Evolution of Genetic Codes through Isologous Diversification of Cellular States

Hiroaki Takagi[1]
Kunihiko Kaneko[1]
Tetsuya Yomo[1,2]
[1] Department of Pure and Applied Sciences
University of Tokyo
Komaba, Meguro-ku
Tokyo 153, Japan
[2] Department of Biotechnology
Faculty of Engineering
Osaka University 2-1 Suita
Osaka 565, Japan

Abstract Evolution of genetic codes is studied as change in the choice of enzymes that are used to synthesize amino acids from the genetic information of nucleic acids. We propose the following theory: the differentiation of physiological states of a cell allows for a choice of enzymes, and this choice is later fixed genetically through evolution. To demonstrate this theory, a dynamical systems model consisting of the concentrations of metabolites, enzymes, amino acyl tRNA synthetase, and tRNA–amino acid complexes in a cell is introduced and studied numerically. It is shown that the biochemical states of cells are differentiated by cell-cell interactions, and each differentiated type starts to use a different synthetase. Through the mutation of genes, this difference in the genetic code is amplified and stabilized. The relevance of this theory to the evolution of non-universal genetic code in mitochondria is suggested. The present theory is based on our recent theory of isologous symbiotic speciation, which is briefly reviewed. According to the theory, phenotypes of organisms are first differentiated into distinct types through the interaction and developmental dynamics, even though they have identical genotypes; later, with mutation in the genotype, the genotype also differentiates into discrete types, while maintaining the "symbiotic" relationship between the types. Relevance of the theory to natural as well as artificial evolution is discussed.

Keywords
isologous diversification, sympatric speciation, evolution of genetic codes, symbiosis

1 Introduction

The protein synthetic system adopted in today's living organisms is a very large and complex network. It consists of over 120 kinds of molecules, such as tRNA, ARS (aminoacyl tRNA synthetase), mRNA, 20 kinds of amino acids, ribosomes, and ATP. In this system, the genetic code plays an important role in linking genetic information in DNA to phenotypic functions, and it has been considered to be stable. In view of its importance and experimental results, the genetic code was once considered to be universal, and the *frozen accident theory* was proposed by Crick [2], in which the genetic code is assumed to have been fixed by accident in the early history of life. In recent studies, however, several variant genetic codes were found, for example, in mitochondrial DNA. Now it is recognized that the genetic code is not universal and can change over the long term. It is therefore important to study the evolution of genetic codes with respect to both their stability and their flexibility, which might seem superficially to conflict.

© 2001 Massachusetts Institute of Technology

To discuss the evolution of genetic codes, it is necessary to point out two of their basic features.

The first concerns the relationship between genetic codes and the molecular structure of proteins. Although tight chemical coupling between codons and amino acids (a key-keyhole relationship) was initially assumed, it is now believed that there is no unique relation between them [25]. Suzuki, Ueda, and Watanabe have recently discovered a *polysemous* codon in certain *Candida* species, where two distinct amino acids are assigned by a single codon [29]. Now looseness in the genetic code is being seriously studied.

Second, the evolutionary change of genetic codes has also been studied in the light of their non-universality. Among these studies, *codon capture theory*, proposed by Osawa and Jukes, is the most popular [28]. The essence of the theory is as follows: If some change in the genetic code occurred without any intermediate stage, a sense codon would be changed to a nonsense one, which would cause lethal damage. Therefore, it is necessary to pass through some intermediate stage in evolution, during which the change of genetic code is not lethal. If the genetic code is degenerate and some specific triplet is seldom used, the tRNA and ARS that correspond to that triplet may be able to change their coding nonlethally.

With these two points in mind, we consider the problem of evolution of genetic codes. First, we expect that the genetic code must have passed through the stage with some ambiguity or looseness in the course of the evolution, since otherwise it is hard to imagine that it could have evolved without causing lethal damage to organisms. How is such looseness supported? How is a different coding for translation supported biochemically? If the difference in genetic codes were solely determined by a genetic system all through the evolutionary process, it would be difficult to imagine how the change from one code to another could occur smoothly, without lethal damage to a cell. Instead, we propose here that the difference in the translation is not solely determined by the nucleus, but is also influenced by the physiological state of a cell, at least at some stage of evolution. Indeed, as will be shown, it is rather plausible that cells with identical genes can have different physiological states. Such differentiation is expected to occur according to the *isologous diversification theory* proposed for cell differentiation [6, 7, 13, 15, 16, 19]. Since the translation system from nucleic acid to amino acid is influenced by several enzymes within a cell, a difference in the physiological state can introduce a change in the translation also. By constructing a model with several biochemicals, we will give an example with non-unique correspondences from nucleic acids to amino acids.

The present paper is organized as follows. In Section 2, we describe isologous speciation theory in some detail, since it gives a basis for the present theory of evolution of genetic codes. In Section 3, we introduce our model of a cell with several chemicals, choosing biochemical reactions that allow for differentiation in physiological states and ARS. In Section 4, we take into account mutation in a genetic system, and study how different genetic codes are established through evolution. Through extensive simulation, we propose the following theory for the evolution of the genetic code: first, phenotypic differentiation occurs for metabolic dynamics through cell-cell interaction. Then each differentiated group of cells starts to use different ARS, and adopt a different way of translating nucleic acid to protein (enzyme). Then, through an evolutionary process with competition for reproduction and mutation of genes, this difference in physiological state results in a difference in genes, and a one-to-one correspondence is established between differentiated phenotype and mutated genes, so that each group can clearly be separated in both phenotype and genotype. After this evolutionary process, the difference in the translation is fixed. Each group finally achieves a different genetic code, which is now fixed in time, and the initial ambiguity or looseness in coding

is reduced. A summary and discussion of the relevance of the present result to cell biology as well as to artificial life are given in Section 5.

2 Isologous Speciation Theory

The background for the present theory for the evolution of genetic codes lies in our isologous symbiotic sympatric speciation theory [17, 18]. Since the theory is essential to the present study, we explain it at length in this section.[1]

2.1 Background of Isologous Symbiotic Sympatric Speciation

The question why organisms are separated into distinct groups, rather than exhibiting a continuous range of characteristics, originally raised by Darwin [3], has not yet been fully answered, in spite of several attempts to explain sympatric speciation. The difficulty in understanding stable sympatric speciation (i.e., the formation of distinct, reproductivly isolated groups) lies in the lack of a known clear mechanism for two groups, which have just started to be separated, to coexist in the presence of mutual interaction and mixing of genes by mating. So far, people have tried to propose some mechanism so that the two groups do not mix and survive independently, as is seen in sexual isolation by mating preference (e.g., [4, 11, 22, 23, 26, 30]). However, this type of theory cannot explain how a mating preference that is conducive to sympatric speciation is selected. Furthermore, if one group disappears by fluctuations due to its finite-size population, the other group does not re-create it. Coexistence with one group is not necessary for the survival of the other. Hence the speciation process is rather weak in the face of the fluctuations that should exist in a population of finite size.

Of course, if the two groups were in a symbiotic state, coexistence would be necessary for the survival of each. However, as long as the phenotype is a single-valued function of the genotype, two groups with little difference in genotypes must have almost the same phenotype. Hence, in the beginning of speciation, it is hard to imagine such a symbiotic mechanism. Accordingly, it is generally believed that sympatric speciation, stable against fluctuations, is rather unlikely.

Recall the standard assumptions for evolution in present-day biology [1, 8]. (a) First, each organism has a genotype and a phenotype. (b) Then, the fitness for survival is given for a phenotype, and Darwinian selection acts for the survival of organisms that have higher fitness. (c) Only the genotype is transferred to the next generation (Weissman's doctrine). (d) Finally, there is direct flow only from genotype to phenotype, that is, a phenotype is determined through a developmental process, given the genotype and the environment (the central dogma of molecular biology). Although there may be some doubt about (c) [and (d)] for some cases, we take this standard viewpoint here.

Note, however, that (d) does not necessarily mean that the phenotype is *uniquely* determined. In the standard population genetics, this uniqueness is assumed, but it is not necessarily postulated within the above standard framework. Indeed, an answer to the speciation problem is provided by dropping this assumption and assuming isologous diversification. Furthermore, there are three reasons to make us doubt the assumption of uniqueness.

First, we have previously proposed an isologous diversification theory, where two groups with distinct phenotypes appear even from the same genotype [6, 7, 13, 15, 16, 19]. In this theory, due to the orbital instability in development, any small difference (or fluctuation) is amplified to a macroscopic level, so that the dynamical state of two

1 This section is fairly independent of the others, and one can skip it or read it by itself.

organisms (cells) can be different even if they have the same set of genes. The organisms are differentiated into discrete types through the interaction, and the existence of each type is necessary to eliminate the dynamic instability in the developmental process, which reappears when the ensemble of one of the types is isolated. Here, the existence of each type is required for the survival of each other, even though every individual has identical, or only slightly different, genotypes.

Second, it is well known experimentally that in some mutants, various phenotypes arise from a single genotype with some probability [10]. This phenomenon is known as low or incomplete penetrance [27].

Last, interaction-induced phenotypic diversification is clearly demonstrated in an experiment on specific mutants of *E. coli*. In fact, the coexistence of (at least) two distinct types of enzyme activity is demonstrated in the well-stirred environment of a chemostat, although they have identical genes [20, 21]. Here, when one type of *E. coli* is removed externally, the remaining type starts to differentiate again and recover the coexistence of the original two types. It is thus demonstrated that distinct phenotypes (e.g., in enzyme activity) appear, owing to the interaction among the organisms, even though they have identical genes.

Here, we take this interaction-induced phenotypic differentiation from a single genotype seriously into account and discuss its relevance to evolution. For this, we have to consider a developmental process that maps a genotype to a phenotype. Consider for example an organism with several biochemical processes that convert external resources into products depending on the internal dynamics.

Here, the phenotype is represented by a set of variables, corresponding to biochemical processes. Genes, since they are nothing but information expressed on DNA, could in principle be included in the set of variables. However, according to the central dogma of molecular biology [requisite (d)], the gene has a special role among such variables. Genes can affect phenotypes (the set of variables), but the phenotypes cannot change the code of genes. During the life cycle, changes in genes are negligible compared with those of the phenotypic variables they control. In terms of dynamical systems, the genes can be represented by control parameters that govern the dynamics of phenotypes, since the parameters in an equation are not changed through the developmental process, and they control the dynamics of phenotypic variables. Accordingly, we represent the genotype by a set of parameters. When an individual organism reproduces, this set of parameters changes slightly by mutation.

Next, there are interactions between individuals through exchange of chemicals. Some chemicals secreted by one organism may be taken up by another, and the two have competitive interactions for resources. This interaction depends on the internal states of the units. In dynamical systems theory, an interaction term is introduced that depends on changes of (biochemical) states of the units.

Then, each individual replicates when some chemicals are accumulated after chemical reactions. Since genotypes are given by a set of parameters representing the reactions, they mutate slightly on reproduction. With each replication, the parameters are changed slightly by adding a small random number.

As the number of organisms grows, not all of them in general survive. The competition for survival is included by random removal of organisms at some rate that has a given dependence on their (biochemical) state.

2.2 Theory of Speciation

We have carried out simulations of several models of the above type, in view of which a speciation theory has been proposed, as follows [17, 18] (see Figure 1 for schematic representation).

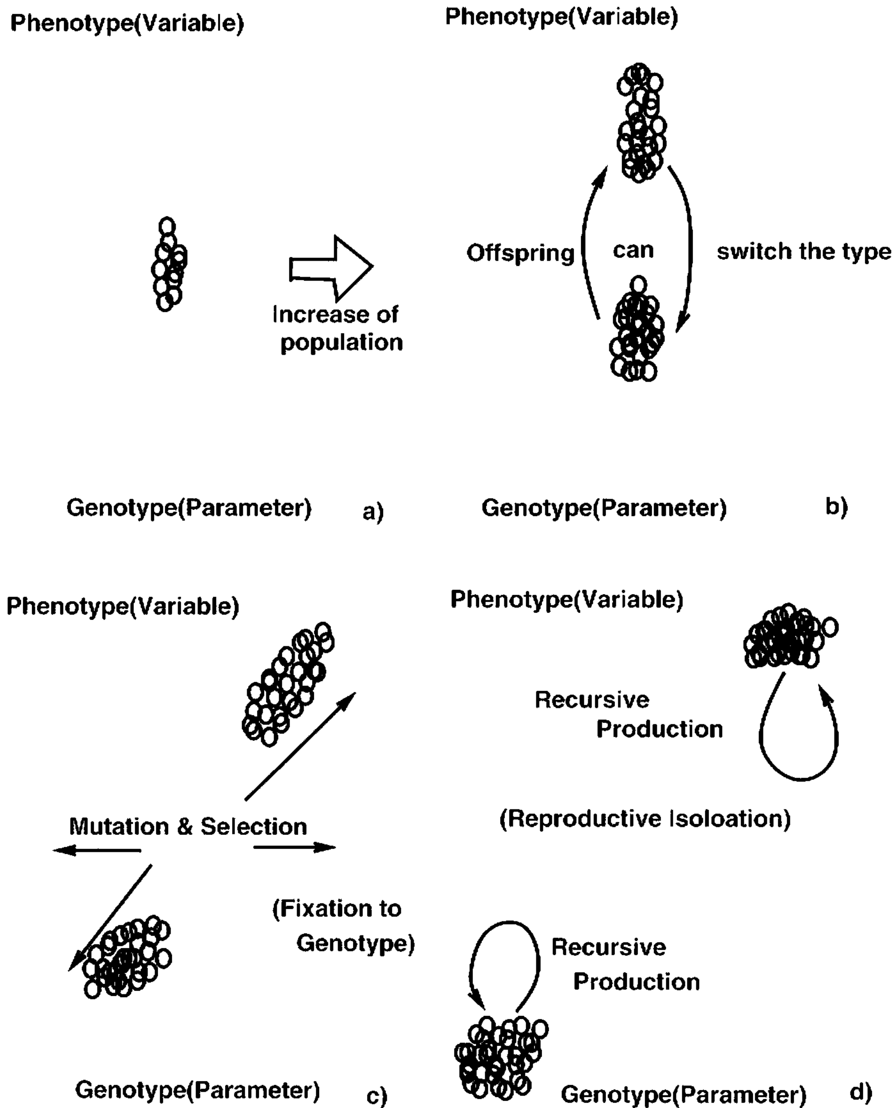

Figure 1. Schematic representation of speciation process, plotted as phenotype-genotype relationship. (a) Initially, there is a group of organisms with distribution centered around a given phenotype and genotype. (b) Then, with the increase of population, the phenotype is differentiated into discrete types. (c) Then according to the difference of phenotypes, genotype is also differentiated. (d) Finally, the two groups differentiate both in genotypes and phenotypes, and form distinct species. Indeed, these two groups are separated also by sexual recombination, since hybrid offspring cannot produce progeny.

2.2.1 Stage 1: Interaction-Induced Phenotypic Differentiation

When many individuals interact by competing for finite resources, the phenotypic dynamics can become differentiated even though the genotypes are identical or differ only slightly. This differentiation generally appears if nonlinearity is involved in the internal dynamics of some phenotypic variables. Slight differences in variables between individuals are amplified by the internal dynamics (e.g., metabolic reaction dynamics). Through interaction between organisms, the differences in phenotypic dynam-

ies are amplified and the phenotype states tend to be grouped into two (or more) types. The dynamical systems mechanism for such differentiation was first discussed with reference to clustering [12], and then extended to the study of cell differentiation [6, 7, 13, 15, 16, 19]. In fact, the orbits lie in a distinct region in the phase space, depending on each of the two groups that the individual i belongs to. Note that the difference at this stage is fixed neither in the genotype nor in the phenotype. The progeny of a reproducing individual may belong to a type distinct from the parent. If a group of one type is removed, then some individuals of the other type change their type to compensate for the missing type. To discuss the present mechanism in biological terms, consider a given group of organisms faced with a new environment and not yet specialized for the processing of certain specific resources. Each organism has metabolic (or other) processes in a biochemical network. As the number of organisms increases, they compete for resources. As this competition becomes stronger, the phenotypes become diversified to allow for different uses in metabolic cycles, and they split into two (or several) groups. Each group is specialized in processing some resources. Here, the two groups realize differentiation of roles and form a symbiotic relationship. Each group is regarded as specialized in a different niche, which is provided by another group.

2.2.2 Stage 2: Coevolution of the Two Groups to Amplify the Difference of Genotypes

At the second stage of our speciation, the difference in both genotypes and phenotypes is amplified. This is realized by positive feedback between the changes in geno- and phenotypes. In general, there is a parameter that has opposite influences on growth speed between the two phenotypes. For example, for the upper group in Figure 1b, assume that the growth speed is higher when the parameter is larger, and the other way around for the lower group. Then, through mutation and selection, the genetic parameters of the two phenotype groups start to separate as shown in Figure 1c.

Indeed, such parameter dependence is not exceptional. As a simple illustration, the use of metabolic processes is different between the two groups. If the upper group uses one metabolic cycle more, then the mutational change of a specific parameter to enhance the use of the cycle is in favor of the upper group, while the change to reduce it may be in favor of the lower group. Indeed, several numerical results support the notion that there always exist such parameters. This dependence of growth on genotypes leads to genetic separation of the two groups.

With this separation, each phenotype (and genotype) tends to be preserved by the offspring, in contrast with the first stage. Now, distinct groups with recursive reproduction have been formed. However, up to this stage, the two groups with different phenotypes cannot exist by themselves in isolation. When isolated, offspring with the phenotype of the other group start to appear. The developmental dynamics in each group, when isolated, are unstable and some individuals start to be differentiated to recover the other group. Each phenotype's dynamics is stabilized by the other through the interaction. Hence, the two groups are in a symbiotic state. To have such stabilization, the population of each group has to be balanced. Even under random fluctuation by finite-size populations and mutation, the population balance of each group is not destroyed. Accordingly, our mechanism of genetic diversification is robust against perturbations.

2.2.3 Stage 3: Genetic Fixation and Isolation of Differentiated Groups

Complete fixation of the diversification in the genes occurs at this stage. Here, even if one group of units is isolated, the offspring of the phenotype of the other group are no longer produced. Offspring of each group keep their phenotype (and genotype) on their own. This is confirmed by numerically eliminating one group of units.

Now, each group has one phenotype corresponding to each genotype, even without interaction with the other group. Hence, each group is a distinct independent reproductive unit at this stage. This stabilization of a phenotypic state is possible because the developmental flexibility at the first stage is lost, due to the shift of genotype parameters. The initial phenotypic change introduced by the interaction is now fixed in the genes.

To check the third stage of our theory, it is straightforward to study the further evolutionary process of only one isolated group. In order to do this, we pick out some population of units of only one type, after the genetic fixation is completed and both the genotypes and the phenotypes are separated into two groups, and start the simulation again. When the groups are picked at this third stage, the offspring keep the same phenotype and genotype. Now, only one of the two groups exists. The other group is no longer necessary to maintain stability.

2.3 Some Remarks

To check the conditions for speciation, we have performed numerical experiments on evolution, by choosing model parameters so that differentiation into two distinct phenotypic groups does not occur initially. In this case, separation into two (or more) groups with distinct pheno- and genotypes is never observed, even if the initial variance of genotypes is large or a large mutation rate is adopted.

Furthermore, the genetic differentiation always occurs when the phenotype differentiates into two (or more) distinct groups, as long as mutation exists. Hence, phenotypic differentiation is a necessary and sufficient condition for speciation in a standard biological situation, that is, a process with reproduction, mutation, and a proper genotype-phenotype relationship. Note that the interaction-induced phenotypic differentiation is deterministic in nature. Once the initial parameters of the model are chosen, it is already determined whether such differentiation will occur or not.

The speciation process is also stable against sexual recombination. In sexual recombinations, two genes are mixed, and the two differentiated groups may be mixed and the speciation destroyed. We have found that our speciation process is stable under sexual recombinations [17, 18]. Hybrids are indeed formed with random mating, but they have lower reproduction rates, and finally they become sterile. Thus the definition of species, with respect to sterility of the hybrid, is satisfied.

In our speciation process, the potentiality for a single genotype to produce several phenotypes decreases. After the phenotypic diversification of a single genotype, each genotype again appears through mutation and assumes one of the diversified phenotypes in the population. Thus the one-to-many correspondence between the original genotype and phenotypes eventually ceases to exist. As a result, one may expect a phenotype to be uniquely determined for a single genotype in wild types, since most organisms at the present time have gone through several speciation processes.

Finally, it should again be stressed that *neither any Lamarckian mechanism nor epigenetic inheritance is assumed in our theory*, in spite of the genetic fixation of the phenotypic differentiation. Only the standard flow from genotype to phenotype is included in our theory. Note also that genetic "takeover" of phenotype change was also proposed by Waddington as genetic assimilation [32], in possible relationship with Baldwin's effect. Using the idea of *epigenetic landscape*, he showed that genetic fixation of the displacement of phenotypic character is in the genes. In our case the phenotype differentiation is not given by an epigenetic landscape, but rather, the developmental process forms different characters through the interaction. Distinct characters are stabilized through the interaction. With this interaction dependence, the two groups are necessary to each other, and robust speciation is possible.

3 Our Model for the Evolution of Genetic Codes

Now, let us come back to the problem of the evolution of genetic codes. Here, we construct an abstract model to demonstrate the theory for the evolution of a genetic code. Of course, it is almost impossible to describe all factors of a complex cellular process. Furthermore, even if we succeeded in doing so, we could not understand how the resulting model worked, since it would be much too complicated. Rather, we extract only some basic features of the problem, and construct a model to clarify certain general aspects of the evolution of the genetic code. In particular, we show how differentiated phenotypes are organized, which adopt a different coding in the translation from nucleic acids to amino acids, based on isologous diversification. Then, with the evolution through mutation of genes, the different translations will be shown to be established, following the theory of the last section.

We start from a cell with a set of variety of biochemicals. Considering the metabolic and genetic process, at least four kinds of basic compounds are necessary, namely, metabolic chemicals, enzymes for metabolic reactions, chemicals for genetic information, and enzymes to translate genetic information to protein. In the present paper, these four kinds of chemicals are chosen as the metabolites (metabolic chemicals), enzymes for metabolites, tRNA–amino acid complexes, and ARS, respectively. Now, as state variables characterizing the cell, we introduce

- $c_i^{(j)}(t)$: concentration of jth metabolic chemical in ith cell
- $d_i^{(j)}(t)$: concentration of jth enzyme for metabolites in ith cell
- $e_i^{(j)}(t)$: concentration of jth ARS in ith cell
- $x_i^{(j)}(t)$: concentration of jth tRNA–amino acid complex in ith cell

For the dynamics of these chemicals, we consider the following processes.

- Intra-cellular chemical reaction network
- Inter-cellular interaction
- Cell division and mutation

Now, we describe each process. See Figure 2 and Figure 3 for a schematic representation of our model.

3.1 Intra-cellular Chemical Reaction Network

In general, each biochemical reaction in cells is catalyzed by some enzymes. Here, each metabolic reaction is assumed to be catalyzed by one specific enzyme, and a simple form of reaction rate is adopted given by just the product of the concentrations of the substrate and enzyme. (This specific form is not essential, and the same qualitative results are obtained by using some other form, such as the Michaelis-Menten.) Here we choose a network consisting of reactions from some metabolite j to another metabolite k catalyzed by the enzyme k. The network is chosen randomly, and is represented by a reaction matrix $W(j, k)$, whose elements take the value 1 if there is a reaction path from j to k, and 0 otherwise. The network is fixed throughout the simulation. Of course, the dynamics can depend on the choice of the reaction network. Here we choose a network that allows for some oscillatory dynamics. Oscillation is rather commonly observed, as long as there are a sufficient number of autocatalytic paths.

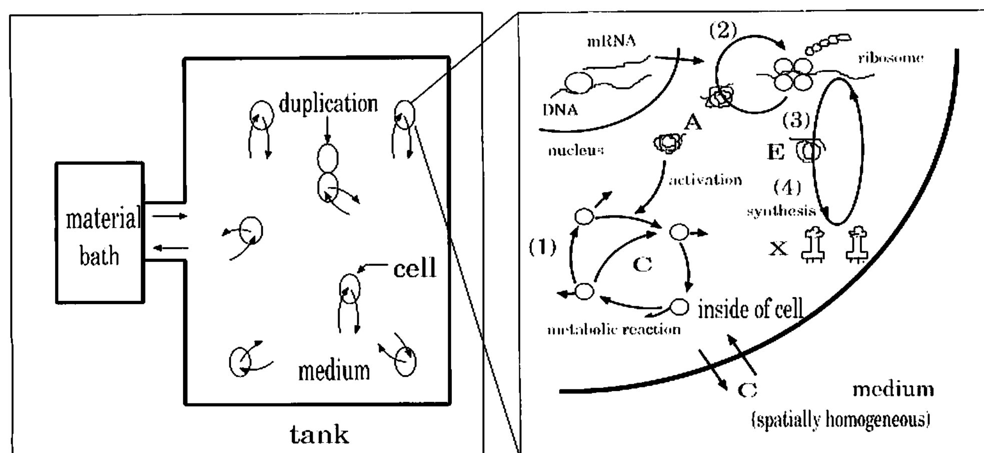

Figure 2. Schematic representation of our model.

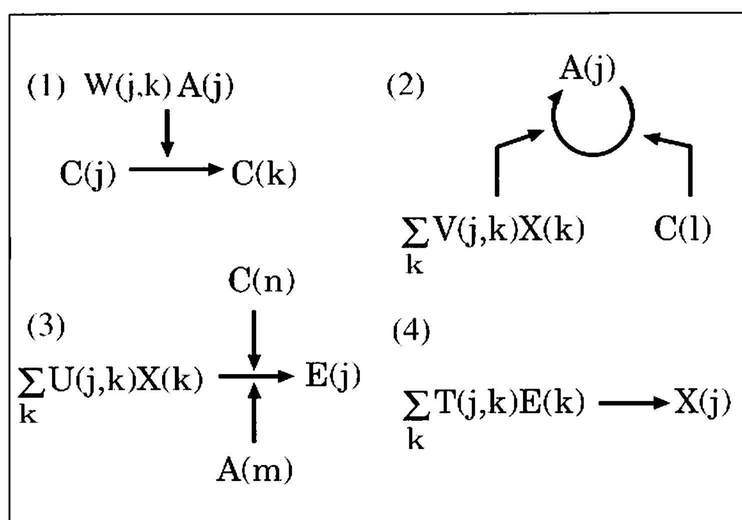

Figure 3. Schematic representation of intra-cellular reaction network.

Next, each enzyme—including ARS, the one for the synthesis of tRNA—is synthesized from amino acids. This synthesis is again catalyzed by some enzyme. It is specified by a resource table. For the enzyme $a^{(j)}$ and the ARS $e^{(j)}$, the tables are given by $V(j, k)$ and $U(j, k)$ respectively. We also set all entries of $V(j, k)$, $U(j, k)$ at random, with the normalization constraint $\sum_k V(j, k) = \sum_k U(j, k) = 1.0$.

Third, we assume that an ARS produces tRNA–amino acid complexes in proportion to its amount. The correspondence between the two is given by a reaction matrix $T(j, k)$, which is 1 if ARS $e^{(j)}$ produces tRNA–amino acid complex $x^{(k)}$, and 0 otherwise. To include ambiguity, we allow one-to-many correspondence between e and x. The matrix $T(j, k)$ is again chosen randomly.

Here we take P (= 16) species of metabolic chemicals (C) and the corresponding enzymes (A), R (= 12) species of ARS (E), and Q (= 6) tRNA–amino acid complexes (X). Accordingly the concentration changes of chemicals due to intracellular processes

are given by

$$\frac{dc_i^{(j)}(t)}{dt} = D_1 \sum_{k=1}^{P} W_i^{(k,j)} a_i^{(k)}(t) c_i^{(k)}(t) - D_1 \sum_{k=1}^{P} W_i^{(j,k)} a_i^{(j)}(t) c_i^{(j)}(t),$$

$$\frac{da_i^{(j)}(t)}{dt} = D_3 \left(\sum_{k=1}^{Q} V_i^{(j,k)} x_i^{(k)}(t) \right) a_i^{(j)}(t) c_i^{(l(j))}(t)$$

(where $l(j)$: $j \to l$ gives a one-to-one mapping), and

$$\frac{de_i^{(j)}(t)}{dt} = D_4 \left(\sum_{k=1}^{Q} U_i^{(j,k)} x_i^{(k)}(t) \right) a_i^{(m(j))}(t) c_i^{(n(j))}(t)$$

(where $m(j)$, $n(j)$: $j \to l$ give one-to-one mappings).

Finally, tRNA–amino acid complexes, which provide the materials of all enzymes, are assumed to change on a faster time scale than the above three types of chemicals. Hence, we adiabatically eliminate its concentration to obtain the equation for it. By setting

$$\frac{dx_i^{(j)}(t)}{dt} = D_5 \sum_{k=1}^{R} T_i^{(k,j)} e_i^{(k)}(t) - D_3 x_i^{(j)}(t) \sum_{k=1}^{P} V_i^{(k,j)} a_i^{(k)}(t) c_i^{(l(k))}(t) - D_4 x_i^{(j)}(t) \sum_{k=1}^{R} U_i^{(k,j)} a_i^{(m(k))}(t) c_i^{(n(k))}(t) = 0,$$

we obtain

$$x_i^{(j)}(t) = D_5 \sum_{k=1}^{R} T_i^{(k,j)} e_i^{(k)}(t) \times \left(D_3 \sum_{k=1}^{P} V_i^{(k,j)} a_i^{(k)}(t) c_i^{(l(k))}(t) + D_4 \sum_{k=1}^{R} U_i^{(k,j)} a_i^{(m(k))}(t) c_i^{(n(k))}(t) \right)^{-1}.$$

Note that the translation of genetic information to proteins is carried out by the reaction between X (tRNA–amino acid complexes) and E (ARS). One can discuss the difference in coding by examining which species of E has nonzero concentration and acts in the translation process. We first study the relevant ways in which the difference in physiological states given by C affects the choice of E.

3.2 Cell-Cell Interaction

According to the isologous diversification theory, cell-cell interaction is essential to establish distinct cell states. Here we consider the interaction as diffusion of some chemicals through the medium. In this model, we assume that only metabolic chemicals (c) are transported through the membrane, which is rather plausible biologically.

Assuming that the cells are in a completely stirred medium, we neglect spatial variation of chemical concentrations in the medium. Hence we need only one new set of concentration variables:

- $C^{(j)}(t)$: concentration of jth metabolic chemical in the medium

Therefore, all the cells interact with each other through the same environment. As a transport process we choose the simplest diffusion process: a flow proportional to the concentration difference between the inside and outside of a cell.

Of course, the diffusion coefficient depends on the metabolic chemical. Here, for simplicity we assume that all the chemicals c are classified as either penetrating or nonpenetrating ones. The former have the same diffusion coefficient D_2, while for the latter the coefficient is set to 0. Here we define the index σ_m, which is 1 if the chemical $c^{(m)}$ can penetrate the membrane, and otherwise 0. Each cell grows by taking in penetrating chemicals from the medium and transforming them to other, nonpenetrating chemicals.

Accordingly, the diffusion term

$$\sigma_j D_2 [C^{(j)}(t) - c_i^{(j)}(t)] \tag{1}$$

is added to the equation for $dc_i^{(j)}(t)/dt$, while the concentration change in the medium is given by

$$\frac{dC^{(j)}(t)}{dt} = D_0[\overline{C^{(j)}} - C^{(j)}(t)] - \sigma_j D_2 \frac{\sum_{k=1}^{N} C^{(j)}(t) - c_k^{(j)}(t)}{\text{Vol}},$$

where the parameter Vol is the volume ratio of the medium to a cell, and N is the number of cells. Since these chemicals in the medium are consumed by a cell, we impose flows of penetrating chemicals from the outside of the medium that are proportional to the concentration differences. These flows are given by the term $D_0[\overline{C^{(j)}} - C^{(j)}(t)]$, where the external concentration of chemical $C^{(j)}$ is denoted by $\overline{C^{(j)}}$.

The variables $c_i^{(j)}$, $a_i^{(j)}$, $e_i^{(j)}$, and $x_i^{(j)}$ stand for the concentrations. Since the volume of a cell can change with a flow of metabolites, its change should be taken into account. Here, we compute the increase of the volume from the flow of chemicals as the sum of the terms in equation (1). The concentration is diluted in accordance with this increase of the volume. With this process the sum

$$\sum_{j=1}^{P} c_i^{(j)}(t) + \sum_{j=1}^{P} a_i^{(j)}(t) + \sum_{j=1}^{R} e_i^{(j)}(t)$$

is preserved through the development of a cell, and it is here set equal to 1.

3.3 Cell Division

Each cell gets resources from the medium and grows by changing them to other chemicals. With the flow into a cell, the chemicals accumulate. As mentioned, this leads to an increase of the volume of the cell. We assume that the cell divides when the volume is twice the original. In the division process, a cell is divided into two almost equal cells, with some fluctuations. Hence, the concentrations of chemicals, $b_i^{(j)}$ (where b represents either c, a, or e), are changed to $(1 + \eta)b_i^{(j)}$ and $(1 - \eta)b_i^{(j)}$, with η as a random number from $[-10^{-2}, 10^{-2}]$. As will be shown later, this fluctuation can be

amplified to a macroscopic level. Its magnitude is not essential, but the existence of the fluctuation itself is necessary for differentiation.

3.4 Mutation

To discuss the evolution of genetic codes in a long run, we need to include mutation of genes. In our model, the process in which the genetic information is translated from DNA into amino acids is represented as enzyme synthesis reactions. Here both the matrices U and V, which are resource tables for enzyme synthesis, correspond to the genotype, while the other chemical concentrations give the biochemical states of the cell. At each division, each element of the matrix U or V is mutated by the random number κ from the range $[-\varepsilon, \varepsilon]$, where ε is the amplitude of the mutation rate, which we later set at 10^{-3} for most simulations.

Since in our model there is no direct process for the concentrations to change the matrix, the "central dogma" of molecular biology is satisfied, that is, the genotype can change the phenotype, but not vice versa. We also assume that mutation of genes affects only the catalytic abilities of the enzymes a, c, and not the specificity of the catalytic reactions.

Recall that the difference in the genetic code is represented by which kinds of ARS are used in a cell, depending on the physiological state of the cell. Here, we are interested in how this difference is fixed genetically through evolution. With the change of the matrix element of U corresponding to the ARS, the use of specific kind of ARS may start to be fixed, with the increase of some matrix element (to approach unity), according to the theory of Section 2. If this is the case, specific mappings between ARS and the tRNA–amino acid complex are selected, to establish a different coding system. We will confirm this argument in the following sections, based on the simulation results of our model.

4 Isologous Diversification of the Genetic Codes

First, we discuss the behavior of the cell system, without introducing mutation. We assume that intracellular chemical dynamics for a single-cell system show oscillation. Since there are many oscillatory reactions in real cells (such as Ca^{2+}, cAMP, NADH) and oscillation is easily brought about by autocatalytic reactions (and also by the hypercycle [5]), the existence of oscillation is a natural assumption [9, 31].

We have carried out several simulations by constructing a variety of reaction networks that produce oscillatory dynamics. In many such examples, we have observed the differentiation process as discussed. Here we focus on such cases, mostly using one typical example network. The oscillation of chemical concentrations at the first stage in this adopted example is shown as type 0 in Figure 4. This oscillation of chemical concentrations is observed for most initial conditions, although for a few there is also a fixed point solution, whose basin volume is very small. Note that the oscillation of chemicals, and accordingly the expressions of genes, show on/off type switching, as is true in realistic cell systems.

Now we discuss the behavior of cells as their numbers increase. As cells reproduce, their chemical states start to be differentiated, in accord with isologous diversification. First, the phase coherence of oscillations is lost in the intra-cellular dynamics as the cells increase in number. Then, the chemical state of cells differentiates into two groups. Each group has a different composition in metabolites and also in other enzymes. In the example shown in Figure 5, the type 1 cell is differentiated from the type 0 cell. Here the type 0 cell has a higher activity, with larger metabolite concentrations, than the type 1. In order for a cell to grow, metabolites, enzymes, and ribonucleic acids are necessary. The growth speed of a cell depends on the balance among the con-

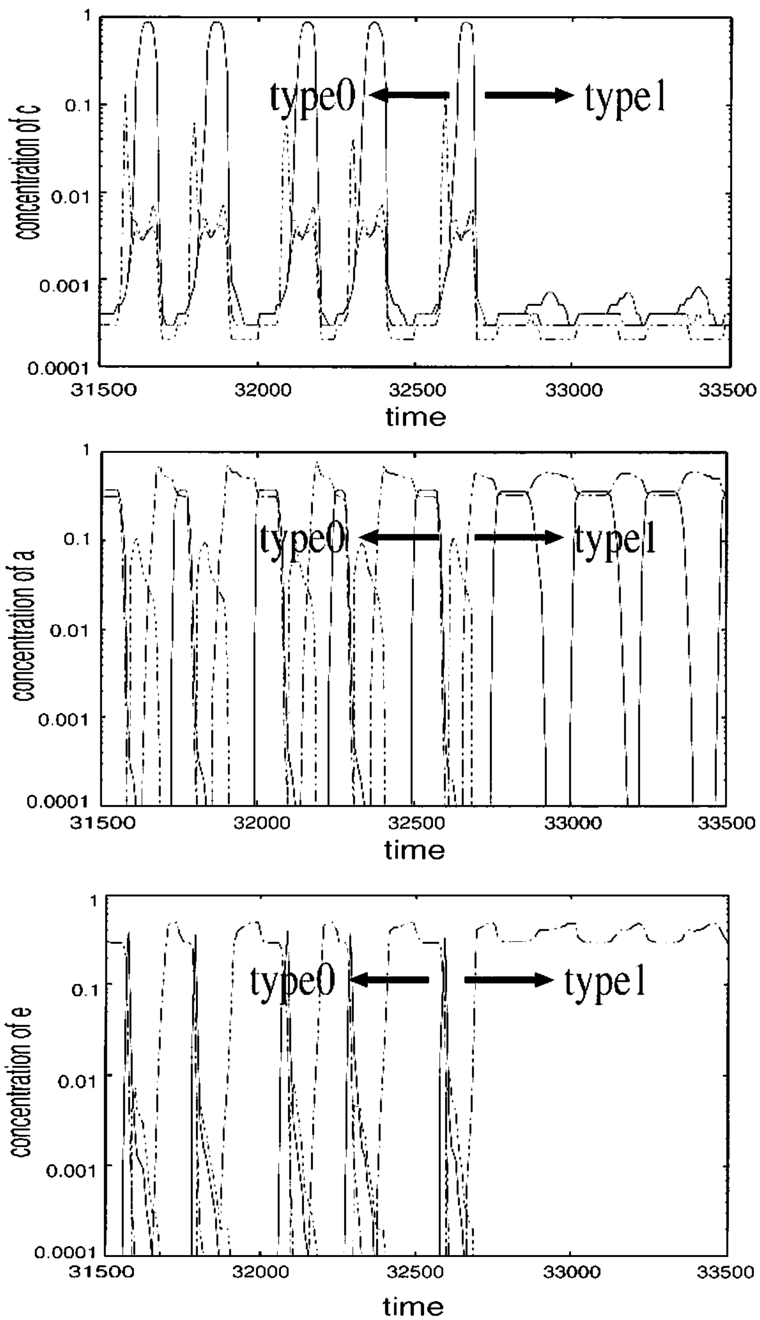

Figure 4. Oscillations of chemicals. The time series of some $c_i^{(j)}$, $a_i^{(j)}$, $e_i^{(j)}$ are plotted on a semi-log scale. The parameters are set at $D_1 = 3.0$, $D_2 = 0.050$, $D_3 = 100.0$, $D_4 = 100.0$, $D_5 = 1.0$, $D_6 = 0.050$, $Vol = 100.0$, $\overline{C^{(j)}} = 0.010$ (for all j) in all the simulations shown in the present paper.

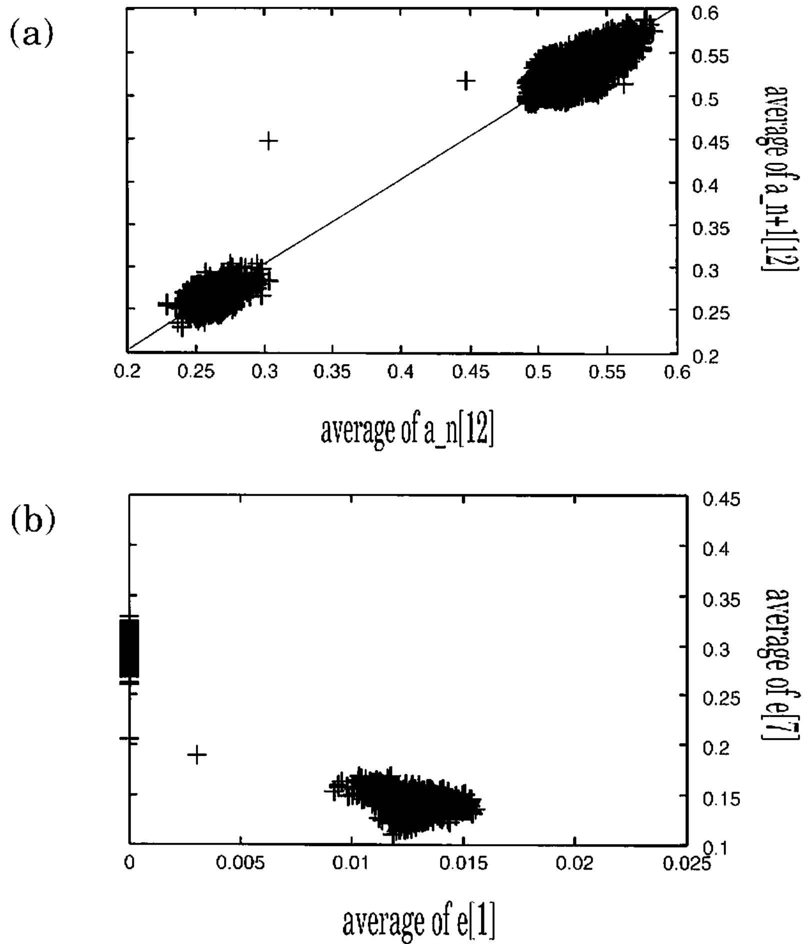

Figure 7. (a) is the return map of the average concentration of $a^{(12)}$, plotted in the same way as Figure 5, while (b) shows the average of ($e^{(1)}$, $e^{(9)}$), plotted at every division event. Each cell displays recursive production.

in the genotype. Hence, we adopt a network such that the chemical states of a cell are differentiated to allow for different uses of ARS to synthesize tRNA–amino acid complexes. Accordingly, we choose the matrix W adopted in the previous section (for Figure 2).

Of course, the evolutionary process depends on the mutation rate, which is given by the magnitude of the noise ε added to the matrix elements. If ε is larger than 10^{-2}, differentiation produced initially is destroyed, and the types are not preserved by cell divisions. With such high mutation rates, the distribution of matrix elements among cells is broader, and both the genotypes and phenotypes are distributed without forming any distinct types. Then the initial loose coupling between genotype and phenotype remains. No trend in the evolution of codes is observed.

When the mutation rate is lower, the genotypes (i.e., the matrix elements) also start to differentiate. Each group, with different compositions of metabolites, starts to take on different matrix element values. An example of the time course of some matrix elements is shown in Figure 8a. Two separate groups are formed according to the differentiated chemical states of metabolites given in Figure 7. Through mutation and

selection, the genotype is also differentiated, following the phenotypic differentiation. This differentiation, originally brought about by the interaction among cells, is embedded gradually in genotypic functions.

Not all the elements of U and V, but only some of them, split. In fact, metabolites or enzymes having higher concentrations are often responsible for the differentiation. To estimate the splitting speed in the genotype space, we have plotted the average differences in the values of elements of U and V between the two types. To be specific, we have measured the following difference between the averaged values of a given matrix element of each type, i.e.,

$$d^{(j,k)} = \left| \frac{1}{N_0} \sum_{i \in \text{type } 0}^{N_0} S_i^{(j,k)} - \frac{1}{N_1} \sum_{i \in \text{type } 1}^{N_1} S_i^{(j,k)} \right|,$$

where S represents either V or U, and N_0 and N_1 are the numbers of type 0 and 1 cells respectively.As shown in Figure 8b, the separation progresses linearly in time, although the mutational process is random. In this sense, this separation process is rather fast and deterministic in nature, once the phenotype is differentiated, as is expected from the theory of Section 2. Furthermore, the slope in the figure differs from chemical to chemical, although the same mutation rate is adopted for all elements. For some of the other matrix elements, no separation occurs.

With this mutation process, the difference in chemical states is also amplified, as shown in Figure 8. As this evolution proceeds, the differentiation starts to be more rigid. In Figure 7, we have plotted the return map of the chemical states. Now, the frequency of the differentiation event from type 0 to type 1 decreases in time. Each type keeps recursive production.

Next, we examine this separation process with a transplant experiment, to see if each group of cells can survive on its own. At the initial stage of evolution, when type 0 cells are extracted, some of them spontaneously differentiate to type 1 cells. Type 0 cells cannot exist by themselves. When evolution changes the genes, the rate of differentiation to type 1 cells from isolated type 0 cells is reduced. Later in the evolution (≈ 600 generations), transplanted type 0 cells no longer differentiate to type 1, and the type 0 cell exists stably on its own.

On the other hand, as already mentioned, type 1 cells, when transplanted, are transformed to type 2 cells, where different ARSs are used in the translation process (see Figure 9). This characteristic feature does not change through the evolution.

With this type of fixation process, the difference in the correspondence between nucleic acid and protein (enzyme) is fixed. For the matrix U, one of the elements $U^{(i,j)}$ for any given j increases through evolution. As shown in Figure 10, $U^{(7,0)}$ increases for a type 1 cell, and $U^{(7,5)}$ decreases, implying that the correspondence between $x^{(0)}$ and the ARS $e^{(5)}$ is stronger. In other words, the loose correspondence between the nucleic acid and amino acid is reduced in time, and a tight relationship between them is established.

Due to the evolution of matrix elements, the difference in the correspondence between the type 0 and type 1 cells gets amplified. Hence, the difference in the correspondence, initially brought about as a difference in metabolic states, is now fixed in genes, and each type of cell, even after isolation, maintains a different use of ARS for the translation of genetic information.

6 Summary and Discussion

In the present paper, we have studied how different correspondences between nucleic acids and enzymes are formed and maintained through evolution. To discuss this

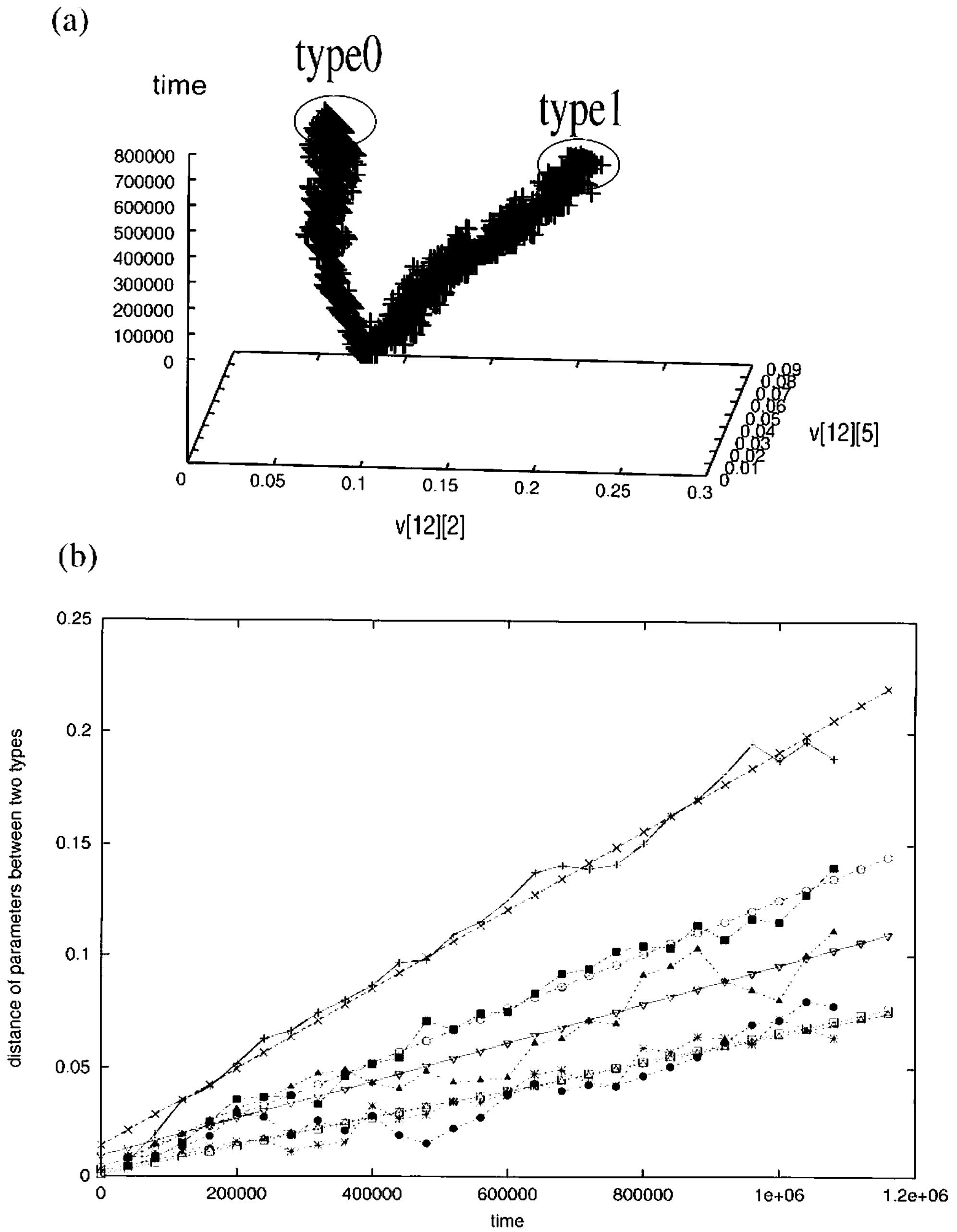

Figure 8. (a) The temporal changes of $V^{(12,2)}$, and $V^{(12,5)}$, namely, the activity of $a^{(12)}$ for the compositions $x^{(2)}$ and $x^{(5)}$. The parameter values are plotted at each division event. (b) The temporal change of the distance between the averages for some matrix elements of each type.

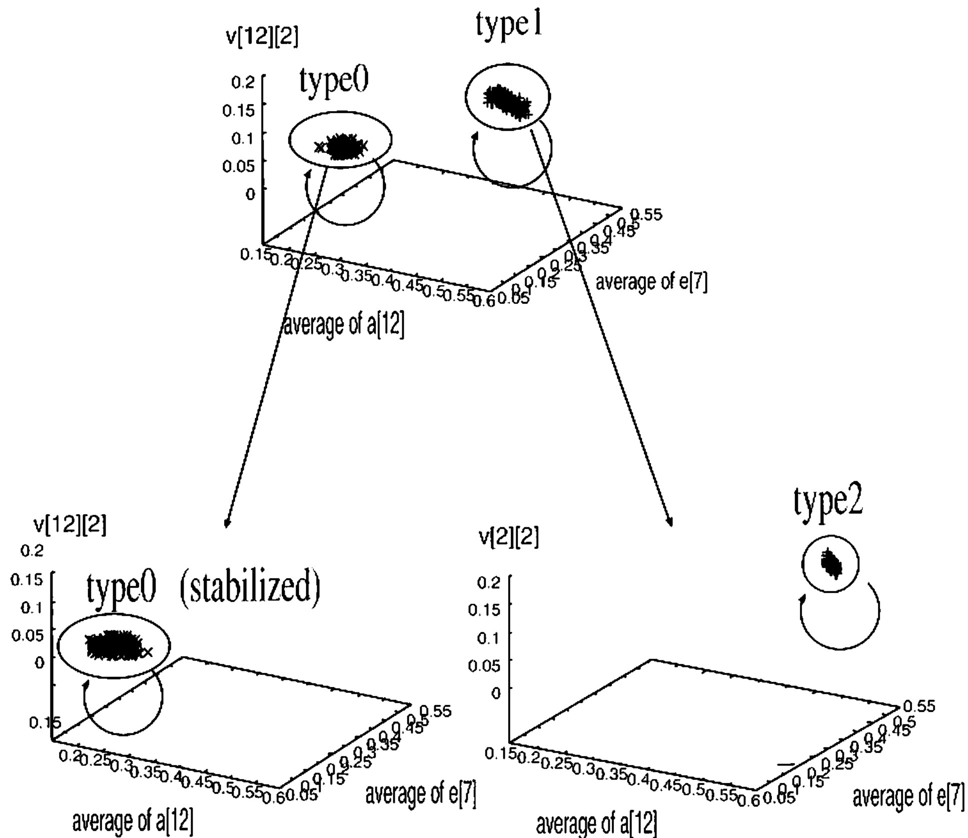

Figure 9. Genotype-phenotype relation after the transplant experiment. The set of values $(a^{(12)}, e^{(7)}, V^{(12,2)})$ is plotted at every division event.

problem we have adopted a model of a cell consisting of

(a) An intracellular metabolic network

(b) Ambiguity in the translation system

(c) Cell-cell interaction through the medium

(d) Cell division

(e) Mutation of the correspondence between nucleic acid and enzyme

According to our theory, the evolution of the genetic code is summarized as follows.

1. First, due to the intra-cellular biochemical dynamics with metabolites, enzymes, tRNA, and ARS, distinct types of cells with distinct physiological states are formed (say, type 0 and type 1 cells). Each cell type has different chemical composition and also uses different species of ARS for the protein synthesis. Hence, each cell type adopts different correspondence between nucleic acids and enzymes. The differentiation at this stage is due to cell-cell interaction. For example, a type 1 cell is differentiated from a type 0 cell, and can maintain itself only in the presence of

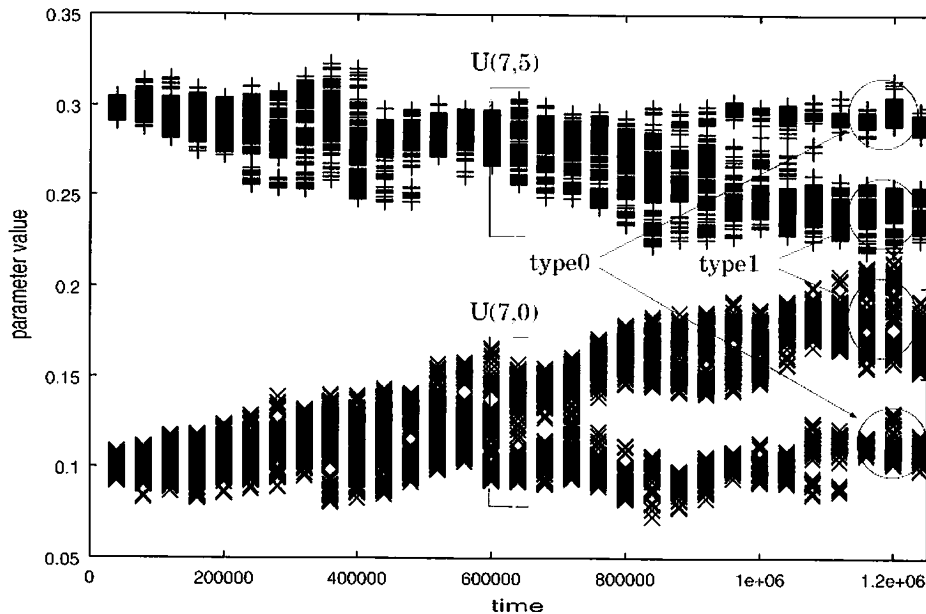

Figure 10. The change of the matrix U through evolution. The parameter values $U^{(7,0)}$ and $U^{(7,5)}$ are plotted at each division event. First, each type starts to take different $U^{(7,0)}$ values, and later, different $U^{(7,5)}$ values. For example, the type 1 cell starts to use more $x^{(0)}$ and $e^{(7)}$.

type 0 cells. The difference in the correspondence, however, is not fixed as yet, and at each cell division, each cell can take a different metabolic state, and the correspondence is changeable.

2. Next, by mutation of the catalytic ability of enzymes at each division of a cell, each distinct cell type starts to be fixed, and keeps its type after the division. The difference in chemical states is now fixed in parameters that characterize the catalytic ability. Accordingly, each cell type with distinct metabolic states is fixed also in the catalytic ability of enzymes represented by genes.

3. After the fixation of distinct types is completed in both phenotypes and genotypes, these types are maintained even if each type of cells is isolated. Each type uses different ARSs for the translation between nucleic and amino acids, and this difference in usage is amplified through evolution. At this stage, one can say that different coding, originally introduced as distinct physiological states of cells through cell-cell interaction, is established genetically.

The result presented here is rather general, as long as cellular states differentiate into a few types, as is generally observed in a model adopting features (a)–(e). Although the network we have adopted is randomly chosen, it is expected that the same evolution of the genetic code will be observed as long as this general setup with (a)–(e) is satisfied. As mentioned in Section 2, the evolutionary process here is based on the standard Darwinian process without any Lamarckian mechanism, although the genetic fixation occurs later from the phenotypic differentiation.

6.1 The Origin of the Mitochondrial Non-universal Genetic Code

Our theory of the evolution of genetic codes can shed new light on the non-universal genetic code of mitochondria. In recent studies in molecular biology, it is suggested that mitochondria had used almost the same code as the universal one before endosymbiosis [24], and its code deviated afterward [28].

According to our theory for the evolution of genetic codes, the coding system can depend on cell-cell interaction. A type 2 cell that is formed by the isolation of a type 1 cell makes different use of ARSs than a cell in coexistence with type 1 cells. With the interaction, the cells take on a different coding system. Furthermore, this difference in the coding is established through evolution. In this sense, it is natural that mitochondria, which start to live within a cell and have strong interactions with the host cell, will establish a different coding system through evolution.

Although it might seem that the switch to a different coding would be fatal to an organism, a cell can survive it via the loose coupling between the genotype and phenotype. The loose coupling produced by the cell-cell interaction is essential to the evolution of non-universal genetic codes.

It should also be stressed that the genetic code is not necessarily solely determined by the genetic system. In a biochemically plausible model, we have demonstrated that a change in the physiological state of a cell can lead to a difference in the genetic code. Based on our theory, we believe that this dependence on the physiological state is essential to the study of non-universal coding in mitochondria and other systems. Furthermore, the possibility of a difference in coding may not be limited to the early stage of evolution. It may even be possible to pursue this possibility experimentally, by changing the interaction among cells or intracellular organs that contain genetic information.

6.2 Relevance to Artificial Life

Discussion of the mechanisms involved in evolution often remains vague, since no one knows for sure what has occurred in history, in view of the limited fossil data. Our theory, on the other hand, is conducive to experimental verification. As mentioned, isologous diversification has already been observed in the differentiation of enzyme activity of *E. coli* with identical genes [20, 21]. We have already started an experiment on the evolution of *E. coli* in the laboratory [33], controlling the strength of the interaction through the population density. With this experiment we can check if the evolution on the genetic level is accelerated through interaction-induced phenotypic diversification, and can see if the evolutionary mechanism of Section 2 really occurs in nature. In this way, our theory is testable in the laboratory, in contrast with many other speculations. Change of genetic code through evolution can also be checked in the laboratory.

In the same way, our study is relevant to the field of artificial life (AL), since AL attempts to understand some biological processes, such as evolution, by constructing an artificial system in a laboratory or in a computer.

A problem in most of the present AL studies lies in their being too much symbol-based. They generally assume some rule, which is represented as manipulation over symbols. A model with such rules can in principle be implemented by a universal Turing machine. Hence it generally faces the problem that the emergence may not be possible in principle in such a system, since emergence means generation of a novel, higher level that is not originally written in a rule. The same drawback appears in any symbol-based study of evolution (i.e., any study starting from the evolution of symbols corresponding to genes), and indeed, an AL study of evolution is often nothing but a kind of complicated optimization problem.

According to our theory, first the phenotype is differentiated, given by a continuous (analog) dynamical system, which is later fixed in genes that serve as rules for dynamical

systems. Now, rules written by symbols (genetic codes) are not necessarily the principal cause of evolution [14].

Acknowledgments

We would like to thank C. Furusawa for stimulating discussions. This work is supported by grants-in-aid for scientific research from the Ministry of Education, Science and Culture of Japan (11CE2006, Komaba Complex Systems Life Project, and 11837004).

References

1. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J. D. (1994). *The molecular biology of the cell*, 3rd ed. New York: Garland.
2. Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology, 38*, 367.
3. Darwin, C. (1859). *On the origin of species by means of natural selection or the preservation of favored races in the struggle for life*. London: Murray.
4. Dieckmann, U., & Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature, 400*, 354.
5. Eigen, M., and Shuster, P. (1979). *The hypercycle: A principle of natural self-organization*. Berlin: Springer-Verlag.
6. Furusawa, C., & Kaneko, K. (1998). Emergence of rules in cell society: Differentiation, hierarchy, and stability. *Bulletin of Mathematical Biology, 60*, 659.
7. Furusawa, C., and Kaneko, K. (1998). Emergence of multicellular organism: Dynamic differentiation and spatial pattern. *Artificial Life, IV*, 79.
8. Futsuyma, D. J. (1986). *Evolutionary biology*, 2nd ed. Sunderland, MA: Sinauer Associates Inc.
9. Hess, B., & Boiteux, A. (1971). Oscillatory phenomena in biochemistry. *Annual Review of Biochemistry, 40*, 237.
10. Holmes, L. B. (1979). Penetrance and expressivity of limb malformations. *Birth Defects Original Article Series, 15*, 321.
11. Howard, D. J., & Berlocher, S. H. (Eds.). (1998). *Endless form: Species and speciation*. New York: Oxford University Press.
12. Kaneko, K. (1990). Clustering, coding, switching, hierarchical ordering, and control in network of chaotic elements. *Physica D, 41*, 137.
13. Kaneko, K. (1994). Relevance of clustering to biological networks. *Physica D, 75*, 55.
14. Kaneko, K. (1998). Life as complex systems: Viewpoint from intra-inter dynamics. *Complexity, 3*, 53.
15. Kaneko, K., & Yomo, T. (1997). Isologous diversification: A theory of cell differentiation. *Bulletin of Mathematical Biology, 59*, 139.
16. Kaneko, K., & Yomo, T. (1999). Isologous diversification for robust development of cell society. *Journal of Theoretical Biology, 199*, 243.
17. Kaneko, K., & Yomo, T. (2000). Symbiotic speciation from a single genotype. *Proceedings of the Royal Society, Section B, 267*, 2367.
18. Kaneko, K., & Yomo, T. (2000). Sympatric speciation from interaction-induced phenotype differentiation. In *Proceedings of Artificial Life VII*. Cambridge, MA: MIT Press.
19. Kaneko, K., & Furusawa, C. (2000). Robust and irreversible development in cell society as a general consequence of intra-inter dynamics. *Physica A, 280*, 23.
20. Kashiwagi, A., Kanaya, T., Yomo, T., & Urabe, I. (1999). How small can the difference among competitors be for coexistence to occur. *Researches on Population Ecology, 40*(2), 223.

21. Ko, E., Yomo, T., & Urabe, I. (1994). Dynamic clustering of bacterial population. *Physica D, 75*, 81.

22. Kondrashov, A. S., & Kondrashov, A.F. (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature, 400*, 351.

23. Lande, R. (1981). Models of speciation by sexual selection on phylogenic traits. *Proceedings of the National Academy of Sciences of the USA, 78*, 3721.

24. Margulis, L. (1981). *Symbiosis in cell evolution*. San Fransisco: W. H. Freeman and Company.

25. Maynard-Smith, J., & Szathmary, E. (1995). *The major transitions in evolution*. San Fransisco: W. H. Freeman.

26. Maynard-Smith, J. (1966). Sympatric speciation. *The American Naturalist, 100*, 637.

27. Opitz, J.M. (1981). Some comments on penetrance and related subjects. *American Journal of Medical Genetics, 8*, 265.

28. Osawa, S., & Jukes, T. H. (1989). Codon reassignment (codon capture) in evolution. *Journal of Molecular Evolution, 28*, 271.

29. Suzuki, T., Ueda, T., & Watanabe, K. (1997). The "polysemous" codon—a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. *The EMBO Journal, 16*, 1122.

30. Turner, G.F., & Burrows, M.T. (1995). A model for sympatric speciation by sexual selection. *Proceedings of the Royal Society of London, Section B, 260*, 287.

31. Tyson, J.J., Novak, B., Ordell, G.M., Chen, K., & Thron, C.D. (1996). Chemical kinetic theory: Understanding cell-cycle regulation. *Trends in Biochemical Science, 21*, 89–96.

32. Waddington, C.H. (1957). *The strategy of the genes*. Bristol, UK: George Allen & Unwin.

33. Xu, W., Kashiwagi, A., Yomo, T., & Urabe, I. (1996). Fate of a mutant emerging at the initial stage of evolution. *Researches on Population Ecology, 38*, 231.