Journal of Bioinformatics and Computational Biology
Vol. 8, No. 3 (2010) 627–643

DOI: 10.1142/S0219720010004653

Imperial College Press
www.icpress.co.uk

COMPARATIVE MODELING OF COEVOLUTION IN COMMUNITIES OF UNICELLULAR ORGANISMS: ADAPTABILITY AND BIODIVERSITY

SERGEY A. LASHIN*, VALENTIN V. SUSLOV† and YURI G. MATUSHKIN‡

Department of Systems Biology
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Faculty of Natural Science
Novosibirsk State University
Novosibirsk, Russia
**lashin@bionet.nsc.ru*
†*valya@bionet.nsc.ru*
‡*mat@bionet.nsc.ru*

Received 19 October 2009
Revised 29 December 2009
Accepted 16 January 2010

We propose an original program "Evolutionary constructor" that is capable of computationally efficient modeling of both population-genetic and ecological problems, combining these directions in one model of required detail level. We also present results of comparative modeling of stability, adaptability and biodiversity dynamics in populations of unicellular haploid organisms which form symbiotic ecosystems. The advantages and disadvantages of two evolutionary strategies of biota formation — a few generalists' taxa-based biota formation and biodiversity-based biota formation — are discussed.

Keywords: Evolution; modeling; bacterial community; horizontal gene transfer; symbiosis.

1. Introduction

Since Charles Darwin formulated his principles of natural selection, the problem of relationship between diversity of living systems (biodiversity) and their evolutionary success has become fundamental in biology. Evolution indeed can be defined as the process of transformation of genetic variability of individuals into variability of a group — taxon.[1,2] Variability occurs at the level of individual genome — as a result of mutations and recombination[2] and then is tested by a selection in groups of two types: in populations[3,2] and ecosystems.[4–6] In ecosystems various species

*Corresponding author.

are organized in linear or branching trophic chains.[7] Conjugation of several linear chains or branches gives rise to trophic cycles, which support the recovery of a certain resource.[7,8]

Metabolic conjugation is a characteristic feature of prokaryotic ecosystems in the range from a close association of organisms belonging to different species[9] to trophically sealed ecosystems of meromictic[10] or soda[11] lakes and, eventually, to global biogeochemical cycles.[8,12] This fact is related to specific prokaryotic nutrition (pinotrophy) and the organization of prokaryotic cells.[8] Closed trophic chains in which some metabolites being synthesized and secreted by one species or strain can be consumed by another species are called *trophic rings*. Such rings can be found in bacterial films and in complex metabolism graphs constructed in metagenomic studies.[13–17] The evolutionary success of a certain allele is related to both its fixation in the population and the influence of this population on the operation of the whole ecosystem. Thus any subspecies group has some ecocenotic features.[5,18–20]

In the course of investigation of the actual evolutionary process, scientists recognize combined groups of organisms: cenopopulations, hemipopulations, guilds, synusiae, etc. In this way, the comprehensiveness of description can be varied. However, evolutionary mathematical modeling methods lack this flexibility. Most such models consider in detail only one level: an individual, together with its metabolism or development; a population; or an ecosystem.[21] The advantageous models consider all or at least two of these levels and primarily use detailed individual-oriented modeling.[22–24] The individual-oriented approach allows consideration of many factors at the expense of low computational efficiency and predetermination of the model, because a change in its structure requires construction of a new model.

In 2007, we developed a modeling method implemented in the program package Evolutionary Constructor (EC). It allows description of objects to be modeled and change of the model structure in the course of model design.[25,26]

We applied EC to *in silico* studies of evolutionary changes in trophic rings: communities of populations of model unicellular haploid organisms possessing genes for synthesis and substrate consumption. These organisms were placed to a flow that would deliver a non-specific substrate consumed by all the organisms (see Sec. 2). The evolutionary consequences of horizontal gene transfer and adaptation to shortage of the nonspecific resource were studied in trophic rings whose populations employed various trophic strategies.

2. "Evolutionary Constructor" Modeling Approach

Taking into account both genetic and ecocenotic factors, we propose the EC modeling approach previously described and being developed in the present paper to be an effective tool for modeling the evolution of unicellular haploid asexual organisms communities. In consideration of modeling object — bacterial populations particularities, such approach should allow to model as communities of high size (10^9–10^{12} and more individuals) so genetic variability.

Genetically identical or almost identical (with allele variations) individuals form *populations*. Populations live in common flow system of fixed volume V_{total} — *environment*, which intermediates relationships between them. Individuals *consume* and *utilize* substrates, *synthesize* and then *secrete* products (partially or completely) into environment where they can be consumed by other individuals as substrates. Some substrates activate population growth while others may inhibit it. The efficiencies of substrate utilization and product synthesis are regulated by corresponding genes on the principle of "one gene — one constant of utilization/synthesis".

Substrates are divided into *non-specific substrates* which are supplied only by inflow (N_i in Fig. 1), and specific substrates which are supplied only by virtue of individual synthesis and secretion (S_i in Fig. 1). Figure 1 shows the main processes and objects of the EC model. During the model calculation graph of substrates-population interactions can change — here may occur both novel vertices (arise of novel populations or substrates), and edges (arise of novel trophic interaction).

2.1. *Environment modeling*

Environment is characterized by the following parameters:

V_{total} — volume of environment (liters);
k_{flow} — flow rate (% of V_{total} per time unit);
$N_{env,i}$ — concentration of i-th nonspecific substrate in environment (mM);
$S_{env,i}$ — concentration of i-th specific substrate in environment (mM);
$N_{flow,i}$ — concentration of i-th nonspecific substrate in inflow (mM).

Inflow of substrates into environment increases the concentration of non-specific substrates, depending on flow rate and corresponding substrate inflow concentration. *Outflow of substrates from environment* decreases the concentration of both non-specific and specific substrates, depending on flow rate. Change in non-specific

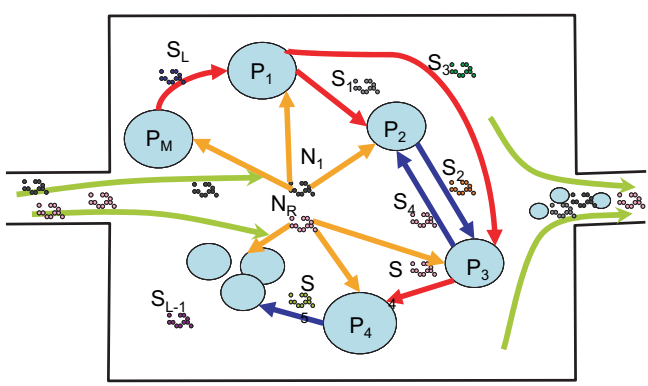

Fig. 1. Main processes and objects of EC model. Red arrows — specific activatory substrate consumption, blue arrows — specific inhibitory substrate consumption, orange arrows — non-specific substrate consumption.

substrates concentration under the action of inflow and outflow is described by the following equation:

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

Outflow of specific substrates if described by the equation below:

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

See Sec. 2.3 for changes in substrate concentration under the action of individual.

2.2. *Populations modeling*

In EC approach individuals are defined as the members of the same population if they consume equal sets of non-specific and specific substrates, synthesize equal sets of specific products, and utilize substrates with the same utilization law (i.e. have the same *trophic strategy*). In EC we assume each process of synthesis and utilization process to be defined by only one *gene*. Gene in this case is considered a heredity unit, with the corresponding process determined by the reaction rate constant. Then *allele* is a gene variant which determines the actual value of corresponding rate constant, and *individual's genotype* is a set of such alleles divided into three groups (vectors). First group (c_i) determines efficiencies of specific substrates' utilization, second group (d_i) — efficiencies of specific products' synthesis/secretion and third group (r_i) – efficiencies of non-specific substrates' utilization. *Mutation* is a change of corresponding constant value which is interpreted as gene substitution.

2.2.1. *Monomorphic populations*

A monomorphic population is univalently determined by three characteristics: *genotype*, *size* (which can be interpreted as a number of cells in a population or cells concentration), and *amount of consumed substrates*. Trophic process of an individual in EC is described at two separate stages: *consumption* (from environment) of substrates and utilization of substrates. During the first stage an uptake of substrates from environment into an individual is modeled; these substrates are utilized during the second stage. As these processes are physically widely different, we model them separately. The amount of substrate consumed depends on its concentration in the environment and the individual's size (cell surface area). Under substrate environmental concentration is low, both inter- and intra-population competition can occur. In the case of substrate excess, the maximum amount which can be consumed by an individual is determined by *maximum substrate consumption rate* that corresponds with, for instance, cell size constraints. Replenishment of intercellular substrates can be provided both by the cell's own synthesis and "substrates import" from environment. A detailed description of substrates consumption stage is presented in Sec. 2.3.

The equations of population size change per one generation — ΔP in dependence on population size, amount of substrates consumed, flow rate and death rate are shown below. These equations are *trophic strategies* of populations[1]:

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

$$\begin{aligned}\Delta P &= F_2(\vec{N}, \vec{S}, \vec{R}, \vec{C}, P) \\ &= P \cdot \frac{\left(\frac{\frac{n_0}{P}}{K_{01}(r_0)}\right)^{\gamma_0}}{1 + \left(\frac{\frac{n_0}{P}}{K_{02}(r_0)}\right)^{\gamma_0}} \cdot \prod_{i \in I_{consumed}} \frac{1 + \left(\frac{\frac{s_i}{P}}{K_{i1}(c_i)}\right)^{\gamma_i}}{1 + \left(\frac{\frac{s_i}{P}}{K_{i2}(c_i)}\right)^{\gamma_i}} \\ &\quad - k_{flow} \cdot P - k_{death} \cdot P^2 \end{aligned} \tag{4}$$

where $I_{consumed}$ is the set of indices of specific substrate consumed by population individuals; $\vec{N}$ and $\vec{S}$ are the respective vectors of amounts of non-specific/specific substrates consumed by population individuals; n_0 is the amount of the unique non-specific substrate consumed by population individuals (as the component of $\vec{N}$); r_0 is the utilization rate of the unique nonspecific substrate (genetically determined); $\vec{R}$ and $\vec{C}$ are the respective vectors of utilization rates of non-specific/specific substrates (genetically determined); P is population size; k_{flow} is flow (washout) rate; k_{death} is death rate; $\gamma, \gamma_0, \gamma_i$ are coefficients of nonlinearity of substrates on population growth; K_{ij} is coefficient of substrates effect on the population growth (depend on corresponding traits). Mortality parameter accords with self-poisoning effect in bacterial populations. This parameter is stereotyped in models of limited population dynamics.[27] A small amount of substrate does not yield population growth and while increases in this amount at a certain range leads to increase of population growth rate, further increases do not give effect to this rate, which remains constant. This nonlinearity is described by Eq. (4). Parameters $\gamma, \gamma_0, \gamma_i$ determine population growth rate increase in corresponding phase (the larger the parameters, the greater the increase); K_{ij} determines substrates concentrations at which growth rate acceleration starts and finishes (like an analogue of Michaelis-Menten constant in enzyme kinetics).

Equation (3) describes the utilization process of several specific and one non-specific substrate. Specific substrates have strong cooperative positive effect. In particular, the deficiencies of one substrate (including non-specific ones) can be compensated by the excess of another. Trophic strategy (3) corresponds to *Rubel's*

[1]Some other trophic strategies are presented in Refs. 25 and 26. Actually, the choice of trophic strategy is the user's prerogative. The only constraint in the EC approach is the taking into consideration of all four characteristics. So while EC can work with wide types of trophic strategies described in literature, it requires isomorphic modification of formulas if one or some characteristic(s) is/are implicit.

law of replaceability of ecological factors[7,28] and is called *compensatory trophic strategy*. Equation (4) describes trophic strategy corresponding to ecological *Liebig's law of the minimum*[7] — deficiency of non-specific substrate cannot be compensated by the excess of specific ones (*noncompensatory trophic strategy*). Our approach also provides usage of other trophic strategies, for instance, strategies of mutual poisoning between populations, which may lead to unexpected chaotic-like behavior of a system.[26]

2.2.2. *Polymorphic populations*

Individuals of a *polymorphic population* have the same set of genes, but their alleles may vary in one or more genes. Polymorphic population is determined by its "*generalized genome*" — multi-dimensional distribution of alleles' frequencies for all genes of its individuals. In order to operate with such distribution, we introduce the notion of *genetic spectrum* — the distribution of alleles' frequencies for one gene (Fig. 3). As we define allele as a numerical value of a trait — rate constant of corresponding reaction (see Sec. 2.2) — the allele numbers are shown in Fig. 2 at X axis, with each number corresponding to a certain constant value. Proportion of each allele is shown at Y axis.

An aggregate of spectrums for all population genes considered as a vector V_{GS} is used for modeling multidimensional alleles distribution for all genes. In our model we assume linkage equilibrium for each gene, i.e. we assume all alleles for all genes to be distributed independently. In such a case we may consider polymorphic population as a set of monomorphic subpopulations. The size of such subpopulation corresponds with product of frequencies of the alleles, determining that subpopulation. Change of polymorphic population size is calculated in the following manner. First of all it is "*spliced*" into a set of monomorphic subpopulations. After that the size change calculation procedure is applied for each of them, with the use

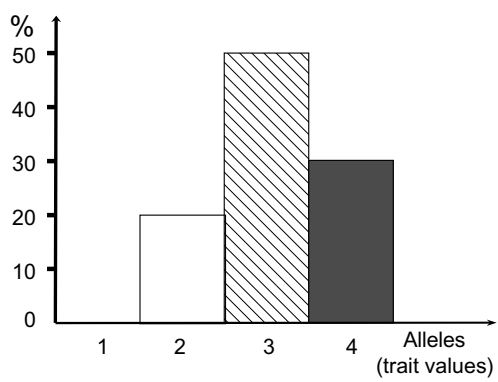

Fig. 2. Example of genetic spectrum. 20% individuals of the population have the allele value of 2 units; 50%, 3 units; and 30%, 4 units.

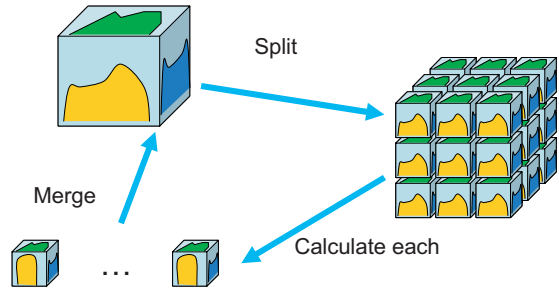

Fig. 3. Calculation of polymorphic population (big cube) size change. Polymorphic population is "split" into a set of monomorphic ones (small cubes). Then we calculate the size change for each monomorphic population and finally "merge" them again into one polymorphic population.

of trophic strategies equations (see Sec. 2.2.1). Finally the changed monomorphic subpopulations are "*merged*" into a polymorphic population (Fig. 3). It should be noted that various combinations of alleles have various fitness. So the growth of a certain monomorphic subpopulation can greatly differ from that of the others. As a result, the genetic spectrums of a polymorphic population change (right up to alleles elimination), which is interpreted as adaptation to environment.

Polymorphic population allows us foremost to model genetic polymorphism and, in the next phase, to vary the level of detail of modeled objects with the necessary flexibility and computational efficiency needed to pass/progress/move/transit from the population-genetic to the ecocenotic level. For this purpose we suggest *threshold value* of a trait in a population. Threshold value defines the lowest trait value, whereby gene is supposed to be "working". In the case of substrate utilization gene it means that population cannot utilize the substrate when it has a corresponding trait value (allele) lower than the threshold.[29] In the case of synthesis gene this means incapability to synthesize a corresponding product. Moreover, we can, at first, subdivide a polymorphic population in a number of ways according to ecological features modeled (Fig. 4(b)); secondly we can join populations or their parts together to consider them as one object. The first case accords with separation of cenopopulations, the second with separation of life forms, guilds, etc.[6,30]

Using genetic spectra we also modeled *mutations* and *horizontal gene transfer* (HGT). Mutation means a change of trait value in one or several individuals. It is modeled through the change of corresponding genetic spectrum profile (not excepting new alleles occurrence) (Fig. 4(a)).

HGT is the transfer of DNA piece from genome of a donor cell and its embedding into an acceptor cell. In our approach HGT is modeled as the sequence of the following steps:

— choice of monomorphic subpopulation P_{MD} from population-donor. P_{MD} has genome $V_{\mathrm{D}} \in V_{\mathrm{GSD}}$ (V_{GSD} is vector of genetic spectra at population-donor);
— choice of subset (vector) $V_{\mathrm{T}} \in V_{\mathrm{D}}$ of genes to be transferred;

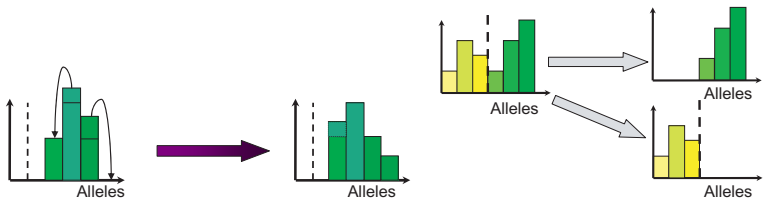

Fig. 4. (a) Mutations modeling — changes of genetic spectrum. (b) Splitting a population with the threshold value relative to the spectrum of a particular gene. Two populations arise. Their spectra are uniform in relation to the threshold value.

— choice of monomorphic subpopulation P_{MA} from population-acceptor, which has genome $V_{\mathrm{A}} \in V_{\mathrm{GSA}}$ (V_{GSA} is vector of genetic spectra at population-acceptor);
— construction of genome of novel type: $V_{\mathrm{N}} = V_{\mathrm{A}} \cup V_{\mathrm{T}}$ (if there are genes present both in V_{A} and V_{T}, trait values are taken from V_{A}, i.e. these genes are not transferred);
— evaluation of initial size of novel population (usually from 1 to 1000 individuals, taking into account the size of the whole metapopulation of a system);
— addition of novel population to the list of ones which exist in a system. In the case where there is already a population having the similar set of genes (not necessarily the same alleles) in a system, novel population "joins" the present one. Genetic spectra are "merged" for this purpose proportionally to sizes of novel and old populations (Fig. 5).

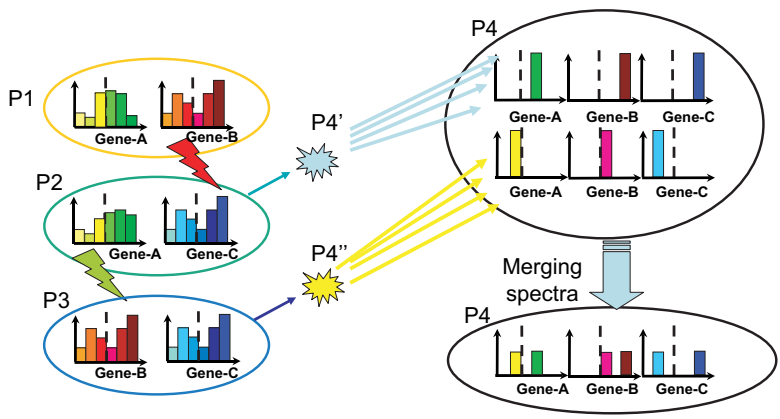

Fig. 5. Modeling of two HGT leading to polyphyletic speciation of population P4. Transfer of gene B from an individual of P1 into an individual of P2 leads to the arising of a new type individual which contains genes A, B and C and then forms novel population P4'; Second transfer — gene A from an individual of P2 into an individual of P3 forms population P4" containing genes A, B and C (alleles differ from P4'). In fact, both novel populations should be considered as the only polymorphic population P4.

Therefore, in EC we can model both mono- and polyphyletic evolution[31] (Fig. 5). Processes of mutations and HGT in EC may be predefined (by virtue of model scenarios, written with special script language) or generated stochastically. Operations with genetic spectra described above compose so-called *genetic spectra arithmetic*, a key feature of EC. The same mathematical apparatus of genetic spectra arithmetic provides modeling of population-genetic and ecological problems, combining both directions in one model of required detail level.

2.3. *Substrates consumption and secretion*

In the current version of EC the process of substrates consumption is modeled on the assumption of prokaryotic pinotrophy whereby metabolic rate directly depends on the ratio of surface area to volume.[32] During the period of cell cycle a certain environmental volume is infiltrated through a cell due to flow and cell movement. That volume V_{consumed} is proportional to the cell's surface area S_{cell}, its relative velocity v_{cell} and its lifetime t.

$$V_{\text{consumed}}(t) = S_{\text{cell}} \cdot v_{\text{cell}} \cdot t \tag{5}$$

Estimation of this variable for *E.coli* is about $5.4 \cdot 10^{-12}$ liters per half-hour (average period of cell cycle): cell surface area is about $6 \cdot 10^{-12}\,\text{m}^2$, movement speed is about 50 micrometer/sec or $9 \cdot 10^{-2}$ meters per half-hour[33]. We used this estimation for V_{consumed} in all numerical experiments. On the assumption of uniform distribution of populations in environment and the same value of V_{consumed} for all individuals, the decrease of substrate concentration $S_{\text{env},i}$ in environment under population consumption is described by the following equation:

$$S_{\text{env},i}(t+1) = S_{\text{env},i}(t) \cdot (1 - P \cdot V_{\text{consumed}}/V_{\text{total}}) \tag{6}$$

where P is the population size (number of cells). The number of substrates molecules $S_{\text{pop},i}$ which population's individuals get with that is calculated using the following equation:

$$S_{\text{pop},i}(t+1) = S_{\text{pop},i}(t) + P \cdot V_{\text{consumed}} \cdot S_{\text{env},i}(t) \cdot N_{\text{A}} \tag{7}$$

where N_{A} is the Avogadro number. If the inequality (8) is satisfied, it means a high number of individuals in an environment and that these individuals overlap each other and cannot "percolate" the whole volume V_{consumed} ($\eta \leq 1$ — normalizing coefficient, I_{POP} — set of all populations in community). In this case in the Eqs. (6) and (7) V_{consumed} should be replaced with $\tilde{V}_{\text{consumed}}$ (9).

$$\sum_{i \in I_{\text{POP}}} P_i \cdot V_{\text{consumed}} > \eta V_{\text{total}} \tag{8}$$

$$\tilde{V}_{\text{consumed}} = \eta V_{\text{total}} \Big/ \sum_{i \in I_{\text{POP}}} P_i \tag{9}$$

In the case where the potential number of substrate molecules which cell can consume exceeds the maximum substrate consumption rate, "surplus" remains in environment and the decrease of substrate concentration is described by Eq. (10) instead of Eq. (6).

$$S_{\mathrm{env},i}(t+1) = S_{\mathrm{env},i}(t) - \frac{P \cdot S_{\mathrm{pop},i}}{V_{\mathrm{total}} \cdot N_A} \tag{10}$$

Products synthesis performed by individuals of polymorphic populations is calculated with the use of the following integral formula:

$$\Delta s_i = P \cdot \sum_{j \in \mathrm{Spectr}_i} d_{i_j} \left(\frac{P_j}{P} \right) \tag{11}$$

where Δs_i is the amount of the synthesized substrate of i-th type; d_{ij} is trait values in genetic spectrum Spectr_i; P is population size; P_j is size of subpopulation having trait value of d_{ij} in relation with the spectrum. The product being synthesized is then secreted into environment. Change of this concentration due to secretion is as follows:

$$S_{\mathrm{env},i}(t+1) = S_{\mathrm{env},i}(t) + \frac{\Delta s_i}{V_{\mathrm{total}} \cdot N_A} \tag{12}$$

2.4. *Iteration process*

Each iteration step of the computation cycle contains the stages of substrate consumption (taking into account competition), utilization of consumed substrates during reproduction process, product synthesis, product secretion and flow simulation. The mutation or HGT stage is not mandatory and is set either deterministically or stochastically (Fig. 6).

2.5. *Implementation*

Current version of EC program is developed using object-oriented paradigm in C++ under Windows and Unix. Program contains the following modules: computational kernel, language of model scenarios, and graphical user interface.

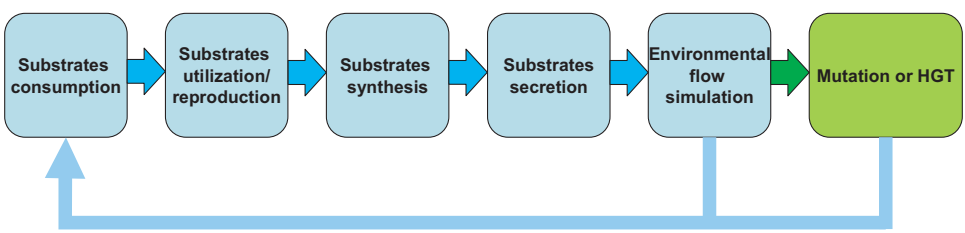

Fig. 6. Flow diagram of one iteration step. Final step (mutation or HGT) is not mandatory.

3. Modeling of Biodiversity and Adaptability Dynamics in Bacterial Communities

We used EC for comparative modeling of stability, adaptability and biodiversity dynamics in trophically closed communities having compensatory (Eq. (3)) and non-compensatory (Eq. (4)) metabolism under sublethal substrate deficiency conditions. Community is represented as a trophic ring-like network (TRLN) which consists of three populations: population P1 utilizes specific substrate S1 being produced by P3 and produces specific substrate S2; P2 utilizes S2 and produces S3 which, in turn, is utilized by P3 (Fig. 7(a)). All populations consume common non-specific substrate N. Flow rate, inflow substrate concentration and other initial data (populations' sizes and substrate concentrations) were chosen in such a way that a short period of population growth is replaced by gradual extinction. Trophic strategies parameters were chosen in such a way that both systems showed the same behavior in the absence of external actions. Mutant-individual had the increased efficiency of specific substrate S1 utilization and begot subpopulation P′1, which competitively displaced "mother" population P1.

It has been shown that mutations fixation in compensatory systems (TRLN-C) causes either significant increase of TRLN-C populations' life-time or saves TRLN-C from extinction (Fig. 7(b)). The size of mother population and its symbionts (which depends on mutation occurrence moment) plays the key role in mutation effect. Mutation occurred just after initial time prevented inevitable extinction of both population-mutant and the whole TRLN-C while later mutation could only extend extinction time (Fig. 7(d)). Similar mutations in non-compensatory systems (TRLN-NC) merely led to local improvements but could not change limiting regime of TRLN-NC functioning, regardless of occurrence moment (Fig. 7(c)).

We also modeled simultaneous functioning and competition of two TRLNs with various combinations of trophic strategies: C–C, NC–NC, C–NC (Fig. 8(a)). We also considered effects of HGT and environmental parameters on systems functioning. Compensatory systems were shown to be more competitive and are characterized by higher biomass growth. However, compensatory systems tend to lose their biodiversity more than non-compensatory ones. In spite of starvation, NC-NC system (Fig. 8(b)) keeps proportions between separate populations sizes. C-C systems are more sensitive to events like HGT or substrate deficiency. Figure 8(c) shows moderately short-time starvation within 300 generations. After the end of this period, the proportions of populations slightly change. A slightly longer starvation period of 360 generations leads to completely different results (Fig. 8(d)) — one of TRNL extincts, besides for long-time period after starvation finish — within about 2500 generations.

4. Discussion

The software package developed by us allows the modeling of evolution of unicellular haploid organisms, which form the greatest part of Earth biota and are the earliest

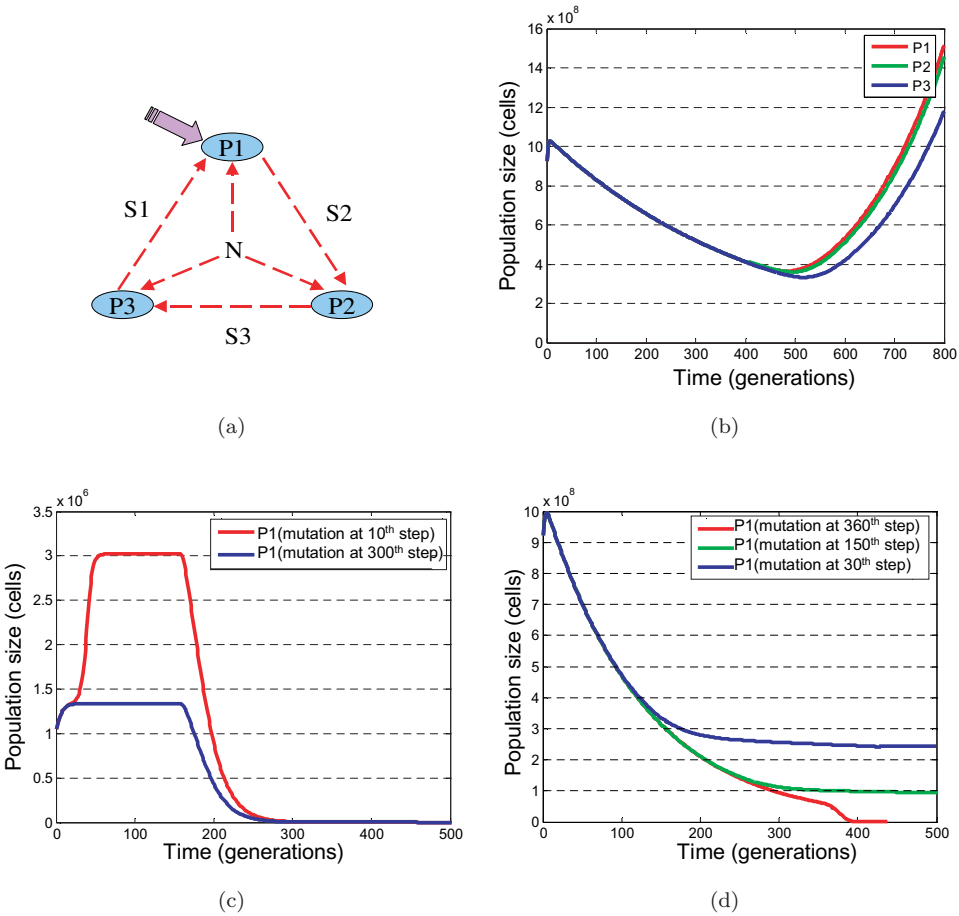

Fig. 7. (a) Diagram of trophic ring-like network (TRLN). Populations P1, P2, P3 are designated as circles. Trophic interactions are shown with dashed lines. Bold arrow shows mutation. (b) TRLN-C dynamics after mutation. Mutation occurs in an individual of P1 population (360th generation), it prevents "extinction tendency" and saves both population P1 and the whole TRLN-C. (c) Population P1 dynamics in two different TRLN-NC (the same initial data) after mutations occurred at 10th and 300th generations, respectively. Neither of them could "save" populations regardless of occurrence moment. (d) Population P1 dynamics in three different TRLN-C (the same initial data — harder than in b case) after mutations occurred at various moments (360th, 150th and 30th steps). Mutations at 30th and 150th steps are shown to delay extinction process (but tendency remains) while mutation at 360th step has almost no effect.

living creatures. Up until the neoproterozoic period (1–3.8 billions years ago) the biosphere remained completely prokaryotic. The total mass of prokaryotes is estimated at 50–90% of the mass of whole biota.[8,11] Cooperating in communities, various prokaryotic taxa form a significant part of the Earth biosphere. They maintain execution of all fundamental global biochemical cycles: sulfur, nitrogen, and a significant part of phosphor cycle. Furthermore, closed metabolic cycles are often formed

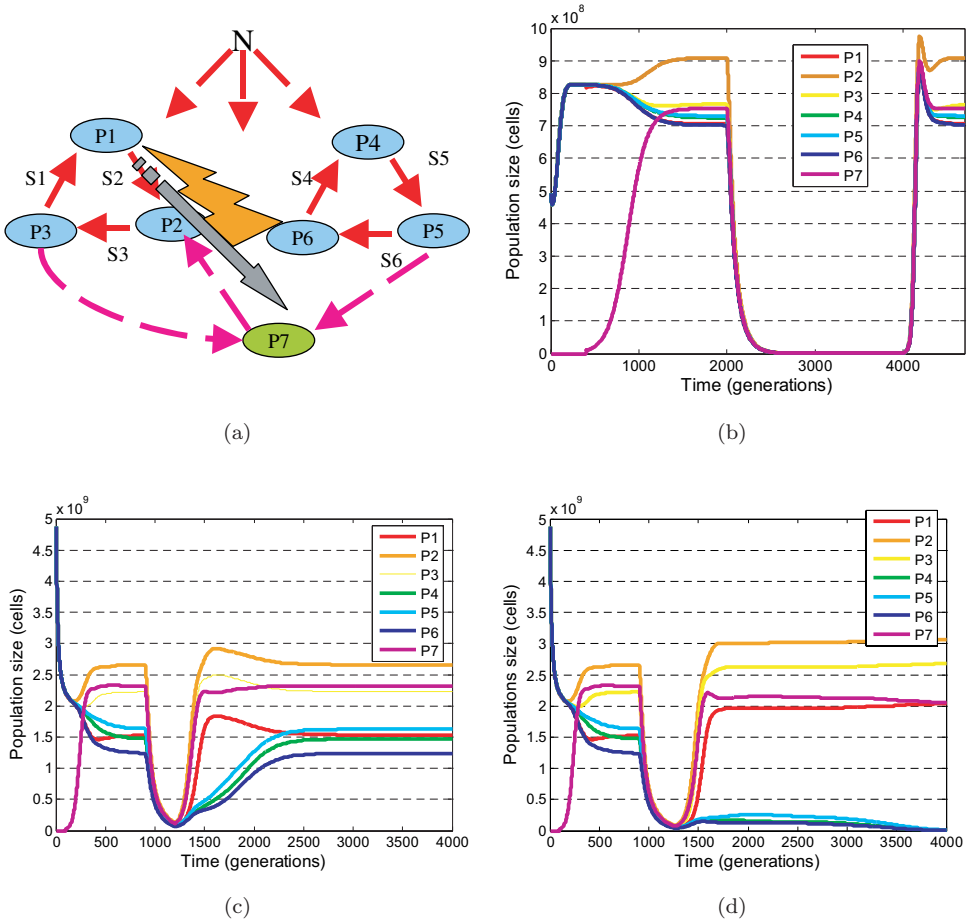

Fig. 8. (a) Diagram of trophic systems containing two TRLNs: P1-P2-P3 and P4-P5-P6. HGT from an individual of P6 into an individual of P1 is also shown. It leads to the arising of novel type population P7. Lightning shows gene transfer of substrate S6 utilization gene; grey arrow – separation of novel population P7. P7 individuals are able to utilize two specific substrates: S1 and S6 (i.e. they connect in some way two trophic rings). They also synthesize and secrete substrate S2 which includes them into acceptor "mother" TRLN. (b) NC-NC system's dynamics after HGT, speciation and consequent starvation during 1700 generations. In spite of a long period of starvation, all populations survived and biodiversity is preserved. (c,d) C-C system's dynamics after HGT, speciation and consequent starvation during 300 (c) and 360 (d) generations. In the first case the system restores completely, in the second biodiversity is partially lost.

in biofilms and mats, which provide high probability of horizontal gene transfer between separate bacteria. For example, about 15–20% of the genomes of thermophilic bacteria *Thermotoga maritima* are typically archaean genes.[34,35] The significance of HGT is confirmed by the recent finding of integrons and superintegrons, natural vectors for cloning alien genes in prokaryotic genomes.[36] The evolution of

such highly-integrated communities has its own qualitative peculiarities and cannot be traced to evolution of separate individuals.

Our approach allows the modeling of the evolution of genetic structure of a large number of populations of unicellular organisms over long periods of time. Except the example described in the present paper — comparison of non-compensatory and compensatory trophism, we also studied the possible evolutionary origin of autonomous bacterial taxa (as a result of prokaryotic cells symbiosis) having rich intracellular metabolism. This corresponds with concepts of eukaryota origin.[37] For this purpose we carried out up to 15000 iterations-generations. The size of final populations varied up to $5 \cdot 10^7$, and the final number of populations was up to 33. The EC software was run on a typical PC (Pentium 4, 3 Ghz, 2 Gb RAM). The algorithm has good abilities for parallelization and provides evolutionary experiments *in silico*, considering populations of natural size and complexity.

The results of modeling presented in the paper indicate that the trophic structure of a community imposes considerable limitation on HGT advantages. Long-term advantage can be achieved only by HGT among populations with initially diverse and flexible metabolism (compensatory feeding). In populations with simple metabolism (non-compensatory feeding), HGT provides only local advantages. This fact contradicts the assumption that the main trend in prokaryotic evolution is simplification of the genome of a particular individual compensated by complex interactions in the bacterial community (metagenome).[31] In the long run, over-simplification will kill the community.

We have shown that a TRLN with compensatory feeding (TRLN-C) was much better adaptable than that with non-compensatory feeding. Fixation of useful mutations even in one TRLN-C significantly increased the stability of the whole community. It either prolonged the life of the community, providing additional chances to pass through starvation, or allowed TRLN-C to overcome the shortage by optimizing the metabolism. Thus, in the long run, non-compensatory Liebig-type systems will gain advantage over compensatory Rübel-type ones. Note that this advantage is stochastic, like all evolutionary processes. Hence, a system can die from stochastic causes even at high biodiversity. This statement is in agreement with paleontological data. The paleontological record demonstrates constant change of leading biotas[5,38] and preservation of a relatively few number of relict ecosystems, which accumulate nearly complete sets of biochemical activities[8] (cyanobacterial mats and alkaliphilic communities[8,11]).

Acknowledgments

The work is supported by the following grants: RAS Program 22. Molecular and cellular biology. Project 8. RAS Program 23. Project 29. Integration project 119. RAS Program 15, Subprogram 2 Biosphere origin and evolution. SciSchool-2447.2008.4. RFBR #10-04-01310.

References

1. Darwin C, *The Origin of Species,* NY, Oxford Univ Press, 1996.
2. Levontin RG, *The Genetic Basis of Evolutionary Change,* NY & London, Columbia Univ Press, 1974.
3. Hedrick PW, *Genetics of Populations,* Jones and Bartlett Publishers, Sudbury, MA, 2000.
4. Kamshilov MM, *Evolution of Biosphere*, Moscow. Nauka, 1979 (in Russian).
5. Krasilov, *Unsolved Problems Of Evolution Theory*, Vladivostok. FERSAS SSSR, 1986 (in Russian).
6. Zherikhin VV, Biocoenotic regulation of evolution, *Paleontol J* **1**:3–12, 1986.
7. Odum E, *Basic Ecology,* Saunders College Publishing, 1983.
8. Zavarzin GA, *Microbial Biosphere*, In book *Biosphere Origin and Evolution,* Dobretsov NL, Kolchanov NA, Rosanov AY, GA Zavarzin (eds.), Springer Science - Business Media, LLC, pp. 25–42, 2008.
9. 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* **70**:1641–1650, 2004.
10. Rogozin DY, Degermendzhy AG, Meromictic lakes as example of ecosystems in which spatial heterogeneity determines biodiversity to a large extend, In book *Biodiversity and dynamics of ecosystems: computational approaches and modeling.* Shumny VK, Shokin YI, Kolchanov NA, Fedotov AM (eds.), Novosibirsk, SB RAS, pp. 436–440, 2006 (in Russian)
11. Zavarzin GA, Zhilina TN, Soda lakes — natural model of ancient continental biosphere, *Priroda* **2**:45–55, 2000 (in Russian).
12. Wilmes P, Bond PL, Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles, *Curr Opin Microbiol* **12**:310–317, 2009.
13. Grube M, Cardinale M, de Castro JV Jr, Müller H, Berg G, Species-specific structural and functional diversity of bacterial communities in lichen symbioses, *ISME J* **3**:1105–115, 2009.
14. Camp JG, Kanther M, Semova I, Rawls JF, Patterns and scales in gastrointestinal microbial ecology, *Gastroenterology*, **136**:1989–2002, 2009.
15. Simões LC, Simões M, Vieira MJ, Biofilm interactions between distinct bacterial genera isolated from drinking water, *Appl Environ Microbiol* **73:**6192–6200, 2007.
16. Kolenbrander PE, Andersen RN, Blehert DS, PG *et al.*, Communication among oral bacteria, *Microbiol Mol Biol Rev* **66**:486–505, 2002.
17. Doolittle WF, Zhaxybayeva O, On the origin of prokaryotic species, *Genome Res* **19**:744–756, 2009.
18. Cohan FM, Koeppel AF, The origins of ecological diversity in prokaryotes, *Curr Biol* **18**:R1024–1034, 2008.
19. Cohan FM, Perry EB, A systematics for discovering the fundamental units of bacterial diversity, *Curr Biol* **17**, R373–R386, 2007.
20. Goldschmidt R, Ecotype, ecospecies and macroevolution, *Experientia Suppl* **35**:140–153, 1980.
21. Jørgensen SE, State-of-the-art of ecological modeling with emphasis on development of structural dynamic models, *Ecol Model* **120**:75–96, 1999.
22. Knibbe C, Fayard JM, Beslon G, The topology of the protein network influences the dynamics of gene order: from systems biology to a systemic understanding of evolution, *Artif Life* **14**:149–156, 2008.

23. Semovski SV, Verheyen E, Sherbakov DY, Simulating the evolution of neutrally evolving sequences in a population under environmental changes, *Ecol Model* **176**:99–107, 2004.
24. Levchenko VF, *Models in the Theory of Biological Evolution*, St Petersburg, Nauka, 1993 (in Russian).
25. Lashin SA, Suslov VV, Matushkin YG, Evolution modeling of trophic web-like communities with compensatory and non-compensatory metabolism, *Inf Vestn VOGiS,* **13**:150–159, 2008 (in Russian)**.**
26. Lashin SA, Suslov VV, Kolchanov NA, Matushkin YG, Simulation of coevolution in community by using the "Evolutionary Constructor" program, *In Silico Biol* **7**:261–275, 2007.
27. Bazykin AD, *Nonlinear Dynamics of Interacting Populations*, Singapore, WSP, 1998.
28. Rubel E, The Replaceability of Ecological Factors and the Law of the Minimum, *Ecology*, **16**:336–341, 1935.
29. Gunbin KV, Suslov VV, Kolchanov NA, Molecular-genetic systems of development: functional dynamics and molecular evolution, *Biochemistry (Mosc)* **73**:219–230, 2008.
30. Zherikhin VV, Evolutionary Biocenology: Problem of Model Choice, In book *Ecosystem Reorganizations and Biosphere Evolution*, **1,** Rozanov AY, Semikhatov MA (eds.), Moscow, Nedra, pp. 13–20, 1994 (in Russian).
31. Thane Papke R, A critique of prokaryotic species concepts, *Methods Mol Biol* **532:**379–395, 2009.
32. Schlegel HG, *General Microbiology*, Cambridge, Cambridge Univ Press, 1993.
33. Sundararaj S, Guo A, Habibi-Nazhad B *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* **32**:D293–D295, 2004.
34. Nelson KE, Clayton RA, Gill SR, *et al.*, Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritime*, *Nature* **399**:323–329, 1999.
35. Nesbo CL, L'Haridon S, Stetter KO, Doolittle WF, Phylogenetic analyses of two "archaeal" genes in thermotoga maritima reveal multiple transfers between archaea and bacteria, *Mol Biol Evol* **18**:362–375, 2001.
36. Il'ina TS, Bacterial superintegrons, a source of new genes with adaptive functions, *Genetika* **42**:1536–1546, 2006 (in Russian).
37. Lashin SA, Suslov VV, Matushkin YG, Modeling cell autonomism origin in prokaryotes using Evolutionary Constructor program, *Proceedings of the Sixth International Conference on Bioinformatics of Genome Regulation and Structure (BGRS 2008)*, 134, Novosibirsk, SB RAS, 2008.
38. Dobretsov NL, Kolchanov NA, Suslov VV, On important stages of geosphere and biosphere evolution, In *Biosphere Origin and Evolution*, Dobretsov NL, Kolchanov NA, Rosanov AY, Zavarzin GA (eds.), Springer Science - Business Media, LLC, pp. 3–23, 2008.

Sergey A. Lashin received his Diploma in Mathematics and System Informatics from Novosibirsk State University (NSU), Novosibirsk, Russia, in 2003, and his Ph.D. in Mathematical Biology and Bioinformatics in 2010. His research is centered on systems biology, bioinformatics, mathematical and computer modeling of evolution, genomics, and gene and metabolic networks.

Valentin V. Suslov received his Diploma in Genetics from Novosibirsk State University (NSU), Novosibirsk, Russia, in 1997. He is currently working on his Ph.D. at the Department of Systems Biology at the Institute of Cytology and Genetics. His research is centered on systems biology, bioinformatics, and problems of evolution.

Yuri G. Matushkin heads the Laboratory of Molecular-genetic Systems at the Department of Systems Biology at the Institute of Cytology and Genetics. He received his Diploma in Mathematics from Leningrad State University (LSU), Saint-Petersburg, Russia, in 1977, and his Ph.D. in Genetics in 1990. His research is centered on systems biology, bioinformatics, mathematical modeling, transcriptomics, and problems of evolution.