frac{N_E}{V}, \Phi \right) = p_s\left(\frac{N_E}{V}, \Phi \right).$$

These functions $L_{\text{off}}(N/V, \Phi)$, $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$ create a significant regulatory mechanisms in populations. Populations with a significant regulatory mechanisms can subsist for a very long time.

2.2 Used functions and parameters

In order to run the simulation the functions $L_{\text{off}}(N/V, \Phi)$, $p_r(N/V, \Phi)$ and $p_s(N/V, \Phi)$ must be changed to adequate formulas. In this paper monotone functions $\Phi \rightarrow p_r(N/V, \Phi)$ and $\Phi \rightarrow p_s(N/V, \Phi)$, and constant $L_{\text{off}}(N/V, \Phi)$ are considered. So the directional natural selection and selected kinds of influence of phenotype on fitness were considered. The following formulas were used:

$$L_{\text{off}}\left(\frac{N}{V}, \Phi \right) = 1,$$

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

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

In this paper the values of parameters of these functions were selected in such a way that for a given $\Phi$ there exists only one positive equilibrium number $N_E(\Phi)$.

In reality a given phenotype influences very rarely both: the probability of death and the probability of reproduction. So in this paper such kinds of natural selection were considered separately: for $b_r>0$ and $b_s=0$ (phenotypes influencing reproduction) and for $b_r=0$ and $b_s<0$ (phenotypes influencing mortality) (Fig.2). For each kind two different regulatory mechanisms were compared. The first one such that all functions $N/V \rightarrow p_r(N/V, \Phi)$ are decreasing, and the second one such that all functions $N/V \rightarrow p_r(N/V, \Phi)$ are constant. The functions $N \rightarrow P_s(N/V, \Phi)$ increased for both (Fig.2).
Parameters $\sigma_g$ and $\sigma_e$ estimate the size of the influence of mutation on the functional mode of genotype and the size of influence of non-genotype factors on phenotype respectively. The values of these parameters depend on the unit of phenotype measurement. The ratio $\frac{\sigma_g}{\sigma_g+\sigma_e} \times 100\%$ determines the degree to which the variability of genotypes explains the variability of phenotypes. 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 $\sigma_e=1$.

The probability of DNA mutation 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 locality of this part in a nucleus. But first of all it depends on the length of this part of DNA. The total probability of mutation exists for each part of DNA and its value can probably be estimated in the future. The effects of increasing of the probability of gene mutation (i.e. gene length increasing) are easy to predict. If the probability
of mutation increases, then the rate of evolution increases. So in this paper the probability of mutation was a constant and equal to \( p_{\text{mut}} = 0.0001 \). The intensity of natural selection is a measure of the influence of the size of phenotype on fitness. It can be measured as a regression slope between the size of phenotype in any generation and the probability of reproduction (Fig.3). The same definition can be used for probability of survival. This slope can be calculated as a derivative of the function \( \Phi \rightarrow p_r(N/V, \Phi) \) or \( \Phi \rightarrow 1 - p_s(N/V, \Phi) \) at point \( \Phi \) equal to the mean value of phenotypes of individuals from a given generation.

![Fig.3. Idea of calculating the intensity of natural selection in each generation of population – it is a slope of tangent to non-linear regression between phenotypes and probabilities (of reproduction or survival) at mean phenotype.](image)

To sum up, different variables which have biological interpretations can be calculated by the formulas:

- fitness = the probability of reproduction during a time-step, \( p_r(N/V, \Phi) \) or the probability of survival a time-step, \( 1 - p_s(N/V, \Phi) \),
- intensity of natural selection = \( \frac{\partial}{\partial \Phi} p_r(N/V, \Phi) \) for a phenotype influencing the probability of reproduction or = \( -\frac{\partial}{\partial \Phi} p_s(N/V, \Phi) \) for a phenotype influencing the probability of survival,
- equilibrium number of population \( N_E = \frac{b_r \Phi + c_r - b_s \Phi - c_s}{a_r - a_s} V \).
2.3 The analysis of the results of simulations

Object-oriented computer models allow to calculate and analyse many numerical variables. In the programme shown in the appendix we calculated: the size of the population at each time-step, the number of mutations that happened from the start (t=0) to time t, the mean phenotype \( \Phi \) from \( t \) - generation and the mean number of mutations on genes of individuals from \( t \)-generation relative to the ancestors from the initial population. All listed variables were written in a text-file.

For each set of parameter values the simulations were done for time equal 1 000 000 time-steps and repeated over 500 times. The results was elaborated using primary statistical methods without testing due to an arbitrarily large number of repetitions of the simulations.

The rate of evolution was calculated after repeating the simulation for \( \Phi_t \) (the mean of the mean phenotype in \( t \)-th generation in populations) as the function:

\[
t \to \Phi_{t+1000} - \Phi_t.
\]

Logarithmic transformation of phenotype values was not used.

3 The results

All initial populations consisted of 100 individuals of genotype AA where A=(0,0,0) living on a territory of area equal to one unit. This area did not change during the simulation. The probability of reproduction and the probability of death had the same value: 0.5. The probabilities of mutation of one gene were equal to \( p_{\text{mut}}=0.0001 \). Only the intensity of natural selection changed from 0.005 to 0.05 for adequate parameters \( b_r \) and \( b_s \) (see legends on Fig. 4 - Fig. 9).

During the simulation the value of the phenotype increased (Fig. 4) and the density of the population increased (Fig. 5). For populations presented in Fig.4A, Fig.4B, Fig.4C the increase of parameter \( b_r \) and the decrease of parameter \( b_s \) caused the increasing of phenotype changes rate in the first part of evolution. After some generations evolution became slower and this change of the rate of evolution became faster and stronger for high values of \( b_r \) and low values of \( b_s \). As a result at the end for the highest values of \( b_r \) and the lowest values of \( b_s \) the phenotype changes were the lowest. For the populations with regulatory mechanisms given by functions: \( p_r = 0.5 \) and \( p_s = \frac{1}{1+\exp(-(0.01N+b_s\Phi-1))} \) the increases of phenotype of individuals are almost linear (due to the fact that they are realisations of stochastic processes).
The described changes were accompanied by a nonlinear growth of population density (Fig. 5A, Fig. 5B, Fig. 5C). For phenotypes influencing the probability of reproduction the increase of density was inhibited when the probability of reproduction and the probability of death were equal to almost 1. Then in each time-step each individual had a descendant and
died. The populations which have the second regulatory mechanism and phenotype influencing the probability of death were characterized by the density dynamics which have the increasing fluctuations around the increasing linear curves (Fig. 5D).

The density increase was a result of a change of the probability of reproduction and the probability of death (Fig.6). These probabilities have similar values due to high speed return to the equilibrium number (higher than the rate of change of a mean phenotype in population). For the second type of regulation the probabilities of death were characterized by larger fluctuations than the probabilities of reproduction. It is very well visible on Fig.6D, where $p_r = 0.5$ was constant and $p_s$ fluctuated from 0 to 1. The increase of the phenotype and population density of successive generations caused a decrease of intensity of natural selection. They remained positive but they did not differ much from the zero at the end.
A phenotype-genotype model of the population and observation of the morphological and molecular evolution due to intensity of natural selection. Faculty of Biology, University of Warsaw. On-line: www.biol.uw.edu.pl/informatyka/phengenmodel

The rate of morphological evolution for a single simulation was characterized by a great volatility. For the mean phenotype calculated after 1000 repetitions of simulation of each variant clear rules of the evolution rate could be seen (Fig. 7). First, this rate increased, reached the extreme and decreased to zero.

The total number of mutations during evolution can be estimated by the formula:

$$N_{mut} = \sum_{t=0}^{1000000} 2 \cdot 0.0001 p_r(N_t, \Phi_t)N_t,$$

where $N_t$ is the population number at time-step $t$ and $\Phi_t$ is a mean phenotype in the $t$-th generation. For the analysed populations this number varied from 2829 to 1698730 (Tab.1). But only a little part of them (from 0.04% to 1.81%, Tab.1) retained on the genes of the final generation. Changes in time of the mean numbers of mutations on the genes of individuals were only a little similar to the mean phenotype changes (Fig.8 vs Fig.4). The rate of molecular evolution was calculated after 1000 repetitions of simulation of each variant (Fig.9). This rate is very similar to the rate of morphological evolution (Fig.9 vs Fig.7).

The number of mutations accumulated on genes always increased, even if the evolution of the phenotype was inhibited. The populations of the same kind of regulatory mechanism and kind of evolving phenotype can be treated as populations derived from the common ancestor. After 1000000 generations the correlation between the mean number of mutations on the genes (a measure of molecular evolution rate) and the mean phenotype (a
measure of morphological evolution rate) were positive and very high (Tab.2). This result justifies the use of this method to compare the morphological and molecular evolution (Seligmann 2010).

Fig. 7. Rate of morphological evolution (changes in mean phenotype per 1000 generations) for different parameters $b_r$ and $b_s$. Means after 1000 repetition of simulation of phen-gen model.

Fig. 8. Molecular evolution (changes of mean number of mutations on genes in given generation) for different parameters $b_r$ and $b_s$. Means after 1000 repetitions of simulation of phen-gen model.
Tab.1. Numbers of mutations during 1000 000 generations and mean number of mutations on the genes of individuals in 1000 000th generation.

<table>
<thead>
<tr>
<th>Phenotype influencing on probability of reproduction</th>
<th>First type of regulation</th>
<th>Second type of regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total number of mutations</td>
<td>Mean number of mutations on genes</td>
</tr>
<tr>
<td>b_r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>11976</td>
<td>62.0 ± 0.0</td>
</tr>
<tr>
<td>0.02</td>
<td>24402</td>
<td>92.0 ± 0.0</td>
</tr>
<tr>
<td>0.04</td>
<td>86717</td>
<td>225.0 ± 1.0</td>
</tr>
<tr>
<td>0.06</td>
<td>242822</td>
<td>238.4 ± 5.4</td>
</tr>
<tr>
<td>0.08</td>
<td>323724</td>
<td>230.7 ± 5.4</td>
</tr>
<tr>
<td>0.1</td>
<td>388415</td>
<td>214.2 ± 6.7</td>
</tr>
<tr>
<td>0.12</td>
<td>440498</td>
<td>206.7 ± 8.4</td>
</tr>
<tr>
<td>0.14</td>
<td>456857</td>
<td>191.0 ± 9.6</td>
</tr>
<tr>
<td>0.16</td>
<td>471937</td>
<td>188.2 ± 11.7</td>
</tr>
<tr>
<td>0.18</td>
<td>477475</td>
<td>172.4 ± 9.7</td>
</tr>
<tr>
<td>0.2</td>
<td>478374</td>
<td>177.4 ± 10.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phenotype influencing on probability of death</th>
<th>First type of regulation</th>
<th>Second type of regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total number of mutations</td>
<td>Mean number of mutations on genes</td>
</tr>
<tr>
<td>b_s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0</td>
<td>11899</td>
<td>64.0 ± 0.0</td>
</tr>
<tr>
<td>-0.02</td>
<td>15000</td>
<td>58.0 ± 0.0</td>
</tr>
<tr>
<td>-0.04</td>
<td>18368</td>
<td>80.0 ± 0.1</td>
</tr>
<tr>
<td>-0.06</td>
<td>13138</td>
<td>87.0 ± 0.0</td>
</tr>
<tr>
<td>-0.08</td>
<td>11732</td>
<td>79.0 ± 0.0</td>
</tr>
<tr>
<td>-0.1</td>
<td>8145</td>
<td>72.0 ± 0.0</td>
</tr>
<tr>
<td>-0.12</td>
<td>6515</td>
<td>67.0 ± 0.0</td>
</tr>
<tr>
<td>-0.14</td>
<td>5790</td>
<td>61.0 ± 0.0</td>
</tr>
<tr>
<td>-0.16</td>
<td>3950</td>
<td>49.0 ± 0.0</td>
</tr>
<tr>
<td>-0.18</td>
<td>4089</td>
<td>51.9 ± 0.3</td>
</tr>
<tr>
<td>-0.2</td>
<td>2829</td>
<td>51.3 ± 0.4</td>
</tr>
</tbody>
</table>

Tab.2. Correlation coefficients between characteristics of morphological evolutions (mean phenotypes) and molecular evolution (mean number of mutations on genes) in the 10000000th generations derived from the common ancestor. The second correlation (for n=9) was calculated with omitting the cases: b_r=0 and b_s=0.02 or b_r=0 and b_s=-0.02.

<table>
<thead>
<tr>
<th>Phenotype influencing on probability of reproduction</th>
<th>First type of regulation</th>
<th>Second type of regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenotype influencing on probability of reproduction</td>
<td>0.950 (n=11)</td>
<td>0.907 (n=11)</td>
</tr>
<tr>
<td></td>
<td><strong>0.987 (n=9)</strong></td>
<td><strong>0.973 (n=9)</strong></td>
</tr>
<tr>
<td>Phenotype influencing on probability of death</td>
<td>0.511 (n=11)</td>
<td>0.996 (n=11)</td>
</tr>
<tr>
<td></td>
<td><strong>0.951 (n=9)</strong></td>
<td><strong>0.951 (n=9)</strong></td>
</tr>
</tbody>
</table>
4 Discussion

The modern definition of a gene is “a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions, and or other functional sequence regions” (Pearson 2006, Pennisi 2007). Such a gene is not a segment of DNA beginning from a promoter and ending at a terminator, which is transcribed to the same type of RNA. It can contain a number of such segments and a significant part of noncoding DNA. Such a definition of a gene that allows to link a genotype with a phenotype was used in this study.

Although the need to distinguish between a genotype and a phenotype of an individual in the analysis of inheritance is very old (Johannsen 1910), its use in models of inheritance caused difficulties. This was due to a very general descriptive definition of the phenotype, which included all morphological and physiological characteristics of the whole organism and its behaviour. In this study the term phenotype was limited to one numerical variable, which may be the result of measurement of selected morphological, physiological or behavioural
characteristics of an individual. The association of such defined phenotype with specific genes is difficult. Almost all DNA is responsible for the numerical values of many phenotypes, but to a different extent. For each part of DNA we can find a percent in so far as this part of the DNA was responsible for the value of the phenotype. Using this assumption we can choose a phenotype (which we can measure) and choose any part of DNA (which we can mutate) and calculate the standard deviation of the phenotype values of all individuals having one mutation on this part of DNA. In the phen-gen model such a standard deviation was noted as $\sigma_g$.

The shaping of phenotype of individual until its sexual maturity is a long process. For many phenotypes (which are measuring variables: length, volume, speed, etc.) this process is only partially explained or not explained at all. But we know for sure, that the value of a phenotype is affected by an individual’s genotype and the environmental factors in all their variability during the development of the individual. The explanation of the phenotype variation is complicated by the phenotypic plasticity (Price et al. 2003, Kelly et al. 2012). The simplification used in this study – determining the finished phenotype as a random number from a normal distribution $N(\mu, \sigma_e)$ may take into account the additional impact on the phenotype of only environmental factors or both environmental factors and omitted genes. The value of the parameter $\sigma_e$ shows the extent to which non-genetic factors can influence the phenotype. Both parameters $\sigma_g$ and $\sigma_e$ are calculated in the same units as the phenotype. It allows their calculation using experimental methods in the future.

The described assumptions are a basis of a phen-gen model. Other assumptions concerning reproduction, heredity and mortality are the same as in all models of population. The used method of modelling – programming individuals as virtual objects – is well known in computer science literature. It is used in the field of computer science called artificial life modelling (Adamatzky and Komosinski 2005, Parisi and Schlesinger 2002, Bedau 2003). During the simulation such individuals exist as materialistic objects. At any time they have a location in the operating memory – they have a size and weight (calculated as the weight of the part of the operating memory or the weight of electrons used for recording the information about individuals). The models described in this work compared with artificial life models are extremely simple. The individuals do not have artificial intelligence, shape, organs allowing movements. But thanks to the simplification the simulations of populations counting several hundred individuals are very fast. It is possible to
observe evolution over millions of generations, and to repeat the simulations. It is possible to analyse the evolution rate depending on different population factors.

This article is the first presentation of phen-gen model in English. So the full source code of this model in C++ language is presented in the appendix. It was a basic model, which can be complicated. It is possible to programme individuals containing multiple genes affecting one phenotype, to consider several phenotypes affecting one of the following: the probability of reproduction, the probability of death or the offspring size. It is possible to add some individual factors (for instance: sex, age) and complicate the regulatory mechanisms in the population. The programme presented in the appendix can be completed with instructions allowing the calculation of number of alleles and the genetic structure of the population at each generation. It does not change the model but allows additional analysis.

For the basic phen-gen model we received a lot of numerical variables as a result (which typical for artificial life modelling). It allows to compare many values which have a biological interpretation.

A very important characteristic of natural selection is its intensity (Fig.3). The idea of measurement of natural selection is not new (Beeton & Pearson 1901, Haldane 1954, Van Valen 1963, Marcus 1964, O’Donald 1973, Lande & Arnold 1983, Demetrius et al. 2007, Portillo 2012). However, the measures described in literature are more complicated than the ones presented in this paper. After studying different definitions of natural selection the following conclusion was drawn: if all individuals in any generation have the same probability of reproduction and the same probability of death and they have the same offspring size, then there is no natural selection in this generation. Therefore, at least one of these factors must depend on individual’s phenotype. The proposed method of measuring natural selection (Fig.3) is the simplest of all possible. Additionally, the intensity of natural selection can be calculated with the use of field studies and experiments. Such research would consist of determining the number of reproductive and dead animals in a particular time period at different population densities. To calculate the functions \( p_r(N/V, \Phi) \) and \( p_s(N/V, \Phi) \) the non-linear regression method can be used. At the end the adequate partial derivative must be calculated.

The intensity of natural selection affected the rate of evolution very strongly. But there were no constant values during evolution. This is due to the fact that the probabilities take values in the interval \([0,1]\) and any function \( \Phi \to p_r(N/V, \Phi) \) and \( \Phi \to p_s(N/V, \Phi) \) cannot be linear. The increase of population density during evolution is the consequence of the fact that
successive generations have a lower probability of death or/and greater probability of reproduction than their ancestors.

\[
\text{Fig. 10. Idea of calculating the influence size of phenotype on fitness in each generation of population.}
\]

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. This means that the dependence of the probability of reproduction and death on an individual phenotype is not the same function in successive generations (Fig. 10). However, this applies only to those phenotypes, which are used to chase away neighbours from environmental resources or from sexual partners. Such a natural selection can be programmed in a phen-gen model. Two different functions \( p_{r,\text{tot}}(N/V, \Phi) \) and \( p_{r,\text{ind}}(N/V, \Phi) \) could be assumed and the probability of reproduction of any individual would be calculated with \( xp_{r,\text{ind}}(N/V, \Phi) \) for such \( x \) that for any generation \( \frac{1}{N} \sum_{i=1}^{N} xp_{r,\text{ind}}(N/V, \Phi_i) = p_{r,\text{tot}}(N/V, \frac{1}{N} \sum_{i=1}^{N} \Phi_i) \). If in such a model \( p_{r,\text{tot}}(N/V, \Phi) \) does not depend on the phenotype \( \left( \frac{\partial}{\partial \Phi} p_{r,\text{tot}}(N/V, \Phi) = 0 \right) \) then the increase of the phenotype is linear and the density of population is fluctuating around the constant \( N_f/V \).

In almost all models of morphological evolution the independence of population number/density on the modelled phenotype was done as assumptions (Magal and Webb 2000, Magal 2002, Raoul 2011). So they are models of the evolution only of a specific type of phenotypes. Only using evolutionary algorithms (Storn and Price 1995) the complexity network between the population size and evolution (Teo 2006) were noticed.

The above considerations and the results of this study show that natural selection cannot be perceived unilaterally. We can distinguish an adaptive and competitive natural
selection. For many phenotypes natural selection can be partially adaptive, partially competitive with domination of one of these types of natural selection. Additionally, each of these natural selections should be considered separately if the phenotype influences the probability of reproduction and if it influences the probability of death (and also if it influences the offspring size). All kinds of natural selection are programmed in another way and they give different results.

In a phen-gen model both processes: the increase of the phenotype value and the accumulation of mutations on individuals’ genes arise spontaneously. Because the model can be only complicated, it is a special case of all more complicated models (created by resetting certain parameters). Therefore, this model is a proof that each system which consists of a finite number of objects which:
- are born
- have genes and their genes influence an individual phenotype
- the phenotype influences individual fitness
- bear descendants
- copy genes (their own or other individual’s) to the descendants
- the mutation of the gene before or during this copying is possible
- die

can evolve. This evolution can run at a very different rate.

The rate of evolution has been interesting for scientists for many years. For some of them this term means an increased numbers of species living on a given territory (Stanley 1985, Pawar 2005). But in paleobiology it means an increasing or decreasing rate of a selected phenotype of individuals including to the same species (Simpson 1944, Haldane 1949, Stanley 1979, Wray 1992, Roopnarine 2003). Haldane (1949) has proposed a measure of the rate of evolution for logarithmically transformed data. This rate of evolution is measured in units called “darwins”. His idea is correct. If a phenotype increases, then its variation increases and the possibility to receive a larger value by mutation also increases. In the phen-gen model it is possible to make an assumption that the variability of a phenotype increases if the mean value of the phenotype increases ($\sigma_g$ or/and $\sigma_e$ are functions of $\Phi$). But then the growth of the phenotype can be inhibited very quickly. In this work it is shown that without this assumption and without logarithmic transformation of the data the inhibition of growth of a phenotype is also present. But at the beginning the rate of evolution increases in accordance with Haldane’s hypothesis.
Because the genetic drift in small populations is faster than in the big ones (Kimura 1968, Tomiuk et al. 1998) many scientists have come to a conclusion that evolution in small populations is faster than in the big ones, though there are no publications confirming this view to be found. There are very few scientific papers analysing the impact of population size on the rate of evolution. We found only one such paper (Smith 1976) from which it also follows that evolution is faster in larger populations, although this relationship is not linear.

Some scientists observing the fast disappearance of genetic diversity in small populations have been of the opinion that the rate of genetic drift affects the population number (Newman & Pilson 1997). But evolution is a process different from the genetic drift. For evolution rate the decay rate of certain alleles does not matter but what is important is the rate of accumulation of new alleles on the genes.

Genetic drift is easy to get with a phen-gen model. It is enough to assume that the initial population is heteromorphic and $p_{mut} = 0$. Then the changes in the genetic structure in a small population are greater than in a big one and in a small population the mean of decay time of certain alleles is shorter than in a big one. If the $p_{mut} > 0$ and the phenotype does not influence the fitness ($b_r = 0$ and $b_s = 0$) then we can see that the rate of accumulation of new alleles on the genes does not depend on the population size. If intensity of natural selection is different than zero then the evolution is faster in bigger population than in smaller. Such dependence causes that in the beginning, when population numbers increases, the rate of evolution increases.

Although the phen-gen model refers to the artificial life models, it was made for a completely different purpose. The reconstruction of processes existing in nature is not its aim. The phen-gen model was made to be able to analyse separately some elements of the really existing processes. It is hard to find a population in which the probabilities of death change but the probabilities of reproduction do not, whose density increases and it does not cause the exploitation of environmental resources. Answers to the questions how fast evolution is if one factor is changed can be provided by models only. So we do not expect results compatible with reality. The obtained results can explain the evolution of such phenotypes as the length of elephant trumpet or the length of giraffes’ neck and give a surprising answer: these phenotypes have obtained such monstrous size because their bigger size only a little increased the probability of survival of the time period (month, year, etc.) compared to the neighbours with a smaller size. But the evolution of other phenotypes has not been explained. Their evolution must have been more complicated than in the presented model. And this is also a result.
References


Sokol, M. 2014. A phenotype-genotype model of the population and observation of the morphological and molecular evolution due to intensity of natural selection. Faculty of Biology, University of Warsaw. On-line: www.biol.uw.edu.pl/informatyka/phengenmodel


4. Appendix

Source code by C++ of phen-gen model. Functions: pr(N,Φ), L(N,Φ) and ps(N,Φ)
and parameters: sigg, sige and pmut can be changed. Initial population consists of N
individuals of alleles ((0,0,0),(0,0,0)) and it also can be changed.

```cpp
#include <iostream>
#include <cstdlib>
#include <fstream>
#include <time.h>
#include <math.h>
using namespace std;

float normal(float mi ,float sig)
{
    float r0,r1,r2,fate;
    fate=rand()%32768;
    fate=fate/32768;
    r0=2*M_PI*fate;
    fate=rand()%32768;
    r1=fate/32768;
    fate=rand()%32768;
    r2=fate/32768;
    if (r2<r1) r2=1-r2;
    return sig*cos(r0)*sqrt(-2*log(r2*r2))+mi;
}

int Loff(int n, float phen, float a, float b, float c)
{
    int L;
    float x,y,z,lambda,fate;
    if (a*n+b*phen+c>50) lambda=a*n+b*phen+c;
    else lambda=log(1+exp(a*n+b*phen+c));
    if (lambda<50)
    {
        L=0;
        x=exp(-lambda);
        y=exp(-lambda);
        fate=rand()%32768;
        fate=fate/32768;
        while (fate>x)
        {
            L=L+1;
            y=y*lambda/L;
            x=x+y;
        }
    }
    else L=(int) round(normal(lambda,sqrt(lambda)));
    return L+1;
}
```
int main()
{
    int n0, timestep, nrepeat, L, nall;
    float aL, bL, cL, ar, br, cr, as, bs, cs, pr, ps, pmut, sigg, sige;
    //loading the values of parameters n0, aL, bL, cL, ar, br, cr, as ,bs, cs, sigg, sige, timestep, nrepeat
    n0=10; aL=0; bL=0; cL=-20; ar=-0.005; br=0; cr=0; as=0.005; bs=0; cs=-1; pmut=0.001;
    sigg=1; sige=10; timestep=100000; nrepeat=1;
    //creation the result-file
    ofstream info("infodyn.txt",ios::out);
    info<<"repeat timestep numbers n_mutations av_phenotyp sd_phenotyp av_mut_on_gene sd_mut_on_gene"<endl;
    //creating the objects "gene" , state of individual at time and populations (initial, at time t and
    //at time t+1)
    srand(time(NULL));
    struct gene
    {
        float phen;
        int all;
        int mut;
    };
    struct stateofindiv
    {
        int nr;
        gene G1;
        gene G2;
        float phenotyp;
    }; stateofindiv indiv, indivfather, potom;
    stateofindiv pop0[10000], pop1[10000], pop2[10000];
    //creating the initial population
    float avphen=0, sdphen=0, avmut, sdmut ;
    for (int i=1; i<=n0; i++)
    {
        indiv.nr=i;
        indiv.G1.phen=0;
        indiv.G1.all=0;
        indiv.G1.mut=0;
        indiv.G2.phen=0;
        indiv.G2.all=0;
        indiv.G2.mut=0;
        indiv.phenotyp=normal(0.5*(indiv.G1.phen+indiv.G2.phen),sige);
        avphen=avphen+indiv.phenotyp;
        sdphen=sdphen+indiv.phenotyp*indiv.phenotyp;
        pop0[i]=indiv;
    }
    avphen=avphen/n0;
    sdphen=sqrt(sdphen/n0-avphen*avphen);
    //loop of repeating
for (int repeat=1; repeat <= nrepeat; repeat++)
{
    int nrlast = n0;
    for (int i = 1; i <= n0; i++) pop1[i] = pop0[i];
    int n = n0;
    int t = 0, nmut = 0;
    info << repeat << " " << t << " " << n << " " << nmut << " " << avphen << " " << sdphen << " " << 0 << " " << 0 << " " << endl;
    // loop of time
    for (t = 1; t <= timestep; t++)
    {
        int licz = 1;
        if (n > 100000) break;
        // reproduction, heredity, survival of each individual
        for (int i = 1; i <= n; i++)
        {
            indiv = pop1[i];
            pr = 1 / (1 + exp(-(ar * n + br * indiv.phenotyp + cr)));
            ps = 1 / (1 + exp(-(as * n + bs * indiv.phenotyp + cs)));
            float fate = rand() % 32768 / 32768.0;
            if (fate < pr)
            {
                L = Loff(n, indiv.phenotyp, aL, bL, cL);
                for (int pot = 1; pot <= L; pot++)
                {
                    potom.nr = nrlast + 1;
                    nrlast = nrlast + 1;
                    int random = rand() % 2;
                    if (random = 1) potom.G1 = indiv.G1; else potom.G1 = indiv.G2;
                    float lot = rand() % 32768 / 32768.0;
                    if (lot < pmut)
                    {
                        potom.G1.phen = normal(potom.G1.phen, sigg);
                        potom.G1.all = nmut + 1; nmut = nmut + 1;
                        potom.G1.mut = potom.G1.mut + 1;
                    }
                    int irandom = rand() % n + 1;
                    indivfather = pop1[irandom];
                    random = rand() % 2;
                    if (random = 1) potom.G2 = indivfather.G1; else potom.G2 = indivfather.G2;
                    lot = rand() % 32768 / 32768.0;
                    if (lot < pmut)
                    {
                        potom.G2.phen = normal(potom.G2.phen, sigg);
                        potom.G2.all = nmut + 1; nmut = nmut + 1;
                        potom.G2.mut = potom.G2.mut + 1;
                    }
                    potom.phenotyp = normal(0.5 * (potom.G1.phen + potom.G2.phen), sige);
                    pop2[licz] = potom;
                    licz = licz + 1;
        }
Sokol, M. 2014. A phenotype-genotype model of the population and observation of the morphological and molecular evolution due to intensity of natural selection. Faculty of Biology, University of Warsaw. On-line: www.biol.uw.edu.pl/informatyka/phengenmodel