

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Evolution, Biodiversity and Ecology in Microbial Communities: Mathematical Modeling and Simulation with the “Haploid Evolutionary Constructor” Software Tool

Sergey A. Lashin, Yury G. Matushkin,
Alexandra I. Klimenko, Valentin V. Suslov and
Nikolay A. Kolchanov

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/58302>

1. Introduction

Life on Earth exists in the form of interacting communities of two types: populations and ecosystems. Populations maintain reproduction and Darwinian selection. Substance and energy exchange both with the environment and between populations is one of the major functions of ecosystems, which, as a result, enables the necessary conditions for organism reproduction [1–3]. Consequently, it is reasonable to consider any evolutionary process both in terms of population and in terms of ecosystems. Therefore, the evolutionary success of any given genotype carrier is related not only to fixation in population, but also the influence of such a population on functioning of the ecosystem. It is impossible to understand patterns of certain species' evolution without considering the trophic structure of the ecosystem, as a part of which they exist. An ecosystem enables the sparing use of environmental resources by setting up various linear or branched trophic chains [4]. When several linear chains or their branches integrate, a trophic cycle can arise, which enables reproduction of one or another resource [5–7]. On the other hand, an unusable resource – a metabolic cul-de-sac – can arise in ecosystems, which is buried or removed by the flow [8–10].

Symbiotic relationships are widespread in ecosystems, and especially specific to prokaryotic communities. These communities often have taxonomical diversity and are characterized by a complicated trophic chain with a certain extent of closeness. Specific traits of

prokaryotes, such as a pinotrophy diet, small cell size and low speed, lead to an organism's inability to escape from the surrounding environment quickly, and to the "biogenic desert" problem [11], when organisms that inhabit the center of a high-density population have to go hungry due to lack of nutrients. Therefore, the characteristic feature of prokaryotic ecosystems is **metabolism integration** – from the close association of single organisms of different species [12] to spatially-structured, trophically highly-closed ecosystems of meromictic [13] and soda lakes [14], and, after all, global biogeochemical cycles of the biosphere [15]. The large majority of prokaryotes in nature exist as a part of communities with a complicated structure – bacterial mats or biofilms, whose common metabolite pool often forms a complete cycle [15,16]. The presence of such cycles optimizes community members' metabolism. Trophic rings, within which metabolic products of one species or strain are used or can in certain conditions be used for food by others, are revealed in bacterial biofilms and in complex metabolic graphs, reconstructed on the basis of metagenomic projects [17,18].

Close bacterial association in biofilms enhances the probability of horizontal gene transfer between different bacteria, which enable them to obtain new features [19–21]. The importance of the horizontal transfer is supported by recently discovered natural vectors for cloning exogenes in prokaryotic genomes – integrons and superintegrons [22,23]. Therefore, most prokaryotes exist as a part of communities. A wide variety of prokaryotes are unable to grow in pure culture (uncultured prokaryotes) confirms this [24]. The evolution of such highly-integrated communities has its qualitative specific factors and cannot be reduced to the evolution of distinct populations composing them. Thus, while the reproduction rate of prokaryotes is exceptionally high, an experimental study of prokaryotic evolution is difficult, as it requires the study of the whole prokaryotic community. Accordingly, the mathematical modeling of evolutionary processes, adjusted for different types of trophic interactions, spatial distribution of organisms, genetic structure of populations, speciation, different reproduction schemes, environmental influence and other factors, is one of the main methods for the study of the evolutionary process. The modeling of evolution is one of the primary challenges of 21st century biology, mathematics and computer science.

Traditional approaches to evolutionary and population process modeling include methods of population dynamics [25,26] and methods of population genetics [27,28]. Population dynamics modeling methods describe population size changes through time subject to environmental conditions, trophic interactions between populations and other features, but as a rule, population genetic structure changes cannot be studied using these methods. As to population genetics methods, they are generally based on methods of the probability theory and mathematical statistics [29–31]. They allow for studying the evolution of population genetic structure, but do not provide means for the modeling of the population dynamics process in detail.

Further development of modeling and simulation methods led to so-called "hybrid" methods [32,33], which allow us to investigate changes both in the size and genetic structure of a

population simultaneously. However, it should be noted that the large majority of these methods (like the other population genetics methods) are focused on diploid organism population modeling (generally, with sexual reproduction).

Methods of “portrait” (individual-oriented, agent-based) modeling [34–36] bridge the gap between individual characteristics and community structure, evolving as a consequence of some or other rules of interactions between individuals. These approaches are illustrative enough, but may have a high computational complexity (ideally, they require a detailed description of every individual in the population). Besides, most of them are static, i.e. during calculations their structure can be changed only within the limits predetermined by the developer. As a result, if modern computer capacity is sufficient for modeling of a population of diploid organisms with sexual reproduction (whereas the effective size of such populations is generally no more than 100-1000 individuals), then direct simulation modeling is often problematic for the modeling of a population of haploid organisms, particularly bacteria (the effective size of bacterial populations is 10^6 - 10^9 individuals).

With due consideration to the abovementioned requirements of the evolutionary and population process modeling tool, we previously developed a modeling method and software package “Haploid evolutionary constructor” (HEC) [37,38]. The HEC provides tools to simulate the functioning of a haploid organism population network, trophically linked with substrate-product relationships under the environmental effect. This modeling approach provides a means of simultaneously describing the prokaryotic community at various levels of its biological organization: genetic, metabolic, population and ecological, flexibly varying the degree of description detail at any level. During the model simulation, each population may vary both its size and genetic diversity due to selection and mutations. The key feature of this methodology is the ability to model such evolutionary and population processes that require an intense structure rearrangement of a model during simulation. Such processes contain, for example, horizontal genetic transfer and speciation. Besides, the methodology offers the possibility to describe the polymorphism of one or several genes in a population, where the number of alleles can be changed during the simulation process.

In summary, the HEC provides comprehensive study of the bacterial community model, analyzing the dynamics of changes in allelic frequency, the population size, the concentration of metabolites, the community trophic structure and its evolution, including stochastic genetic factors.

During model simulation via the HEC, the number of populations reached 300, with an approximate total 10^{20} individuals, the size of distinct population came to 10^{18} , with around $10^6 - 10^7$ different genotypes (up to 4000–15000 different genotypes within one population in case of multiple polymorphism of 10–15 genes).

2. Haploid Evolutionary Constructor methodology description

The HEC methodology provides modeling and simulation of biological and evolutionary processes in trophically linked communities of unicellular haploid organisms. Figure 1 shows

the scheme of the main HEC objects and processes. The HEC simulates trophic interconnected haploid organism networks that are combined into populations according to genetic proximity and reside in one whole volume termed *environment*. Organisms may consume and *utilize* substrates and synthesize, and then *secrete* products into the environment, which inversely can be used by other organisms as substrates. Some substrates favorably affect population growth, others, on the contrary, may have an inhibiting effect. The efficiencies (as rate constants) of substrate utilization and product production are controlled by certain genes.

So-called *nonspecific* substrates are reproduced by the in-flow (Figure 1 shows scheme with one nonspecific substrate N_1). By contrast substrates running into the environment only through cell activity (synthesis and secretion) are known as *specific* (represented by S_i in Figure 1).

2.1. Environment

Environment is a bounded flow system of fixed volume V_{total} containing all populations and substrates. The environment is also a mediator of relationships between populations (and substrates); the inflowing and outflowing processes of both substrates and cells are connected with it. The environment is characterized by following variables:

- V_{total} – environment capacity (in liters);
- $k_{env,flow}$ – flow rate (in % V_{total} per unit time);
- $N_{env,i}$ – concentration of the i^{th} nonspecific substrate in the environment (in mM);
- $S_{env,i}$ – concentration of the i^{th} specific substrate in the environment (in mM);
- $N_{flow,i}$ – concentration of the i^{th} nonspecific substrate in the flow (in mM);

and the following processes:

substrates inflow into the environment – increases a nonspecific substrate concentration according to the flow rate and concentration of these substrates in the flow;

substrates outflow from the environment – reduces both nonspecific and specific substrate concentrations according to the flow rate;

the inflow/outflow of nonspecific substrates follows the formulas:

$$N_{env,i}(t+1) = N_{env,i}(t) + k_{env,flow} \cdot (N_{flow,i} - N_{env,i}(t)) \quad (1)$$

the specific substrates outflow follows the formula below (specific substrates flow into the environment associated with these substrates through cell synthesis as described below):

$$S_{env,i}(t+1) = S_{env,i}(t) \cdot (1 - k_{env,flow}) \quad (2)$$

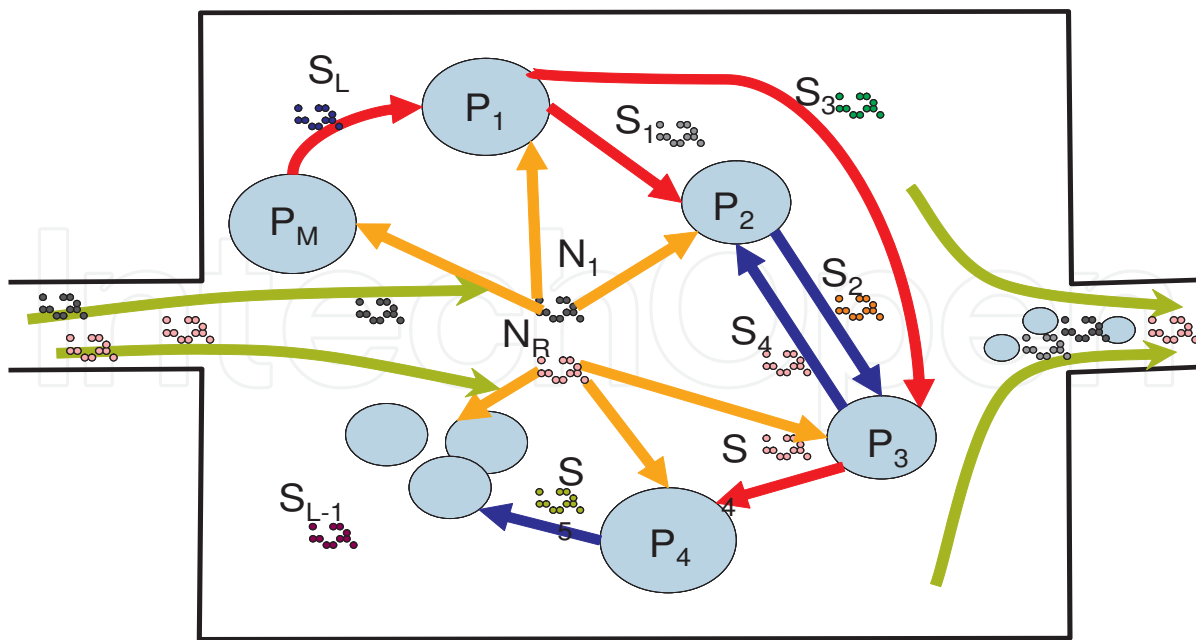


Figure 1. Scheme of HEC objects and processes. Blue circles represent populations P_1, P_2, \dots, P_M . Groups of colored dots represent substrates: specific S_1, S_2, \dots, S_L and nonspecific N_1 . Arrows between populations represent trophic relationships: the population, from which the arrow comes, produces the substrate, used by the population to which this arrow leads. Red arrows represent a substrate's activating effect, blue arrows – an inhibiting effect. The usage of a nonspecific substrate is indicated by orange arrows. The flow action is represented by the thick arrow in the lower left-hand corner of the figure. The bounded area represents the environment.

2.2. Population modeling

We consider populations to be a set of cells, that have common substrate using and producing properties. Cells are considered to belong to the same population (same species) if they:

1. utilize the same variety of nonspecific and specific substrates;
2. produce the same variety of specific products;
3. have the same trophic strategy;
4. have the same synthesis strategy.

We define basic terms and notions, used hereinafter as follows:

- **trait** – any given substrate synthesis or utilization rate constant. Every trait is considered as unambiguously controlled by one **gene**. In this particular case, the **gene** is considered a unit of inheritance;
- **allele** – a gene variant, i.e. a particular value of the corresponding constant;
- **individual's genotype** is the set of alleles, divided into five groups. The first group (c_i) characterizes the efficiency of specific substrate utilization (s_i), the second group (d_i) – substrate production rate, the third group (r_i) – efficiency of nonspecific substrate utilization,

the fourth group (m_i) – efficiency of immunity against phages, and the fifth group (v_i) – phage (viruses) genes;

- **mutation** is a change of a corresponding trait value, which can be interpreted as a gene shifting into another state (allele).

Using the introduced terminology, the concept of a **monomorphic population** can be formed – a population of “genetically identical” cells, where all cells have corresponding genes represented by the same alleles. The genotype, common for all the cells of such population, is called the **monomorphic population genotype**.

In order for cells to use substrates for their reproduction and population size growth along with products synthesis, at first they have to get these substrates from environment. In the HEC, the process of substrate consumption is described by the particular step, where a cell's requirement of various substrates and the availability of substrates in the environment are taken into consideration. In case of any substrate deficiency, a competition for this substrate may occur either intrapopulation or interpopulation (when cells of several populations may use the same substrate at the same time). In case of substrate excess, the maximum amount of the substrate consumed by one cell is defined by the value of the substrate **consumption rate**. This value is species-specific and cannot be changed due to mutation (in this HEC version) – this may be equivalent to, for example, the size limit for one cell. Hence, monomorphic population is additionally characterized by the amount of substrate molecules consumed.

Trophic strategies is a term for the formulas and laws, describing population changes in a single generation depending on population size, the amount of consumed substrates, the flow rate, mortality rate and other factors. Examples of trophic strategies are illustrated by the equations below (other formulas may be also used, including those defined by the user):

$$\Delta P = F_1(\vec{N}, \vec{S}, \vec{R}, \vec{C}, P) = \sqrt{r_0 n_0(P) \cdot \sum_{i \in I_{consumed}} c_i s_i(P)} - k_{flow} \cdot P - k_{death} \cdot P^2 \quad (3)$$

$$\Delta P = F_2(\vec{N}, \vec{S}, \vec{R}, \vec{C}, P) = P \cdot \frac{\left(\frac{n_0}{P}\right)^{\gamma_0}}{1 + \left(\frac{n_0}{P}\right)^{\gamma_0}} \cdot \prod_{i \in I_{consumed}} \frac{\left(\frac{s_i}{P}\right)^{\gamma_i}}{1 + \left(\frac{s_i}{P}\right)^{\gamma_i}} - k_{flow} \cdot P - k_{death} \cdot P^2 \quad (4)$$

$$\Delta P = F_3(\vec{N}, \vec{S}, \vec{R}, \vec{C}, P) = a_{basal}(n_0) \cdot P - \sqrt{\sum_{i \in I_{consumed}} c_i s_i(P)} - k_{death} \cdot P^2 \quad (5)$$

where

$I_{consumed}$ – set of indices of substrates consumed;

n_0 – amount of nonspecific substrate, consumed by the cells of population from the environment (in proportion to the population size);

\vec{N} – vector of specific substrates, consumed by the cells of the population from the environment (in proportion to the population size) values;

r_0 – utilization rate for the unique nonspecific substrate (trait, controlled by the corresponding gene);

\vec{S} – vector of corresponding specific substrate utilization rates (traits, controlled by correspondent genes);

P – population size;

k_{flow} – flow rate in the environment (“washout” rate);

k_{death} – population mortality rate;

a_{basal} – “natural increase” of the population;

$\gamma, \gamma_0, \gamma_i$ – coefficients, describing the nonlinear nature of substrate influence on population growth;

K_{ij} – coefficients, describing the efficiency of substrate influence on population growth (depends on corresponding traits).

The equation [eq.3] governs the utilization process of several specific and one nonspecific substrate, where the latter is essential for cells – when it is not available, the population does not grow. Substrates have strong cooperativity. Besides, some substrates are able to compensate in some degree the lack of other necessary substrates, including a nonspecific substrate deficiency. The trophic strategy described in [eq.3], satisfies the **Rubel’s law of replaceability of ecological factors** [39], and is called the **compensatory trophic strategy**.

The equation [eq.4] again governs the utilization process of several specific and one nonspecific substrate. Nevertheless, every substrate is essential. A deficiency of one substrate cannot be compensated by an excess others. The trophic strategy described by [eq.4] satisfies the ecological **Liebig’s law of the minimum** [4] and called the **noncompensatory trophic strategy**.

The equation [eq.5] governs the utilization process of one nonspecific substrate coupled with the inhibiting effect of specific substrates on population growth. Besides, this effect is cooperative. This trophic strategy is called the **inhibitory trophic strategy**.

Similar to the concept of a monomorphic population, we define the concept of **polymorphic population**, which can be regarded as a set of monomorphic subpopulations. Cells in a polymorphic population have the same gene variety, while different cells can have different alleles of one or several genes. The polymorphic population is characterized by the “generalized genome” – a set of population’s **genetic spectra**. The **genetic spectrum** is the distribution of allele occurrence frequencies in a population (for one gene) (Figure 2).

Mutation in terms of a genetic spectrum means the change of its profile (thus, the formation of a new allele is possible, Figure 3).

The change of the polymorphic population size is calculated according to the following scheme. The polymorphic population is split into many monomorphic subpopulations. Then,

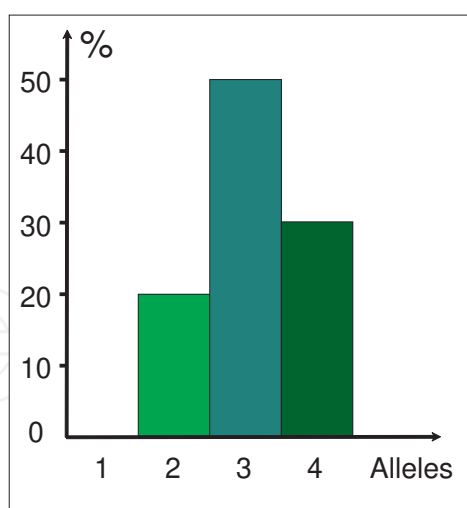


Figure 2. The genetic spectrum shows that the given trait value is 2 for 20% of individuals in this population, 3 for 50%, and 4 for 30%.

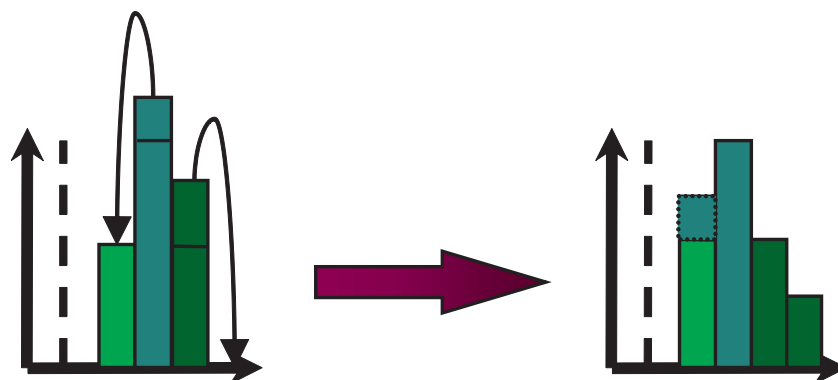


Figure 3. Change of genetic spectrum due to mutations.

the growth for each subpopulation is calculated according to formulas similar to the above-mentioned (eq.3-5). Next, the monomorphic populations are merged into a polymorphic population. The polymorphic population is split based on the proportion of every allele for all genes in the population genotype. The substrates consumed by the polymorphic population are divided proportional to the sizes of the single monomorphic populations. It noteworthy, that each single monomorphic population growth may differ markedly from the growth of other subpopulations. It depends on the monomorphic population genotype, population size and the amount of certain substrates consumed by the population. Consequently, the proportion of alleles in a population may change (which may be interpreted as the adaptation of the population to certain conditions).

2.3. Metabolism

When a cell synthesizes a product, that it can utilize itself, it is obvious that there is “no use” in secreting this product into the environment and then competing for it on “equal terms” with

the other populations. For this reason, we consider two forms (states) of substrates in the modeling of the internal cell substrates. The first form are the substrates that are “**ready for utilization**”, the second form are **synthesized** substrates. The substrates of the first group are replenished by substrate consumption from the environment and through transition of synthesized substrates (if the cell can synthesize it). During reproduction and product synthesis, the substrates of the first group are expended. The substrates of the second group are replenished only by synthesis. A principal scheme of cell metabolism in a population, consisting of the stages of substrate consumption, utilization, synthesis, and secretion is illustrated by (Figure 4).

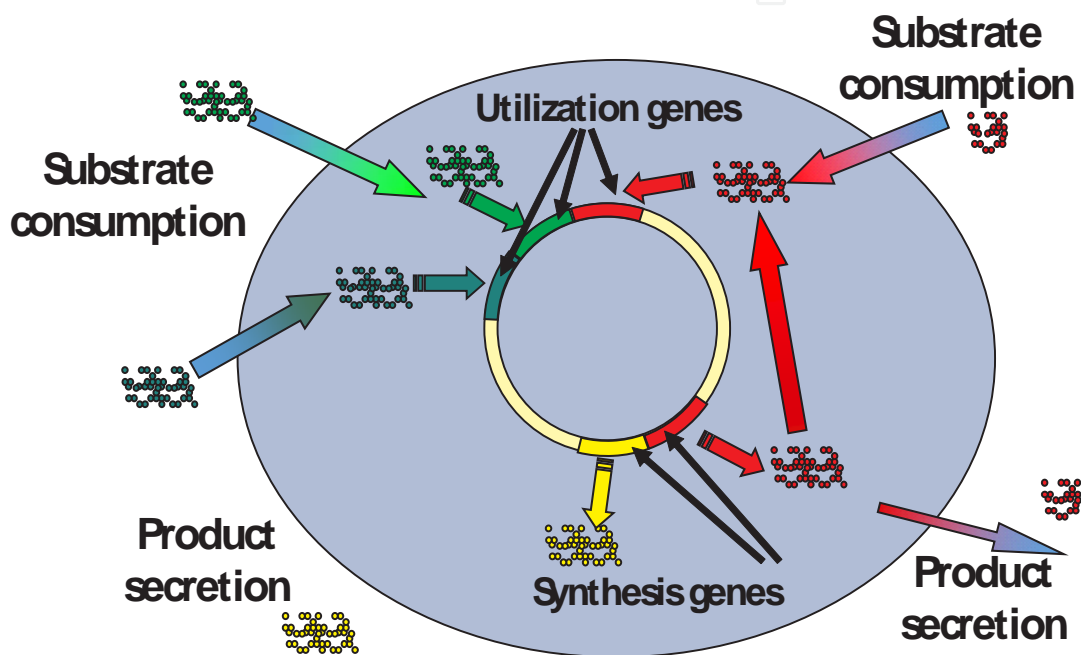


Figure 4. A scheme of trophic processes through the example of one cell. The cell utilizes three types of substrates and synthesises two types of substrates, one of which it can utilize itself. A synthesized substrate, which cannot be utilized by the cell (yellow) is comprehensively secreted into the environment, while the second substrate (red) is partly used by the cell itself (consequently, there is less or no secretion into the environment).

Product synthesis by the cells of a polymorphic population is described by the gene network model of metabolite synthesis, which we call the **synthesis strategy** (Figure 5).

An example of a formula describing a simple synthesis strategy:

$$\Delta s_i = P \cdot \sum_{j \in Spectr_i} d_{ij} s_j \left(\frac{P_j}{P} \right) \quad (6)$$

where

Δs_i – amount of synthesized i-type substrate;

d_{ij} – trait value in genetic spectra $Spectr_i$;

P – population size;

P_j – proportion of individuals having d_{ij} trait value in population (in the genetic spectra).

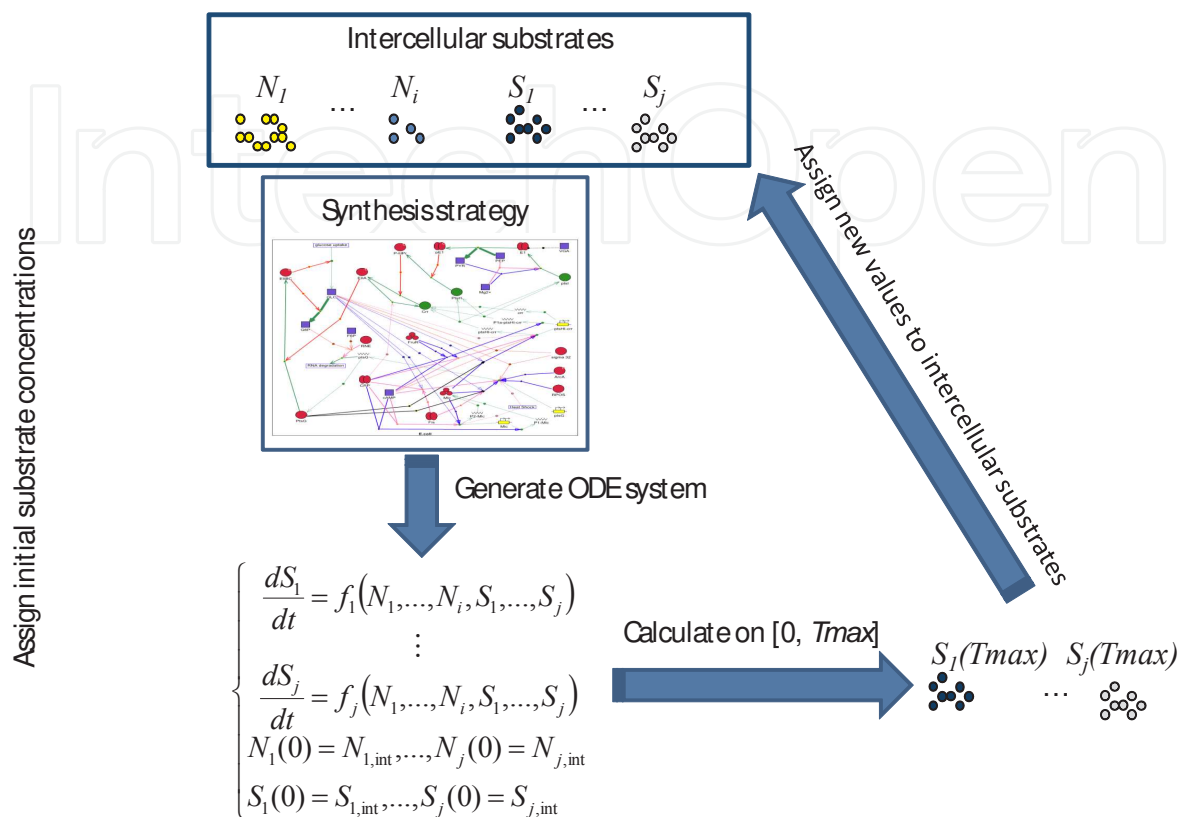


Figure 5. Common pattern of synthesis strategies calculation.

2.4. Infection process modeling

To model a phage infection, the HEC provides extended objects of polymorphic populations: *polymorphic phage populations* and the *polymorphic population of infected cells* (the “normal” polymorphic populations are therefore regarded as “healthy”).

The infection modeling includes the following phases: infection of the healthy cells through phage penetration from the environment into one part of the cell population, phage reproduction inside of the infected cells, and finally, the phages burst into the environment after the lysis of cells. The infected cells form polymorphic populations, further reproduction of which may follow a lytic or a lysogenic pathway. The lytic pathway means death (lysis) of the infected cells with synchronous phage formation and their transportation into the environment (the number of phages depends on their profusion). The lysogenic pathway means prophage formation and no phage transportation into the environment follows. At the same time the population of the infected cells reproduces like an ordinary polymorphic population in the HEC (i.e. according to the trophic strategy), acquiring immunity to that type of phage (Figure 6).

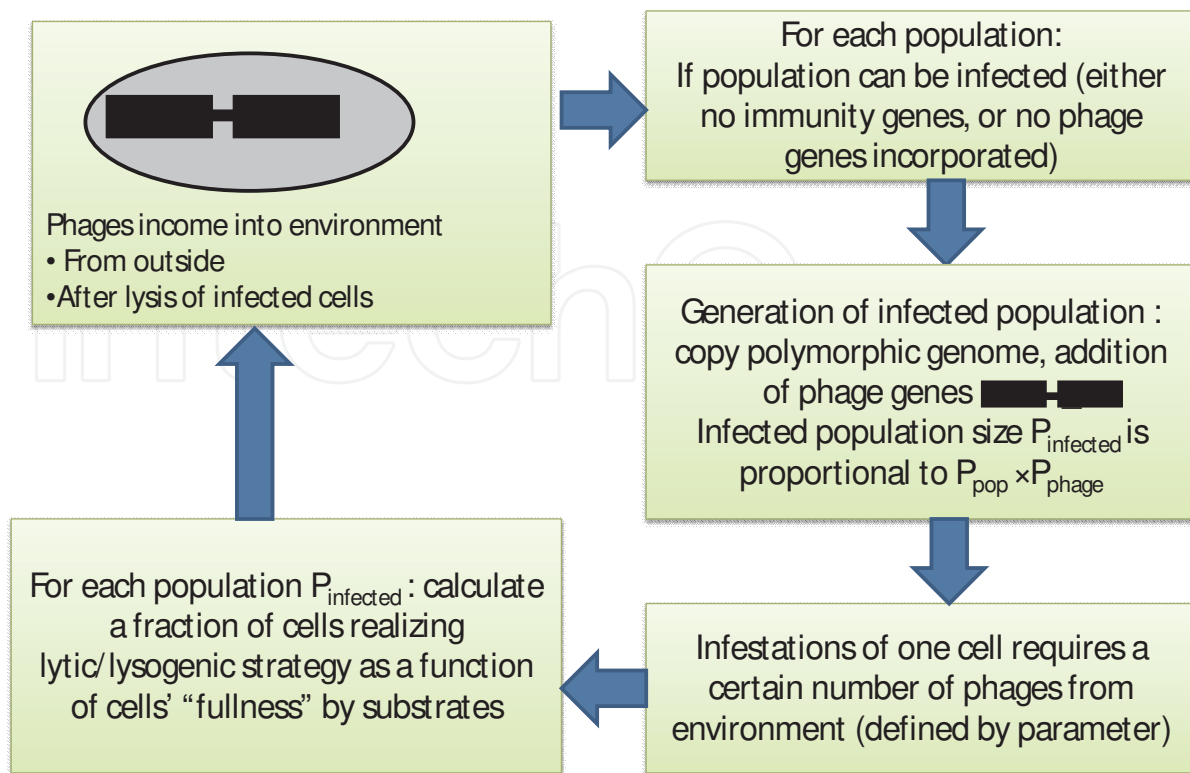


Figure 6. Scheme of phage infestation.

The choice of the lysogenic or the lytic scenario depends on the conditions of the cell population at the moment of infection: optimum conditions lead to a lytic type, pessimal conditions– to a lysogenic. In the latter case, a part of the population randomly switches to the lytic form if conditions improve, causing the death of this part and phage generation (Figure 7).

The polymorphism of phages, the formation of new strains owing to mutations and competition between strains are also described via the genetic spectra arithmetics. However, in the phage populations unlike prokaryotic populations, genes define a specific virulence (an ability to infect certain populations) and abundance (the number of copies per lysed cell).

3. Simulation of prokaryotic communities via the Haploid Evolutionary Constructor

Through the use of the HEC software tool, we have simulated a number of biological models of the functioning and evolution of unicellular haploid organism communities [37]. Inter alia, the correspondence of the modeling results to biological data as well as previously published mathematical models has been illustrated. We estimated the key parameters of the model, regarding cell size, number of substrates required for cell division and other factors based on *E.coli* cell information [40].

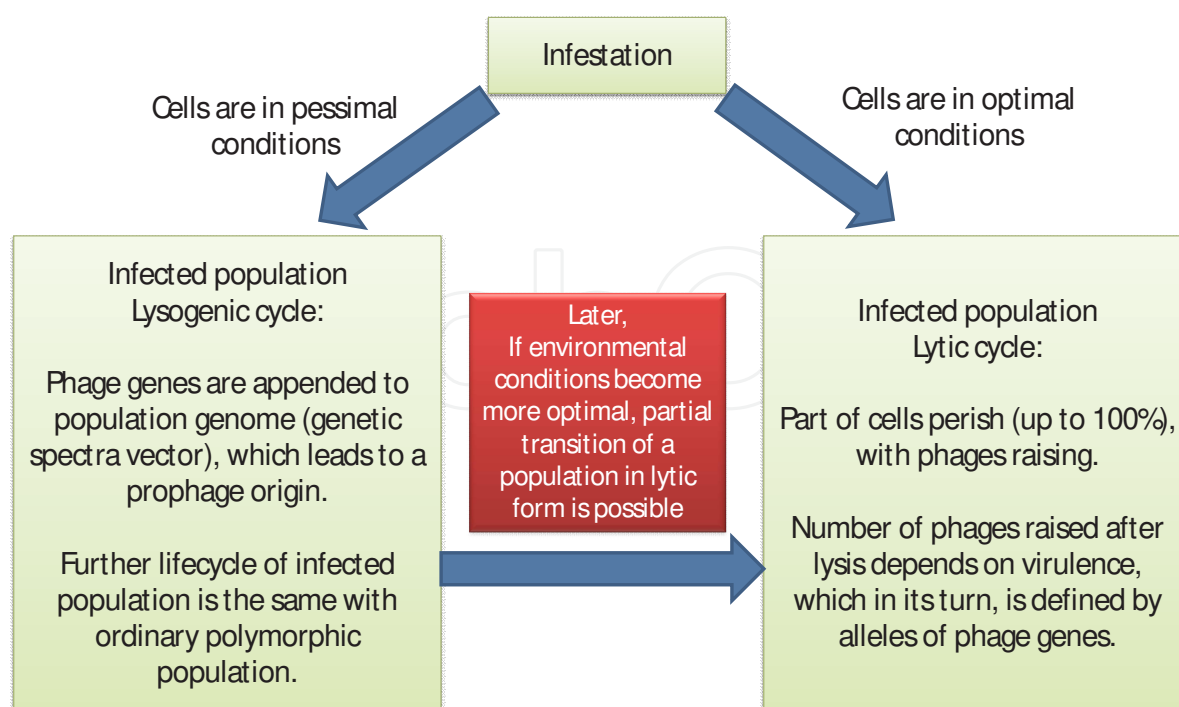


Figure 7. Scheme of the choice of the lysogenic or lytic scenarios.

To verify the modeling approach, a number of basic models were considered, and the obtained results confirmed correspondence to both classical models and experimental information [38]. Thus, the models examined proved the biological adequacy of the HEC and its applicability for a wide range of population, genetic and ecological problems.

3.1. Modeling of biodiversity dynamics and adaptability in bacterial communities

The evolutionary success of a biological system relates to the balance of two characteristics: its stability (ability to preserve its structure and homeostasis despite changes in the environment) and its adaptability –ability to preserve an evolutionary flexibility in response to uncontrolled environmental changes. Traditionally, communities that are more complex are considered to be more stable, but there can be exceptions. The study of model dynamics, when the population size growth of one species depends on the population size of others species in the community, confirms that increased complexity leads to the improvement of stability only when community connectivity increases at the same time.

We maintained a comparative simulation of stability dynamics, the adaptability and biodiversity of trophic closed communities with compensatory and non-compensatory metabolism (according to Rubel and Liebig respectively). The trophic strategies formulas representing these laws [eq.3,4] are mentioned in the HEC models description. The deficiency or low concentration of a nonspecific substrate in the environment leads to population extinction for both strategies. However, in case of compensatory nutrition strategy, the deficiency (a low concentration, but not a complete lack) of a nonspecific substrate in the environment can partly be compensated by the high concentration of specific substrates. This satisfies Rubel's law of

replaceability. In case of non-compensatory nutrition strategy, the deficiency of any substrate cannot be compensated by the extra concentration of other substrates. This case satisfies Liebig's law of the minimum.

It is known that the change of conditions on Earth shows a certain cyclism. Its sources are geophysical and astronomical cyclic processes [41,42]. As a result, the input of matter and energy into the ecosystem tends to change. To keep the common value of this flow constant, the ecosystem must conduct an evolutionary search for new sources of matter and energy [6,43]. Hence, the "learning ability" of the ecosystem becomes a critical parameter. The horizontal gene transfer (HGT) in prokaryotic populations is a relatively easy way to carry out such "learning". The simulation showed that the trophic structure of a community imposes substantive restrictions on the benefits of the HGT – a long-term effect is possible only in case of HGT between populations, whose metabolisms are sufficiently rich and flexibly adjustable (compensatory trophism). For populations with simplified metabolisms (non-compensatory trophism), the HGT offers only localized advantages. This contradicts the assumption that the major trend of prokaryotes evolution is individual genome simplification, compensated by relationship amplification in the bacterial community. Such simplification in the long term leads to the death of the community.

The simulation showed dramatically better adaptivity of trophic rings with compensatory (TRC) trophism in comparison with non-compensatory ones. The fixation of beneficial mutations even in one TRC population improved stability of the entire system, significantly extending its lifetime, i.e. offering an additional chance to wait out the starvation, or even completely saved the TRS from extinction due to the metabolism optimization. Nevertheless, many taxa in the world are unable to compensate one resource with another on a broad scale. Why is the major trend of the evolution on Earth increasing biodiversity by means of progressive specialization, rather than the formation of biota based on several taxa-generalists? Let us assume that "learning ability" is a critical parameter for evolution within an ecosystem. Subsequently, in case of a low biodiversity level in an ecosystem with compensatory trophism, it is entirely possible that new sources of energy and matter will never be found. In fact, the probability of finding such sources is higher in ecosystems that preserve a high level of biodiversity until extinction. Then, if the value of the initial nutrient flow is recovered, a new source will be added, providing the ecosystem with resources for further progressive evolution. Therefore, in the long term Liebig's systems have an advantage over Rubel's. It should be noted that this advantage has, as with all evolutionary processes, a nondeterministic, stochastic nature, while the stability of the advantage is determined. Thus, when biodiversity is high, the system may die out for stochastic reasons. This matches paleontological information. A permanent rotation of hegemonic biotas, without sacrificing comparably small amounts of epibiotic ecosystems, accumulating virtually the entire range of biochemical activities (cyanobacterium tufts, alkalophilic communities), is observed in the fossil records.

3.2. Evolutionary trends of genome sophistication and simplification

One of the most important periods in the evolution of life on Earth is eukaryotic cell formation [44]. The comparative study involving genome-wide information for bacteria, archaea and

eukaryotes suggest that the microbial eukaryotic domains could not be inherited from the ancestors of mitochondria and plastids, but were borrowed from other bacteria [45]. In such a case, the origin of eukaryotes is a consequence of the autonomism of one member of the complex syntrophic prokaryotic community through the looping back of major regulatory interactions. It is considered that such autonomism is the result of symbiosis of several types of prokaryotes, and it is fairly probable that the entire series of HGT between symbionts took place.

3.2.1. Genome amplification

By applying the HEC, we investigated the evolution of a community, which at the beginning of computations constituted a trophic ring, in which each of the three populations consumed and produced precisely one specific substrate. For example, the first population produced the substrate consumed by the second, and the second produced the substrate consumed by the third population, etc. (Figure 8).

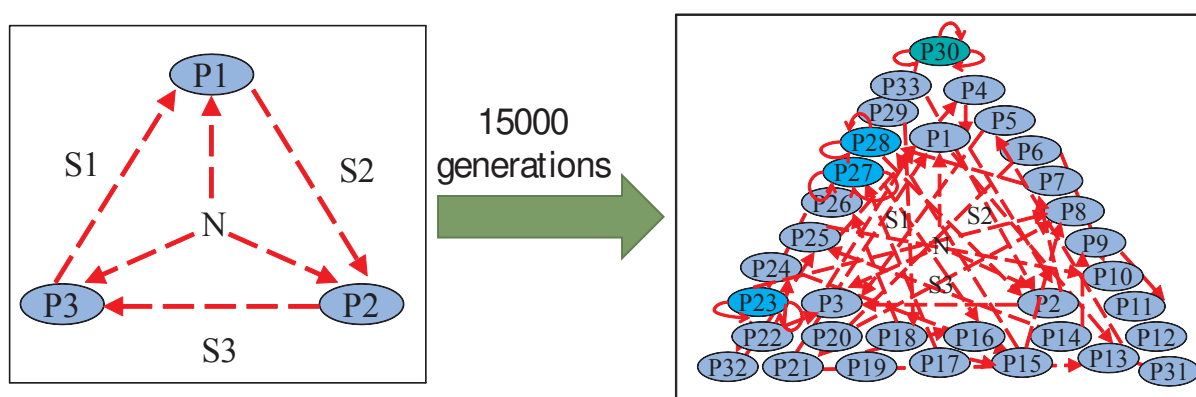


Figure 8. The trophic ring scheme consisted of three populations at start time (left) and after 15000 generations (right).

During calculations with a certain probability between cells of different populations (10^{-7} per generation per cell), HGT could take place. As a result, new populations, either consuming a larger number of specific substrates or producing a larger number of specific products would be formed. It was found, that in the longer term (after 10000–15000 generations), the population with the “most complete genome” (i.e. populations consuming and producing the maximum possible number of specific substrates in the given trophic system), or the population with an “almost complete genome” gain an extreme biomass advantage at the specified conditions. Namely, at moderate genome-length-penalty level values (0.01–0.05), and with stable poor environmental conditions (nonspecific substrate concentration in the inflow is around 10^{-4} mM, i.e. at the survival minimum for parent population cells) (Table 1). In the long-term, such populations replaced all other populations from the trophic system (“outsider” populations either died or reached a maximum number in the environment that hovered around 10–100 individuals) (Figure 9).

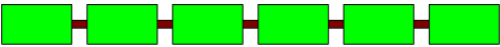
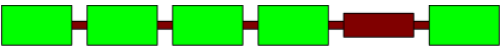



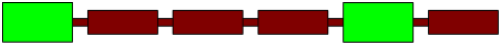

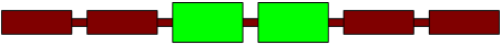
Final population size level	Number	Genotype (scheme)	Total size (individuals)
1	P30		$4.86 \cdot 10^7$
2	P24		$4.85 \cdot 10^7$
3	P28		$2.94 \cdot 10^7$
4	P23		$2.94 \cdot 10^7$
5	P27		$2.94 \cdot 10^7$
---	---	---	---
18-33	P1		0
18-33	P2		0
18-33	P3		0

Table 1. The final distribution of the population sizes after long-term evolution (over 15000 generations) of the trophic system. In the genotype scheme, a green bar represents gene presence in this position (first 3 positions represent 1-3 specific genes of substrate utilization, latter 3 positions represent 1-3 specific genes of substrate synthesis). The table shows that the first and most populous populations are the populations with a “complete genome” P30, next are populations with an “almost complete genome”. Some start populations (P1-P3) became extinct, as did 12 other populations.

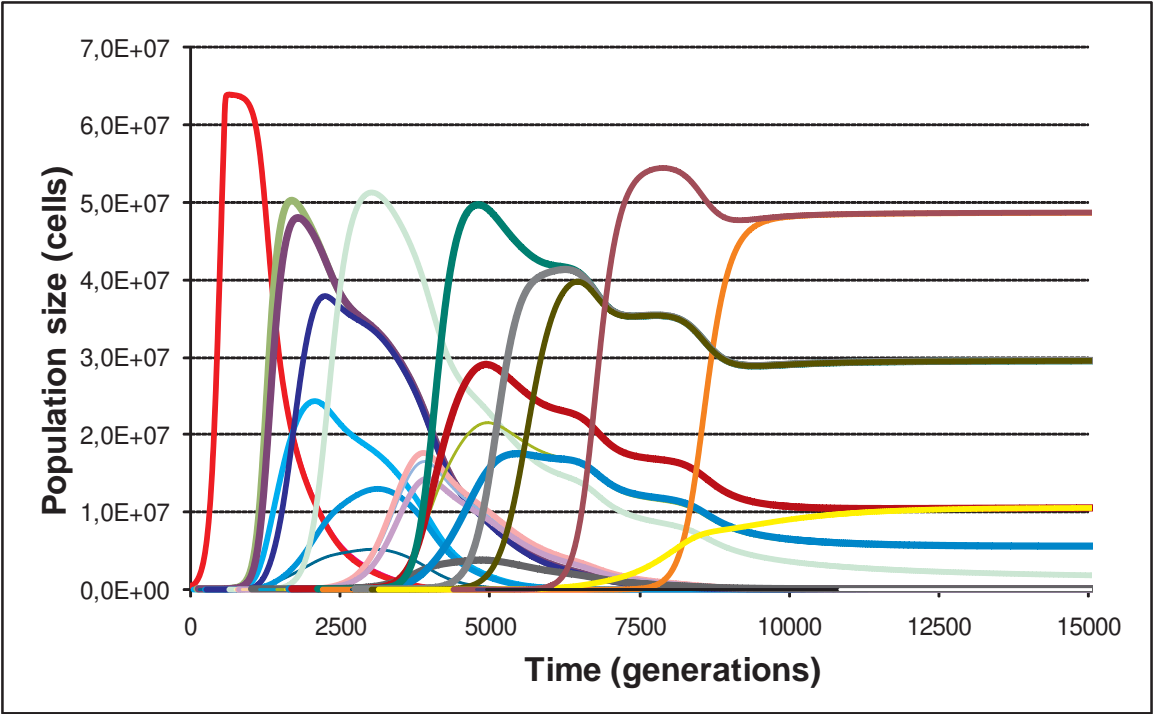


Figure 9. The evolution of a trophic system where new populations form due to HGT. After 7000-8000 generations, populations with metabolically rich genomes that quickly replace other populations have formed.

3.2.2. *Genome simplification*

We have also simulated the processes of gene loss due to a higher genome-length penalty (0.1-0.25) for the environment in which a single metabolically complete population is presented *ab initio* (the genetic analogue of P30 from Table 1). During the iteration process, both gene loss and HGT compensating deletions could occur. It was demonstrated, that a strong genome reproduction tendency is observed in both suitable and unsuitable environmental conditions. The most primitive populations, possessing just two utilization genes (one for nonspecific and the other for specific substrates), replaced other populations.

Therefore, the selective advantage of the “metabolically rich” populations emerged under pessimal conditions and was a combination of autonomism (lower dependence on specific substrate concentration in the environment) and egoism (cells of an autonomous population on average secrete less specific substrates into the environment and consume them independently). The selection supports autonomizing populations, however their formation causes the degradation of the trophic ring into an alliance of autonomous populations, accompanied by a “trail” of small nonautonomous populations. The simulation showed that HGT, in case of poor environmental conditions, actually transforms the collective metabolism of a trophic ring to the metabolism of separate populations, which in general conforms the scenario [46] of eukaryotes formation in symbiogenesis.

However, if conditions are optimal, cells reduced their genome successfully, as well as reducing the time necessary for the reproduction. This fact was observed during the experiment repeatedly [47,48].

3.3. **Evolutionary trends in prokaryotic communities influenced by phages**

To study phage infection influence on possible evolutionary tendencies, models of prokaryotic community infection were constructed. The model from 2.2 (a trophic ring consisting of three populations) was used as a basic model. The horizontal transfer and gene loss processes were stochastically generated during simulations. The addition of a phage population to the community led to an infestation of all populations, while the proportion of infected cells depended on the phage concentration in the environment. The infection fundamentally changed the dynamics of the community, inhibiting the growth or even destroying fast growing populations (following the infection of a lytic pathway), and as a result supported less competitive populations in such conditions. For example, a series of numerical experiments showed that, in pessimal conditions, environments can displace the populations that are far from having a metabolically complete genome (Figure 10). This presents a contrast to the tendency of genome amplification in such conditions as noted above (ref. 2.2).

Possible changes in evolutionary tendencies in case of a phage infection have also been discovered for communities in optimal environmental conditions. Figure 11 shows the survival of a population with a metabolically complete genome – the worst adapted population in the community before the moment of infection. The periodic nature of both prokaryotic and phage population size curves is explained by spontaneous transition of a part of cells to a lytic form, followed by death or the forming of new phages due to lysis.

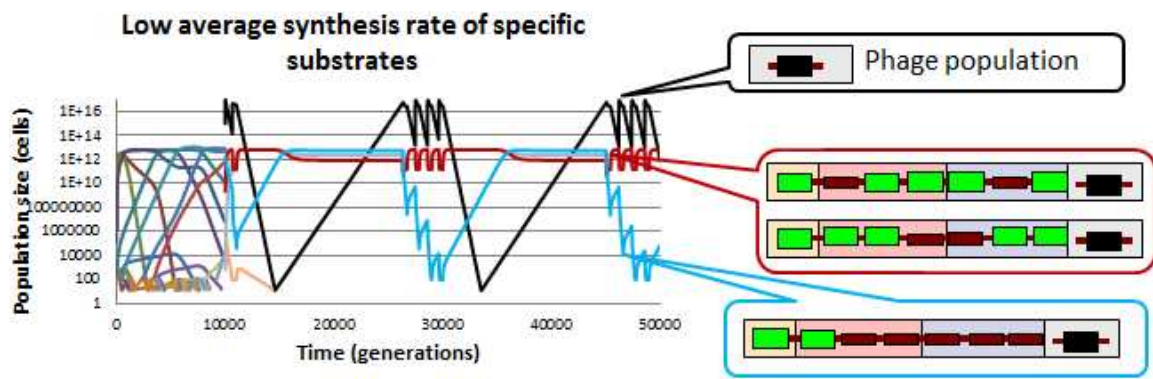


Figure 10. Population size dynamics in pessimal environmental conditions. For some, a species formation takes place in the system due to HGT and gene loss, but an infection occurred at the 10000th iteration. Shortly afterwards, most of the populations became extinct. Surviving populations did not have a metabolically complete genome. Furthermore, among the survivors there were populations with extremely primitive genomes (lower scheme in figure). In the genome figures, the black bar represents the phage gene, i.e. all survived populations are phage carriers.

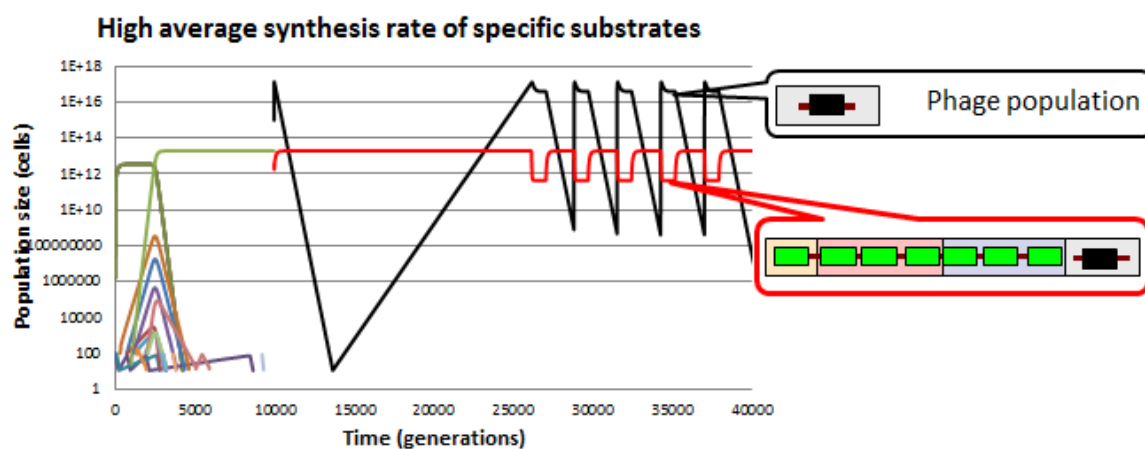


Figure 11. Population size dynamics of a community in optimal conditions. For some time species formation takes place within the system due to HGT and gene loss, but at the 10000th iteration an infection occurs. Shortly afterwards, all populations become extinct, except the population with a metabolically complete genome infected by the phage.

It should be noted that the given results of numerical simulations are stochastic in nature. In a number of numeric experiments, infection led to death of the entire community, or the community died before the infection due to the fast growth of unduly primitive populations. Changes in evolutionary tendencies do not always take place either. Consequently, our results show that phage infection of a community can, but does not necessarily, change its evolutionary tendency.

3.4. Simulation of prokaryotic communities implementing synthesis strategy according to the molecular triggers principle

The model of genetic regulation of protein biosynthesis (Figure 12), suggested by F. Jacob and J. Monod [49], and mathematically studied by D.S. Chernavsky and his colleagues [50], is of a trigger nature.

This model describes a system consisting of two genes, G_1 and G_2 , which synthesize regulators P_1 and P_2 , which are suppressors of G_2 and G_1 respectively. This genetic system has at least two stable stationary states – gene G_1 or gene G_2 is expressed within the system. The probability of transit (a switch) from one stationary state to another is caused by disturbing factors (for example, by adding of one of the regulators to the system), which enables us to suggest the trigger nature of this system.

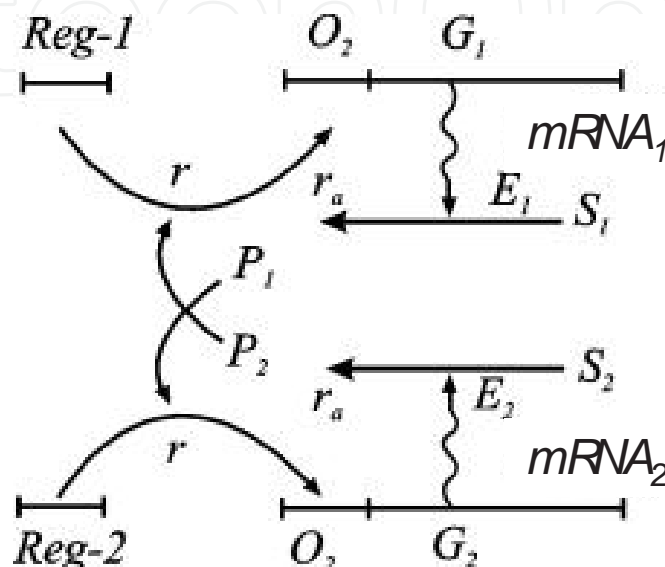


Figure 12. Scheme of synthesis of two enzymes according to Jacob and Monod.

By applying the HEC, a model of a trophic community with one nonspecific substrate N_1 , consisting of one population and synthesizing two specific substrates S_1 and S_2 has been built. The synthesis constants for these substrates are noted as d_1 and d_2 respectively. The synthesis strategy for the cells in the population is described through the gene network – the molecular trigger. The mathematical model of the corresponding gene network is given below [eq.7]. Liebig's non-compensatory strategy has been used as the trophic strategy.

$$\begin{cases} \frac{dS_i}{dt} = \frac{\bar{d}_i}{1 + S_i^\gamma} - S_i \\ \frac{dS_j}{dt} = \frac{\bar{d}_j}{1 + S_i^\gamma} - S_j \end{cases} \quad (7)$$

Where \bar{d}_i and \bar{d}_j – are mean values of the d_i and d_j traits in the population.

The parametric analysis of this model has been reported in many studies e.g. in [51]. In particular, it is shown that if $\gamma \geq 2$ and certain values of $\bar{d}_i/\bar{d}_j > g$, the system takes on the properties of trigger. There are two stable critical points on the phase plane of the system, between which a labile saddle point is situated. The meaning of the g parameter is bifurcational, while the bifurcation is of a trigger nature (saddle forms).

We have conducted a series of computer simulations and have illustrated the possibility of a trigger mode appearing in this model, depending on the gene network parameter values [eq. 7], and physiological, population and ecological parameters. We have demonstrated that the modes of gene networks functioning inside the organism taking into account restrictions on organism, population and ecological levels, may considerably differ from the modes predicted based on analysis of mathematical models of these networks. For instance, in the molecular trigger model [eq. 7] the saddle point ($S_1=S_2=c$) is an unstable stationary state. However, in our model we showed the probability of the system stabilizing precisely in such a state despite the fact that the initial data differed from this state. Such stabilization is possible due to additional factors, including limited cell wall permeability, which limits the effect of substrate switchers.

As can be seen from the above, the models we built suggest an **additional epigenetic mechanism of the functioning mode stability of persistence gene networks**. Such mechanisms were theoretically and experimentally developed by, for example, R.N. Tchuraev and colleagues [52,53]. Via our models, we have also obtained examples the stability persistence of gene network functioning. That is true even in cases when the gene network structure in itself supposes both the presence of several such modes. The possibility of switching between these modes have also been obtained through this models. **Such mechanisms can, on the one hand, explain the “nonworking” of artificial genetic constructions during experiments, when they should work according to the calculations *in silico*.** On the other hand, these mechanisms are of significant evolutionary importance and require further study.

4. Conclusion

The “Haploid Evolutionary Constructor” (HEC) software provides modeling of evolutionary and population processes in prokaryotic communities adjusted for the genetic structure of the population, trophic relationships between populations and the influence of environmental conditions (Figure 13). The special feature of the HEC is the approach that enables the modeling of structure variable systems, which in fact provides for the possibility to vary the number of populations, genes, and other variables and parameters immediately during simulations. This makes it possible to model the processes of the gene loss and horizontal transfer between cells, as well as the companion processes of speciation. Together, they open up possibilities for the modeling of bacterial community evolution and their population and ecological dynamics.

We have compared two trophic strategies of prokaryotes: compensatory and non-compensatory nutrition. It has been shown that compensatory systems are more stable under hard environmental conditions. The beneficial mutations in such prokaryotic communities often extend the lifetime of the community significantly. The compensatory systems are more stable in continuously varying conditions from the viewpoint of preserving and growing biomass, while the non-compensatory systems are more stable when it comes to conserving biodiversity [54]. In this way, both strategies have their evolutionary advantages and disadvantages. Neither of them dominate absolutely.

We have studied the importance of gene horizontal transfer and loss during the evolution of the prokaryotic communities depending on ecological conditions. The models have shown the genesis of autonomous populations with rich intrinsic cellular metabolism in pessimal conditions. However, their genome is unstable, and the metabolically complete populations lose genes if conditions improve. This evolutionary tendency may be overcome by the addition of phages to the system. Moreover, the result is of a stochastic nature. We suggest that this shows the potential role of bacteriophages in the genesis of eukaryotes [55].

The study of a trigger-type gene network (two mutually repressing operons) has shown that special characteristics of intracellular factors are able to stabilize states that are defined as being unstable during the mathematical analysis of continuous gene network models. Therefore, alongside ecological and population genetics modeling, the HEC can be used for research on the competition and evolution of gene networks in populations, as well as for the optimization of gene network parameters for certain environmental conditions. The gene network model can be described as a synthesis strategy, while the criterion of optimality and selection mechanisms are trophic strategies.

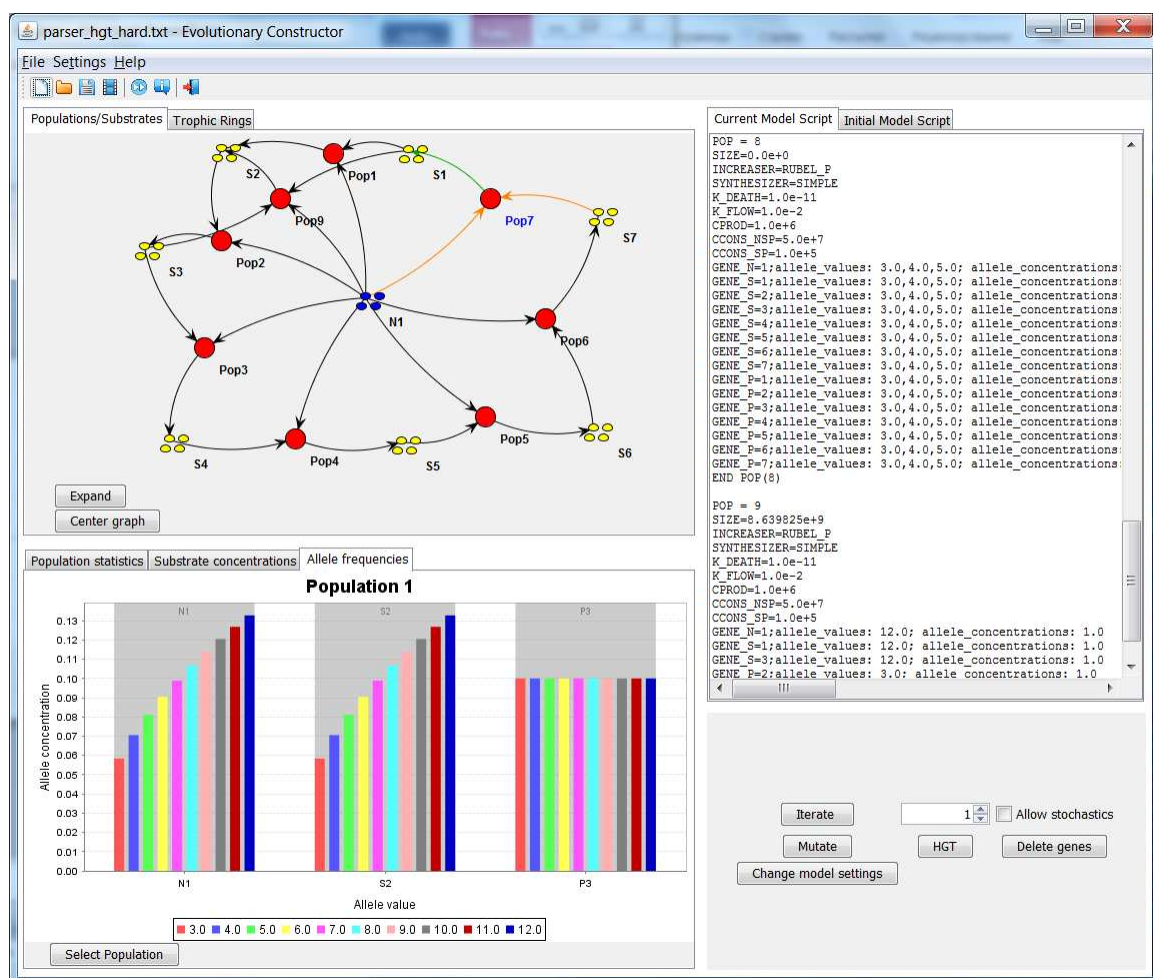


Figure 13. Graphical user interface of the HEC.

Abbreviations

HEC – Haploid Evolutionary Constructor

HGT – Horizontal Gene Transfer

Acknowledgements

This work is supported by the following grants: VI.61.1.2, Russian Foundation for Basic Research (12-07-00671, 13-04-00620), SB RAS integration project №47.

Author details

Sergey A. Lashin^{1,2*}, Yury G. Matushkin¹, Alexandra I. Klimenko^{1,2}, Valentin V. Suslov¹ and Nikolay A. Kolchanov^{1,2}

*Address all correspondence to: lashin@bionet.nsc.ru

1 Department of Systems Biology, Institute of Cytology and Genetics, Novosibirsk, Russia

2 Faculty of Natural Science, Novosibirsk State University, Novosibirsk, Russia

References

- [1] Kamshilov MM. Evolution of the Biosphere. Mir Publishers; 1976. p. 269.
- [2] Margalef R. The Looks of the Biosphere. Moscow: Nauka; 1992.
- [3] Zherikhin V V. Biocoenotic regulation of evolution. Paleontol. J. 1986;(1):3–12.
- [4] Odum EP. Basic ecology. Saunders College Pub.; 1983 [cited 2013 Nov 5]; Available from: <http://agris.fao.org/agris-search/search/display.do?f=1989/US/US89468.xml;US8818555>
- [5] Begon M, Harper J, Townsend C. Ecology. In 2 volumes Moscow: Mir. 1989 [cited 2013 Nov 5]; Available from: http://scholar.google.ru/scholar?cluster=6815279472378442327&hl=ru&as_sdt=2005&sciodt=0,5#0
- [6] Vernadsky V. Chemical Structure of the Earth's Biosphere and Its Environment. 1965 [cited 2013 Nov 5]; Available from: http://scholar.google.ru/scholar?cluster=3535744253151054862&hl=ru&as_sdt=2005&sciodt=0,5#0

- [7] Zavarzin G a. Prokaryotic biosphere. Moscow Univ. Biol. Sci. Bull. [Internet]. 2007 Dec [cited 2013 Jan 31];62(4):141–52. Available from: <http://www.springerlink.com/index/10.3103/S0096392507040013>
- [8] Pianka ER. Evolutionary Ecology [Internet]. 2011 [cited 2013 Nov 5]. Available from: <http://www.google.ru/books?hl=ru&lr=&id=giFL5bonGhQC&pgis=1>
- [9] Allen AP, Gillooly JF. Towards an integration of ecological stoichiometry and the metabolic theory of ecology to better understand nutrient cycling. Ecol. Lett. [Internet]. 2009 May [cited 2013 Nov 5];12(5):369–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19379132>
- [10] Razumovsky S. On the Dynamics of Biogeocenoses. 1981 [cited 2013 Nov 5]; Available from: http://scholar.google.ru/scholar?cluster=15794047151490741694&hl=ru&as_sdt=2005&sciodt=0,5#0
- [11] Ponomarenko A. Evolution of continental aquatic ecosystems. Paleontol. J. C/C ... [Internet]. 1996 [cited 2013 Nov 5]; Available from: <http://www.palaeontomolog.ru/Publ/PALJ705.pdf>
- [12] Kindaichi T, Ito T, Okabe S. Ecophysiological Interaction between Nitrifying Bacteria and Heterotrophic Bacteria in Autotrophic Nitrifying Biofilms as Determined by Microautoradiography-Fluorescence In Situ Hybridization. Appl. Environ. Microbiol. [Internet]. 2004 Mar 8 [cited 2013 Nov 5];70(3):1641–50. Available from: <http://aem.asm.org/content/70/3/1641.short>
- [13] Savvichev AS, Rusanov II, Rogozin DY, Zakharova EE, Lunina ON, Bryantseva IA, et al. Microbiological and Isotopic-Geochemical Investigations of Meromictic Lakes in Khakasia in Winter. Microbiology [Internet]. 2005 Jul [cited 2013 Nov 5];74(4):477–85. Available from: <http://link.springer.com/10.1007/s11021-005-0092-x>
- [14] Garnova ES, Zhilina TN, Tourova TP, Lysenko AM. Anoxynatronum sibiricum gen.nov., sp.nov. alkaliphilic saccharolytic anaerobe from cellulolytic community of Nizhnee Beloe (Transbaikal region). Extremophiles [Internet]. 2003 Jun [cited 2013 Nov 5];7(3):213–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12768452>
- [15] Zavarzin G. Microbial Cycles. Glob. Ecol. [Internet]. 2010 [cited 2013 Nov 5]; Available from: <http://www.google.com/books?hl=ru&lr=&id=JzffxAwj36IC&oi=fnd&pg=PA183&ots=rBbXTYb-WEy&sig=Y1QL4aWM3Qc2Ug6iT0XScvj1aGU>
- [16] Zavarzin G a. A planet of bacteria. Her. Russ. Acad. Sci. [Internet]. 2008 Aug 10 [cited 2013 Jan 31];78(2):144–51. Available from: <http://www.springerlink.com/index/10.1134/S1019331608020056>
- [17] Daniel R. The metagenomics of soil. Nat. Rev. Microbiol. [Internet]. 2005 Jun [cited 2013 Nov 4];3(6):470–8. Available from: <http://dx.doi.org/10.1038/nrmicro1160>
- [18] Riesenfeld CS, Schloss PD, Handelsman J. METAGENOMICS: Genomic Analysis of Microbial Communities. Annual Reviews; 2004 Nov 29 [cited 2013 Nov 5]; Available

from: <http://www.annualreviews.org/doi/abs/10.1146/annurev.genet.38.072902.091216?journalCode=genet>

- [19] Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature* [Internet]. Macmillian Magazines Ltd.; 2000 May 18 [cited 2013 Nov 5];405(6784):299–304. Available from: <http://dx.doi.org/10.1038/35012500>
- [20] Shestakov S. On the Early Biological Evolution from the Viewpoint of Genomics. *Palaeontol. J.* [Internet]. 2003 [cited 2013 Nov 5];(6):50–7. Available from: http://www.maik.ru/abstract/paleng/3/paleng6_3p609abs.htm
- [21] Gogarten JP, Townsend JP. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* [Internet]. Nature Publishing Group; 2005 Sep [cited 2013 Oct 30];3(9):679–87. Available from: <http://dx.doi.org/10.1038/nrmicro1204>
- [22] Rowe-Magnus DA, Guerout A-M, Mazel D. Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Mol. Microbiol.* [Internet]. 2002 Mar [cited 2013 Nov 5];43(6):1657–69. Available from: <http://doi.wiley.com/10.1046/j.1365-2958.2002.02861.x>
- [23] Ilyina TS. Bacterial Superintegrons, a Source of New Genes with Adaptive Functions. 2006;42(11):1294–302.
- [24] Amann R, Ludwig W, Schleifer K. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* [Internet]. 1995 Mar 1 [cited 2013 Nov 5];59(1):143–69. Available from: <http://mmbr.asm.org/content/59/1/143.short>
- [25] Pielou EC. An introduction to mathematical ecology [Internet]. 1969 [cited 2013 Nov 5]. Available from: http://books.google.ru/books/about/An_introduction_to_mathematical_ecology.html?id=apprpg3cdfgC&pgis=1
- [26] Murray JD. *Mathematical Biology®: I. An Introduction*, Third Edition. 3rd ed. Antman SS, Mardsen JE, Sirovich L, Wiggins S, editors. New York: Springer; 2002. p. 551.
- [27] Ewens W. *Mathematical population genetics: I. Theoretical introduction*. 2004 [cited 2013 Nov 5]; Available from: <http://www.google.com/books?hl=ru&lr=&id=twXIyX-yod2MC&oi=fnd&pg=PR7&ots=DMOTg871XN&sig=8mpfOFr83jewa-bR1OKPCr3a28s>
- [28] Neuhauser C. *Mathematical Models in Population Genetics* [Internet]. *Handb. Stat. Genet.* Third Ed. John Wiley & Sons, Ltd; 2004 [cited 2013 Nov 5]. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/9780470061619.ch22/summary>
- [29] Weir BS. *Genetic data analysis. Methods for discrete population genetic data*. Sinauer Associates, Inc. Publishers; 1990 [cited 2013 Nov 5]; Available from: <http://www.cabdirect.org/abstracts/19900180990.html;jsessionid=D364ACD3FCDB6360579FC4C81317E39F>

- [30] Balloux F. EASYPOP (Version 1.7): A Computer Program for Population Genetics Simulations. *J. Hered.* [Internet]. 2001 May 1 [cited 2013 Nov 5];92(3):301–2. Available from: <http://jhered.oxfordjournals.org/content/92/3/301.short>
- [31] Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* [Internet]. 2005 Jan [cited 2013 Oct 31];1:47–50. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2658868&tool=pmcentrez&rendertype=abstract>
- [32] Moskaleichik FF. Normal and Adverse Genetic Processes in Subdivided Populations of the Island Type: Computer Simulation. *Russ. J. Genet.* [Internet]. 2004 Aug [cited 2013 Nov 5];40(8):938–44. Available from: <http://link.springer.com/10.1023/B:RUGE.0000039729.14117.e0>
- [33] Semovski S., Verheyen E, Sherbakov D. Simulating the evolution of neutrally evolving sequences in a population under environmental changes. *Ecol. Modell.* [Internet]. 2004 Aug [cited 2013 Jan 31];176(1-2):99–107. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0304380003005593>
- [34] DeAngelis DL, Gross LJ. Individual-based models and approaches in ecology: populations, communities and ecosystems. Chapman & Hall; 1992 [cited 2013 Nov 5]; Available from: <http://www.cabdirect.org/abstracts/19931977135.html>
- [35] Grimm V, Railsback SF. Individual-based Modeling and Ecology [Internet]. 2005 [cited 2013 Nov 5]. Available from: <http://www.google.ru/books?hl=ru&lr=&id=12MvUbMeog8C&pgis=1>
- [36] Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, et al. A standard protocol for describing individual-based and agent-based models. *Ecol. Modell.* [Internet]. 2006 Sep [cited 2013 Jan 29];198(1-2):115–26. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0304380006002043>
- [37] Lashin S a, Suslov V V, Kolchanov N a, Matushkin YG. Simulation of coevolution in community by using the “Evolutionary Constructor” program. *In Silico Biol.* [Internet]. 2007 Jan;7(3):261–75. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18415976>
- [38] Lashin S a, Matushkin YG. Haploid evolutionary constructor: new features and further challenges. *In Silico Biol.* [Internet]. 2012 [cited 2013 Jan 31];11(3-4):125–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22935966>
- [39] Rübel E. The replaceability of ecological factors and the law of the minimum. *Ecology* [Internet]. 1935 [cited 2013 Nov 5]; Available from: <http://www.jstor.org/stable/10.2307/1930073>
- [40] Sundararaj S, Guo A, Habibi-Nazhad B, Rouani M, Stothard P, Ellison M, et al. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of *Escherichia coli*. *Nucleic Acids Res.* [In-

- ternet]. 2004 Jan 1 [cited 2013 Nov 5];32(Database issue):D293–5. Available from: http://nar.oxfordjournals.org/content/32/suppl_1/D293.short
- [41] Bennett K. Milankovitch cycles and their effects on species in ecological and evolutionary time. *Paleobiology* [Internet]. 1990 [cited 2013 Nov 5]; Available from: http://marineecology.wcp.muohio.edu/climate_projects_04/glacial_cycles/web/pdf/MilankovitchCycles.pdf
- [42] Dobretsov N, Kolchanov N, Suslov V. On important stages of geosphere and biosphere evolution. *Biosph. Orig. Evol.* [Internet]. 2008 [cited 2013 Nov 5]; Available from: http://link.springer.com/chapter/10.1007/978-0-387-68656-1_1
- [43] Levchenko VF, Starobogatov YI. Origin of biosphere. *J. Evol. Biochem. Physiol.* [Internet]. 2010 May 20 [cited 2013 Jan 31];46(2):213–26. Available from: <http://www.springerlink.com/index/10.1134/S0022093010020122>
- [44] Markov A, Kulikov A. Origin of Eukaryota: Conclusions Based on the Analysis of Protein Homologies in the Three Superkingdoms. *Paleontol. J.* [Internet]. 2005 [cited 2013 Nov 5]; Available from: <http://www.maikonline.com/maik/showArticle.do?pii=S0031030105040015&lang=en>
- [45] Markov A V., Kulikov AM. Homologous Protein Domains in Superkingdoms Archaea, Bacteria, and Eukaryota and the Problem of the Origin of Eukaryotes. *Biol. Bull.* [Internet]. 2005 Jul [cited 2013 Nov 5];32(4):321–30. Available from: <http://link.springer.com/10.1007/s10525-005-0108-0>
- [46] Cavalier-Smith T. Principles of Protein and Lipid Targeting in Secondary Symbiogenesis: Euglenoid, Dinoflagellate, and Sporozoan Plastid Origins and the Eukaryote Family Tree, 2. *J. Eukaryot. Microbiol.* [Internet]. 1999 Jul [cited 2013 Nov 5];46(4):347–66. Available from: <http://doi.wiley.com/10.1111/j.1550-7408.1999.tb04614.x>
- [47] Olson M V. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* [Internet]. 1999 Jan [cited 2013 Nov 5];64(1):18–23. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1377697&tool=pmcentrez&rendertype=abstract>
- [48] Ochman H. Genes Lost and Genes Found: Evolution of Bacterial Pathogenesis and Symbiosis. *Science* (80-.). [Internet]. 2001 May 11 [cited 2013 Nov 5];292(5519):1096–9. Available from: <http://www.sciencemag.org/content/292/5519/1096.short>
- [49] Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* [Internet]. 1961 [cited 2013 Nov 5];3(3):318–56. Available from: <http://www.sciencedirect.com/science/article/pii/S0022283661800727>
- [50] Ebeling W, Schimansky-Geier L, Romanovsky YM. Stochastic Dynamics of Reacting Biomolecules [Internet]. 2002 [cited 2013 Nov 5]. Available from: <http://www.google.ru/books?hl=ru&lr=&id=ieVmkC6tY0MC&pgis=1>

- [51] Riznichenko GY. Mathematical models in biophysics [Internet]. 2012. Available from: <http://www.biophysics.org/Portals/1/PDFs/Education/galina.pdf>
- [52] Tchuraev RN. On a stochastic model of a molecular genetic system capable of differentiation and reproduction of the initial state. Biometrical J. [Internet]. 1980 [cited 2013 Nov 5];22(2):189–94. Available from: <http://doi.wiley.com/10.1002/bimj.4710220212>
- [53] Tchuraev R., Stupak I., Tropynina T., Stupak E. Epigenes: design and construction of new hereditary units. FEBS Lett. [Internet]. 2000 [cited 2013 Nov 5];486(3):200–2. Available from: <http://www.sciencedirect.com/science/article/pii/S0014579300023000>
- [54] Lashin S a., Suslov V V., Matushkin YG. Comparative Modeling of Coevolution in Communities of Unicellular Organisms: Adaptability and Biodiversity. J. Bioinform. Comput. Biol. [Internet]. 2010 Jun [cited 2013 Jan 31];08(03):627–43. Available from: <http://www.worldscientific.com/doi/abs/10.1142/S0219720010004653>
- [55] Lashin S a., Matushkin YG, Suslov V V., Kolchanov N a. Evolutionary trends in the prokaryotic community and prokaryotic community-phage systems. Russ. J. Genet. [Internet]. 2011 Dec 8 [cited 2013 Jan 31];47(12):1487–95. Available from: <http://www.springerlink.com/index/10.1134/S1022795411110123>