The Evolutionary Theory along the Phylogenetic Tree of Unicellular Organisms

ABSTRACT

In parallel to the advancement of molecular biology suggesting the evolutionary route from the RNA world to the DNA-RNA-protein world through the RNA-protein world, the analysis on the nucleotide base changes in ribosomal RNA genes revels that the divergence of prokaryote and eukaryote first occurred in the DNA-RNA-protein world but the divergence of present-day lineages in eukaryotes occurred after the ancestral eukaryote has acquired the mitochondria as the endosymbionts of O_2-respiratory eubacteria. These results strongly suggest that the divergence of unicellular organisms is closely related with the drastic changes in cell contents such as the formation of ribosome, the generation of RNA genes, the conversion from the RNA genes to DNA genes, gene duplication or the endosymbiosis. In the present paper, such drastic changes in evolution of unicellular organisms are theoretically formulated on the basis of the concept of biological activity which consists of the acquired energy, stored energy and systematization. Upon the drastic change in cell contents, the biological activity of an organism is first lowered by the increase in stored energy, but it is gradually recovered as the systematization advances to increase the acquired energy overcoming the increased stored energy and negative entropy due to the systematization itself. Throughout this process, the divergence of new style organisms occurred utilizing new material and energy sources. Such evolutionary process will be explained in detail at the main divergence points in the phylogenetic tree. The relation of such drastic changes in evolution with Darwinian evolution is also clarified.

Keywords: Biological activity; divergence; DNA-RNA-protein world; endosymbiosis; gene duplication; RNA-protein world.

1. INTRODUCTION

The gradual accumulation of slight variants by selection is proposed for the origin of new species from the observation of animals and plants living in different regions and of domestic ones [1]. The core of this proposal has become evidenced, after the re-discovery of Mendelian heredity and the detection of hereditary variants, i. e., mutants. Thus, extensive investigations have been carried out for the behavior of mutants, especially in *Drosophila* population [2,3,4]. In parallel, the population genetics is mathematically formulated to estimate the probability that a newly arisen mutant is fixed in a population depending on its selective advantageousness and population size [5,6]. This study also finds that a selectively neutral mutant can be fixed with the probability equal to the mutation rate independently of population size.

The gene and genome sequencing has brought new information about the evolution of organisms. First, the number of synonymous base changes at the third codon positions in orthologous protein genes between different species is found to be approximately proportional to the divergence time of corresponding species estimated from the fossil record, and the change rate is estimated to be about 10^{-9} per site per year independently of the life spans of organisms [7,8]. This probably reflects the accuracy of repairing damaged DNA bases, implying the higher accuracy of DNA replication through proof-reading. Although this base change rate is only useful for reconstructing the phylogeny of organisms diverged during the period from 10^8 to $4x10^8$ years ago, the simple enumeration of base changes over a whole region of ribosomal RNAs detects the divergence of three kingdoms, archaebacteria, eubacteria and eukaryote [9] as well as the reconfirmation that mitochondria and chloroplasts are the endosymbionts of O_2-respiratory Proteobacteria and Cyanobacteria, respectively [10,11]. This is due to the situation that the nucleotide bases in the stem regions of ribosomal RNAs are under the functional constraint of forming stable base-

pairs while most of the nucleotide bases in the other regions are under the stronger constraint of functions such as the interaction with the transfer RNAs, polymerization of amino acids, kingdom specific base sequence complementary to the translation initiation sequence in mRNA. Therefore, the more careful analysis on the base-pair changes in the stem regions of small and large ribosomal subunit RNAs, whose rate is estimated to be ~$8x10^{-11}$ per pair per year, leads to the clearer divergence pattern of organisms with the time scale during the period from 10^9 to $5x10^9$ years ago [12,13]. The second result of gene and genome sequence is the finding of amino acid sequence similarities between paralogous proteins, proposing the expansion of repertoire of protein functions by gene duplication and by the succeeding nucleotide base substitutions, partial insertion and deletion and further domain shuffling in the counterpart of duplicated genes [14,15,16,17].

In parallel, the molecular biology has revealed the molecular organization of the organisms in the three kingdoms, suggesting that the organism first started in the RNA world, then evolved to the RNA-protein world and finally to the DNA-RNA-protein world [18,19].

In the present paper, it is theoretically explained how the molecular organization has been formed in the unicellular organisms of three kingdoms along the phylogenetic tree. This explanation is based on the mathematical formulation of evolution and innovation proposed previously [20] and also provides the molecular basis for the evolution towards the multicellular diploid eukaryotes, which will be investigated in the next paper.

2. BIOLOGICAL ACTIVITY

The selective advantageousness of a mutant used in the population genetics is a superficial parameter for investigating the behavior of a mutant in a population. The free energy gives a measure for a physicochemical reaction to proceed towards the equilibrium with its minimum value, and it is not suitable for characterizing an organism that grows taking material and energy source from the outside and self-reproduces. As proposed previously [21], the more adequate thermo-dynamical quantity characterizing the state of an organism is the biological activity *BA*, which is defined by

$$BA(M; N_i, S_{Ni}) \equiv E_a(M, N_i, S_{Ni}) - E_s(N_i, S_{Ni}) - TS_{Ni} \quad (1)$$

The first term E_a on the right hand side is the energy acquired by the organism and is a mathematical function of material and energy source M available from the outside, the size N_i of information carrier i and its products, and the systematization S_{Ni} of them. The second term E_s is the energy stored in information carrier and its products, and the third term is the negative entropy – S_{Ni} due to the systematization of them multiplied by temperature T. The difference between the acquired energy E_a and the stored energy E_s is released as heat and this entropy production must compensate for the negative entropy – S_{Ni}, according to the second law of thermodynamics. Thus, the biological activity *BA* is positive definite, and is considered to be proportional to the self-reproducing rate of an organism as the first approximation. The evolution of organisms is considered to be the process that the negative entropy – S_{Ni} is maintained and extended to increase the difference between the acquired energy and stored energy through self-reproduction and selection [22].

Although the information carrier is well defined as the DNA genome of an organism in the DNA-RNA-protein world, (N_i, S_{Ni}) of an organism in the RNA-protein world is denoted as the internal variable x_i of cell contents in the next section and its changes, especially from the RNA-protein world to the DNA-RNA-protein world, are indicated explicitly in the section 4.

3. THE POPULATION BEHAVIOR OF UNICELLULAR ORGANISMS

This formulation is carried out on the larger time scale than the scale treated by the population genetics to investigate the longer scale of evolution. In the population of unicellular organisms taking a common material and energy source M, the number $n(x_i)$ of variants with the internal variable x_i obeys the following time-change equation:

$$\frac{d}{dt}n(x_i; t) = \{R(M; x_i) - D(x_i)\}n(x_i; t) + \sum_j q_{xi,xj}(t)R(M; x_j)n(x_j; t) \quad (2)$$

Here, the self-reproducing rate and the death rate of a variant x_i are denoted by $R(M; x_i)$ and $D(x_i)$, respectively, and $q_{xi,xj}(t)$ is the mutation term from the variant x_j to the variant x_i , $q_{xi,xi}(t)$ being defined by.

$$q_{xi,xi}(\mathrm{t}) \equiv -\sum_{j\neq i} q_{xj,xi}(t) \quad (3)$$

The behavior of the above population becomes transparent when Eq. (2) is transformed into the following two types of equations; one concerning the total number of organisms defined by $B(\mathrm{t}) = \sum_i n(x_i\,;t)$ and another concerning the fraction $f(x_i;t)$ of the variants x_i defined by $n(x_i;t)/B(t)$. By a simple calculation, these two types of equations are obtained in the following forms, respectively.

$$\frac{d}{dt}B(t) = W_{av}(M;t)B(t) \tag{4}$$

$$\frac{d}{dt}f(x_i;t) = \{W(M;x_i) - W_{av}(M;t)\}f(x_i;t) + \sum_j q_{xi,xj}(t)R(M;x_j)f(x_j;t) \tag{5}$$

Here, the increase rate $W(M; x_i)$ of a variant x_i and the average increase rate $W_{av}(M; t)$ in the population are defined by the followings, respectively.

$$\mathrm{W(M};x_i) \equiv R(M;\,x_i) - D(x_i) \tag{6}$$

$$W_{av}(M;t) \equiv \sum_i W(M;\,x_i)f(x_i;\,t) \tag{7}$$

In the case when Eq, (5) is treated by the first order of approximation concerning the mutation term, the set of Eqs. (4) and (5) represents the Darwinian evolution in the following way. If the increase rate $W(M; x_i)$ of the variant x_i generated from the second term on the right side of Eq. (5) is higher than the average increase rate, i. e., $W(M; x_i) - W_{av}(M; t) > 0$, the fraction $f(x_i; t)$ of the variants x_i increases with time according to the first term of Eq. (5). The increase in the fraction of such variants x_i gradually raises the average increase rate $W_{av}(M; t)$, resulting in the increase in the total number $B(t)$ of organisms according to Eq. (4). On the other hand, the fraction $f(x_i; t)$ decreases if $W(M; x_i) - W_{av}(M; t) < 0$. Thus, the organisms taking a common material and energy source M are elaborated by mutation and selection, and most of them finally reach the ones with the optimum increase rate $W(M; x_o)$ each characterized by x_o.

However, even the unicellular organisms have experienced more drastic change in cell contents that transiently decreased the increase rate but then recovered it as a new style of organisms, diverging from the original style of organisms, as will be indicated in the next section. To derive such a drastic change in evolution, Eq. (5) will be formally integrated with respect to time t.

$$f(xi;t = \exp\,[\int_0^t\{W(M;\,x_i) - W_{av}(M;\,\tau)\}d\tau][\int_0^t \sum_j q_{xi,xj}\,(\tau)R\big(M;x_j\big)f\big(x_j;\,\tau\big)\exp\big[-\int_0^\tau\{W(M;\,x_j\big) - W_{av}(M;\,\tau^{'})\}d\tau^{'}\big]d\tau + f(x_i;0)] \tag{8}$$

By setting $x_i = x_{d1}$, $x_j = x_o$, and $W_{av}(M; t) \cong W(M; x_o)$ and by denoting the time average of $q_{xd1,xo}(t)$ as the transition probability $q_{xd1,xo}$, Eq. (8) is reduced to the following form for the sufficiently long time $t >> 0$.

$$f(x_{d1}) = \frac{q_{xd1,xo}R(M;\,x_o)}{W(M;\,x_o) - W(M;x_{d1})}f(x_o) \tag{9}$$

This equation indicates that the variant x_{d1} with the increase rate $W(M, x_{d1})$ lowered by a drastic change from x_o to x_{d1} in the internal variable continues to be present as a minor member in the population. By inserting the expression (9) into the second term on the right side of Eq. (5) and integrating with respect to time t, the fraction $f(x_{d2})$ of variants x_{d2} which have suffered the second step of drastic change is also expressed as

$$\mathrm{f}(x_{d2}) = \frac{q_{xd2,xd1}R(M;xd1)}{W(M;x_0) - W(M;x_{d2})}\frac{q_{xd1,xo}R(M;x_{o)}}{W(M;x_o) - W(M;x_{d1})}f(x_o) \tag{10}$$

Here, $q_{xd2,xd1}$ is the transition probability from x_{d1} to $x_{d2.}$ Such procedure can be continued to obtain the fraction $f(x_{dk})$ of variants x_{dk} that have experienced k steps of drastic changes in the internal variable.

$$\mathrm{f}(x_{dk}) = \prod_{m=2}^{k}\frac{q_{xdm,xdm-1}R(M;x_{dm-1})}{W(M;x_0) - W(M;x_{dm})}\frac{q_{xd1,x0}R(M;x_0)}{W(M;x_0) - W(M;x_0)}f(x_0) \tag{11}$$

Among such minor members in the population, new styles of organisms can appear. When the new style organisms take a common material and energy source M more efficiently than the original style organisms, the new style organisms become predominant, compelling the latter to extinction. When the new style organisms utilize the new material and energy source different from M, on the contrary, both styles of organisms coexist, showing the divergence of them in the phylogenetic tree. In this case, total number B of organisms increases by expanding the repertoire of available material and energy sources. Although this formulation is proposed previously for the evolution by the DNA gene duplication [23], this also provides a guide to explain the other drastic evolution of organisms, as will be shown in the next section.

4. EXPLANATION OF MAIN STAGES IN THE EVOLUTION OF UNICELLULAR ORGANISMS

The essence of the phylogeny of organisms is shown in Fig. 1, which is reconstructed by the analyses on the base-pair changes in the stem regions of small and large ribosomal subunit RNA genes [12, 13]. In this figure, the time scale is given on the basis of the divergence time of animal and green plant, which is estimated to have been 1.2×10^9 years ago [24,25]. The divergence times of many lineages in the three kingdoms are mostly consistent with the geological records of O_2 concentration on the Earth as indicated later but the divergence time of prokaryote and eukaryote is assigned to have occurred much more anciently than that shown by broken lines. This strongly suggests that the divergence of prokaryote and eukaryote occurred in the self-reproducible unicellular organisms upon the conversion from the RNA-protein world to the DNA-RNA-protein world. Thus, the application of the present theory starts from the birth of self-reproducible unicellular organisms in the RNA-protein world.

4.1 The Generation of Ribosome and RNA Protein Genes in the Self-Reproducible Organisms

The studies of molecular biology suggest that the organisms have started from self-replicative RNAs called the RNA replicases [26] and then the variant RNA replicases interacted with amino acids to form polypeptides [18]. In such RNA-protein world, some of the polypeptides would have catalyzed the synthesis of lipids to produce primitive cells. Among such primitive cells, some would have further become self-reproducible by the antagonistic balance of lipid synthesis and cell wall construction [27]. The self-reproduction of a cell is essentially important in improving the cell contents consisting of variants RNA replicases and polypeptides. First, the activity of variant RNA replicases to polymerize amino acids produces primitive RNA polymerase, and the stronger activity of RNA polymerase increases the concentration of various kinds of RNAs in the self-reproducible cell by Darwinian evolution according to Eq. (5). But, the higher concentration of RNAs brings about the interference between them including those trapping amino acids. The cells suffering such interference are then declined to the minor members in the population as shown in Eqs, (9) ~ (11). Among such minor members, however, a new style of self-reproducing unicellular organisms appear yielding primitive ribosome, transfer RNAs and initial complex of translation, and their fraction gradually become predominant in the population. If the cell contents consisting of ribosome and other RNAs are newly denoted as x_i again, the behavior of new style organisms is also expressed by Eqs. (4) and (5) Throughout this process, the translation apparatus is elaborated to establish the universal codon usage and the RNAs besides those associated with the translation are converged into the RNA genes of proteins suitable for raising the self-reproducing rate of the organism [27]. In addition to the elaborated RNA genes of RNA polymerase and of the proteins for lipid synthesis and cell wall construction, the RNA genes of enzyme proteins in and around the pathway of glycolysis from glyceraldehyde 3-phosphate to pyruvate would have appeared at this stage. In practice, the considerable regions of these enzyme proteins show the similarities to the proteins concerned with the synthesis of cytoplasmic membrane [28].

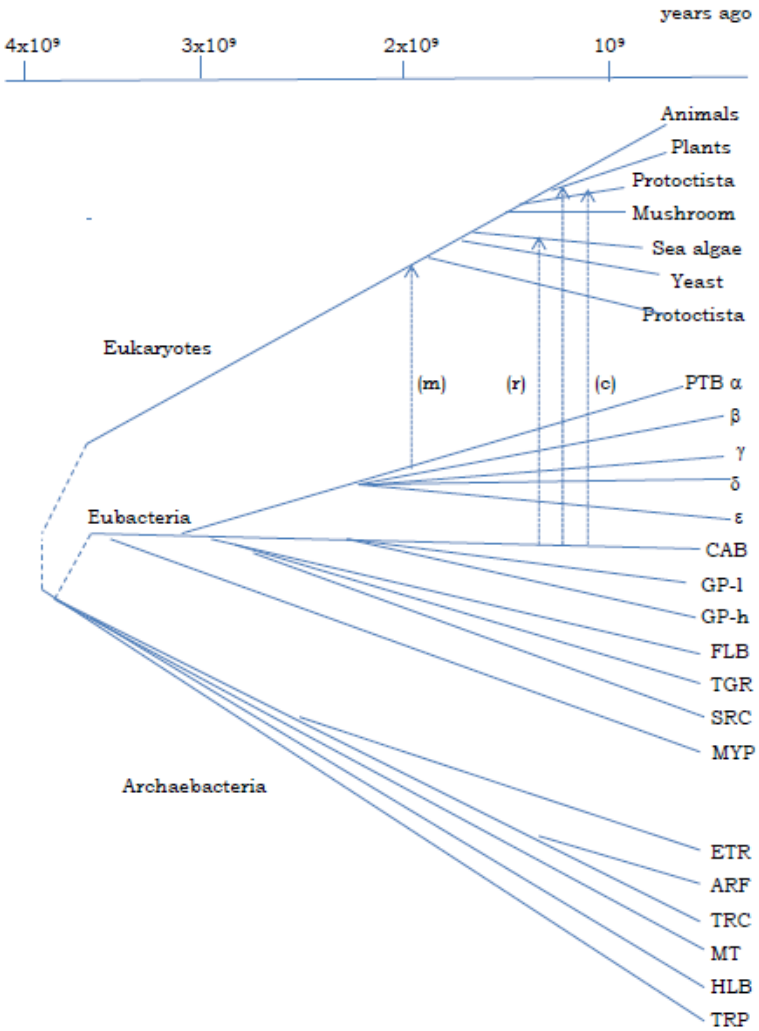

Fig. 1. The Phylogeny of Organisms in the Three Kingdoms

After the first divergence of prokaryote and eukaryote, the divergence of archaebacteria and eubacteria occurred, although these divergence times are estimated to have been more ancient than those shown by broken lines in the original studies [12,13]. Then, the divergence of many lineages occurred in archaebacteria and slightly later in eubacteria. On the contrary, the divergence of many lineages in the eukaryotes occurred after the ancestral eukaryote acquired the mitochondria as the endosymbionts of O_2-respiratory Proteobacteria α at the time denoted by a broken arrow (m). Among the lineages diverged in eukaryotes, the ancestor of sea algae, some of the unicellular eukaryotes called the Protoctista, and green plants further acquired the rhodoplast and chloroplast, respectively, as the endosymbionts of Cyanobacteria separately at the times denoted by broken arrows (r) and (c). Abbreviation of lineages in eubacteria and archaebacteria: PTB; Proteobacteria, CAB; Cyanobacteria, GP-l; Gram Positive low G+C, GP-h; Gram Positive high G+C, FLB; Flavobacteria, TGR; Thermotogales, SRC; Spirochates, MYP; Mycoplasmas, ETR; Extreme thermophiles, ARF; Archaeoglobus fulgidus, TRC; Thermococcus, MT; Methanobacteria, HLB; Halobacteria, TRP; Thermoplasma

4.2 The Conversion of RNA Genes and Ribosomal RNAs to DNA Genes

The glycolysis converts ADPs to ATPs, releasing protons. The production of ATPs is favorable for the further biosynthesis, but the released protons expose the RNAs to deoxidization. The DNAs thus generated would have been first rubbish only to increase the stored energy. Thus, the organism accumulated DNAs again decline to the minor members in the population, their fraction $f(x_{d1})$ being expressed by Eq. (9), if the optimal cell contents consisting of ribosome and RNA genes are denoted again by x_o. Some of such variants further decline to the lower fraction $f(x_{d2})$ in deriving DNA dependent RNA polymerase and DNA polymerase from the RNA gene of RNA polymerase. In fact, the principal component analysis of amino acid sequences reveals that these polymerases form a superfamily with the mutually similarities between them [29]. Although the variants having acquired

such DNA-dependent polymerases turn to raise slightly their increase rate using the RNA genes and DNA genes in parallel, together with the appearance of protein to regulate and control the deoxidization of ribonucleotides, the fractions of these intermediate variants are also expressed in the form of Eqs. (10) ~ (11), confronting with the problem how the replicated DNAs are equi-partitioned into daughter cells. This is because a single set of stable DNA genes is sufficient for the survival of an organism, although many sets of separate RNA genes are almost equivalently partitioned into daughter cells without a special apparatus in the RNA-protein world. This problem has been resolved in two different ways. The ancestral prokaryote is directed to the evolution that the DNA genes are fused to a single circular molecule of double stranded chain and the bi-directional replication started from a specific point (replication origin) of such a DNA molecule takes place attaching to the dividing cell membranes. The ancestral eukaryote, on the other hand, is directed to the evolution that the replicated DNAs in each chromosome are separated to the opposite sides of primitive spindle poles inside the cytoplasmic membrane by the microtubules upon cell division, at the stage when DNA genes have been fused to plural number of linear chains. The phosphorylation signals would have been associated with the formation of such microtubules and spindle pole upon cell division.

This divergence of prokaryote and eukaryote is mathematically described in the following way. By denoting the DNA genomes of ancestral prokaryote and eukaryote as y_p and y_{eu}, respectively, the fractions $f(y_p, t)$ and $f(y_{eu}, t)$ of these two types of organisms turn to increase according to the following equations,

$$\frac{d}{dt}f(y_p,t) = \{W(M;y_p) - W(M;t)\}f(y_p t) + Q_{yp,xdk,x0}(t)R(M;x_0)f(x_0,t) \quad (12)$$

$$\frac{d}{dt}f(y_{eu},t) = \{W(M;y_{eu}) - W(M;t)\}f(y_{eu},t) + Q_{yeu,xdk',x0}(t)R(M;x_0)f(x_0,t) \quad (13)$$

Here, $Q_{yp,xdk,x0}(t)$ and $Q_{yeu,xdk',x0}(t)$ are the transition terms from the optimal organism x_0 in the RNA-protein world to the ancestral prokaryote y_p and eukaryote y_{eu}, respectively, through the intermediate variants x_{dk} and $x_{dk'}$ consisting of the mixture of RNA and DNA genes. The increase rates $W(M;y_p)$ and $W(M;y_{eu})$ of prokaryote and eukaryote becomes larger than the average increase rate $W(M;t)$ of organisms defined by

$$W(M;t) \equiv W(M;y_p)f(y_p;t) + W(M;y_{eu})f(y_{eu};t) + \sum_i W(M;x_i)f(x_i;t) \quad (14)$$

This is because such organisms in the DNA-RNA-protein world can raise the biological activity by decreasing the death rate and by controlling the amount of transcribed RNAs depending on the environment.

At the stage of this divergence, the accuracy of repairing the damaged bases and/or of proof-reading upon DNA replication would have been lower than those at the later time. This may cause the overestimation of the divergence time of prokaryote and eukaryote in the previous papers [12,13].

4.3 Subsequent Evolution of Prokaryotes

It is the prokaryote that first shows the divergence, as seen in Fig. 1. This means that the biological activity of prokaryote first became high enough to allow the existence of variants carrying duplicated DNA genes. From such variants, the divergence of archaebacteria and eubacteria occurs to utilize new material and energy sources *L* and *K*, respectively, instead of the organic compounds *M* synthesized non-biologically which are decreased as most surfaces of the Earth become cooler. The fractions $f(y_{ab},t)$ and $f(y_{eb},t)$ of archaebacteria and eubacteria increase with the transition terms $q_{yab,yp}$ and $q_{yeb,yp}$ from the ancestral prokaryote y_p by the following equations, respectively.

$$\frac{d}{dt}f(y_{ab},t) = \{W(L;\, y_{ab}) - W_p(t)\}f(y_{ab},t) + q_{yab,yp}(t)R(M,y_p)f(y_p,t) \quad (15)$$

$$\frac{d}{dt}f(y_{eb},t) = \{W(K;y_{eb}) - W_p(t)\}f(y_{eb},t) + q_{yeb,yp}(t)R(M;y_p)f(y_p,t) \quad (16)$$

Here, $W(L;y_{ab})$ and $W(K;y_{eb})$ are the increase rate of an archaebacterium and that of a eubacterium, respectively, and the average increase rate of prokaryotes is defined by

$$W_p(t) \equiv W(L;y_{ab})f(y_{ab},t) + W(K;y_{eb})f(y_{ab}) + W(M;y_p)f(y_p;t) \quad (17)$$

The archaebacteria that stay in the region under relatively high temperature are favorable for the initiation of simple chemical syntheses from energy rich compounds by the small steps of nucleotide base changes in the counterparts of duplicated genes. The fraction $f(y_{abLhdk};t)$ of

intermediate variants y_{abLhdk} carrying *h* th set of *k* kinds of duplicated genes is expressed as Eq. (11), if the internal variable x_{dk}, x_o and material and energy source *M* are replaced by y_{abLhdk}, y_{abL} and *L*, respectively. Then, a new lineage y_{abLh} utilizing new material and energy source L_h appears by the following time-change equation of its fraction $f(y_{abLh};t)$.

$$\frac{d}{dt}f(y_{abLh};t) = \{W(L_h;y_{abLh}) - W_{ab}(t)\}f(y_{abLh};t)$$

$$+Q_{yabLh,yabLhdk,yabL}R(L,y_{abL})f(y_{abL};t) \quad (18)$$

Here, $Q_{yabLh.yabLhdk,yabL}$ is the transition term from y_{abL} to y_{abLh} through the intermediate variant y_{ablhdk}, and the average increase rate $W_{ab}(t)$ in the population of archaebecteria is defined by

$$W_{ab}(t) \equiv \sum_h W(L_h, y_{abLh})f(y_{abLh};t) + W(L,y_{abL})f(y_{abL};t) \quad (19)$$

In this way, the divergence of different lineages y_{abLh} of archaebacteria occurs seeking for different new material and energy sources L_{Lh} (h=1, 2, 3,⋯⋯).

Meanwhile, the eubacteria have adapted themselves to the climate change in cooling surface of the Earth by evolving the ingenious regulatory mechanism of repressors and operons in the transcription [30] as well as the expansion of DNA dependent DNA polymerases to families *A* and *C*, while the archaebacteria and eukaryote retain the family *B* of DNA polymerase. This means that the decrease in biological activity due to the advance of systematization is smaller at the lower temperature. The divergence of different lineages then occurs to acquire the material and energy from various sources by gene duplication and succeeding base changes in the counterpart of duplicated genes. Each process of such divergence from the ancestral eubacteria in Eq. (16) is also expressed mathematically by the same form as Eq. (18) if y_{ab}, *L*, and $W_{ab}(t)$ are replaced by y_{eb}, *K* and $W_{eb}(t)$ respectively. Among such diverged lineages, the Proteobacteria α performing O_2-respiration and the Cyanobacteria attaining the O_2-releasing photosynthesis are especially noticeable in the sense that they have brought about much influence upon the evolution of eukaryote. The amino acid sequence similarities suggest that the electron transfer proteins each suspending a heme in the O_2-respiratory system and the proteins suspending chlorophylls in photoreaction centers are derived from the ubiquitous permeases and that the ATP synthetase is derived from the DNA-associated proteins *RecA* and *Rho* by gene duplication [31].

The O_2-respiration in Proteobacteria α transacts the protons and electrons released from the glycolysis including the TCA cycle in the following way; a series of membrane protein complexes pumps out the protons coupled with the electron transfer between them and the final electron acceptor, cytochrome c oxidase, converts the remained protons into H_2O molecules with the oxygen molecules in the cytoplasm. Using the lower proton concentration thus produced in the cytoplasm, the ATP synthetase produces the larger amount of ATPs from ADPs. However, the lowered proton concentration in the cytoplasm restricts the further extension of synthetic pathways.

The O_2-releasing photosynthesis in Cyanobacteria takes place by the photosystems I and II in the thylakoid membrane; the photosystem I forms NADPH from $NADP^+$ and H^+ to fix carbon-dioxides outside the thylakoid through the membrane-bounded iron sulfur protein by accepting electrons from the photosystem II where H_2O molecules in the thylakoid are decomposed into O_2 molecules and protons using the photon energy. Moreover, the proton concentration in the thylakoid is further raised by pumping protons into the thylakoid coupled with the electron transfer from the photosystem I to the photosystem II. The ATP synthetase converts ADP to ATP by passing through the high concentration of protons in the thylakoid to the cytoplasm. However, the outer cytoplasmic membrane also contains the O_2-respiration system to lower the proton concentration in the cytoplasm.

Thus, the genome sizes of Cyanobacteria and Proteobacteria remain around 10^6 *bp*, which are almost the same as those of other eubacteria slightly larger than those of archaebacteria [32]. This indicates the limit of the cell structure of a prokaryote, although the increasing O_2 concentration on the Earth has made many other prokaryotes facultative anaerobic and aerobic, bringing about the second stage of divergence, e. g., the divergence of subdivisions α, β, γ, δ and ε in Proteobacteria and the divergence of *Extreme thermophyles*, *Archaeoglobus fulgidus* and *Thermococcus* in archaebacteria.

4.4 Evolution of Eukaryotes; Acquirement

of Mitochondria and Chloroplasts

In spite of the early appearance of ancestral eukaryote, the lineage leading to the present-day eukaryotes remained single until it acquired the mitochondria as the endosymbionts of O_2-respiratory Proteobacteria α. During this long period, the ancestral eukaryote has probably survived as the predator of prokaryotes, first archaebacteria and then eubacteria, after the decrease in the organic compounds synthesized non-biologically, consuming the energy for phosphorylation signals to evolve the cytoskeleton that retains the cell shape instead of cell wall, endocytosis and exocytosis as well as the nuclear membrane. Such cell structure and living style of ancestral eukaryote, however, have the chance to acquire the mitochondria. This time denoted in Fig. 1 corresponds to the stage 2 of geological period (2.45×10^9 ~ 1.85×10^9 years ago) during which O_2 molecules were produced but absorbed in oceans and seabed rock [33].

When some of the eukaryotes y_{eu} taking the material and energy source P captured the O_2-respiratory eubacteria y_{eb} as the endosymbionts, its nutrient would have been first stolen by the endosymbionts and such variants $y_{eu}y_{eb}$ have been declined to a minor member in the population, their fraction $f(y_{eu}y_{eb})$ being expressed as

$$f(y_{eu}y_{eb}) = \frac{q_{yeuyeb,yeu}R(P,y_{eu})}{W(p\ ,y_{eu})-W(P,y_{eu}y_{eb})} f(y_{eu})f(y_{eb}) \qquad (20)$$

Here, $W(P,y_{eu})$ and $W(P,y_{eu}y_{eb})$ are the increase rate of the eukaryote without an endosymbiont and that of the eukaryote carrying the endosymbiont, respectively, and $q_{yeuyeb,yeu}$ is the probability of capturing endosymbiont y_{eb}. In the host-endosymbiont relationship, the survival of the host eukaryote is a necessary condition, because the extinction of host leads to the extinction of endosymbionts themselves. The first way for the survival of host eukaryote is to utilize the ATP's produced in the endosymbiont by connecting the glycolysis in the host with the O_2-respiration in the endosymbionts. The second way is to delete the genes in the endosymbiont overlapping those in the host genome and the third is to carry out the gene transfer from the endosymbiont genome to the host genome to regulate and control the proliferation of endosymbiont. In fact, such gene transfer is reconfirmed [34]. The DNA dependent RNA polymerase is degenerated to be monomeric in the mitochondria. Throughout such drastic changes in endosymbiont from y_{eb} to z, the fraction $f(y_{euz};t)$ of eukaryotes having succeeded to obtain the mitochondria z turns to increase by the following equation.

$$\frac{d}{dt}f(y_{eu}z;t) = \{W(P,y_{eu}z) - W_{eu}(t)\}f(y_{eu}z;t) + q_{yeuz,yeuyeb}(t)R(P;y_{eu}y_{eb})f(y_{eu}y_{eb};t) \qquad (21)$$

Here, the increase rate $W(P,y_{eu}z)$ of a new style eukaryote with the mitochondria is larger than the average increase rate $W_{eu}(t)$ of eukaryotes defined by

$$W_{eu}(t) \equiv W(P,y_{eu}z)f(y_{eu}z;t) + W(P,y_{eu}y_{eb})f(y_{eu}y_{eb};t) + W(P,y_{eu})f(y_{eu};t) \qquad (22)$$

The eukaryotes having acquired the mitochondria not only become predominant by raising their biological activity under the supply of abundant ATPs and protons from the mitochondria but also allow the existence of variants carrying duplicated genes to yield new genes responsible for the divergence of unicellular eukaryotes called the Protoctista, fungi, animals and plants, as seen in Fig. 1. The repertoire of DNA dependent DNA polymerases would have been expanded within the family B after the acquirement of mitochondria.

Among the diverged lineages, some of Protoctista, some of the fungi to become sea algae and the ancestor of higher plants have further acquired as Cyanobacteria as the endosymbionts so as to cooperate with the mitochondria in the host cell [35]. These endosymbionts become the rhodoplasts in red algae and chloroplasts in green plants by deleting the considerable kinds of protein genes including the O_2-respiratory system in the outer membrane and by transferring many genes to the host genome [34]. The times estimated for these eukaryotes to have acquired rhodoplasts or chloroplasts in Fig. 1 fall into the stage 3 (1.85×10^9 ~ 0.85×10^9 years ago) during which O_2 molecules started to gas out of the ocean and to form the ozone layer [33]. This process of eukaryote of acquiring the chloroplast or rhodoplast w as the endosymbionts of Cyanobacteria is also expressed by Eqs. (20) and (21), if $y_{eu}y_{eb}$, y_{eu} and y_{euz} are replaced by $y_{euz}y_{cy}$, y_{euz} and $y_{euzw,}$ respectively. The fraction $f(y_{euzw},t)$ of eukaryotes succeeded to acquire these endosymbionts begins to increase with the increase rate $W(y_{euzw})$ larger than the average increase rate $W_{eu}(t)$ defined by

$$W_{eu}(t) \equiv W(P,y_{euz})f(y_{euz},t) +$$

$$W(y_{euzw})f(y_{euzw},t) \quad (23)$$

The rigid cell wall is recovered in these eukaryotes that have acquired chloroplasts or rhodoplasts. By the prosperity of such autotrophic eukaryotes, the atmosphere of the Earth has entered into the stage 4 (from $8.5x10^8$ years ago to the present) during which O_2 molecules are accumulated in the atmosphere [33]. However, all eukaryotes do not accept the Cyanobacteria to become autotrophic but some others remain as either animals or decomposers together with some aerobic eubacteria. This is due to the ecological relation to raise the circular flow of materials [36]. Under such ecological relations, the genome sizes of eukaryotes are expanded to 10^7 *bp* even in yeast and to 10^9 *bp* or more in the higher animals and plants together with the increased number of chromosomes [32]. In particular, the repertoire of kinase genes is expanded to evolve the phosphorylation signals for the multicellularity and cell differentiation in addition to the receptors and ligands. The details of this evolution of eukaryotes will be investigated in the next paper.

5. DISCUSSION AND CONCLUSIONS

The biological activity is a useful measure for evaluating the drastic changes in evolution of organisms. Upon the drastic change in cell contents such as the conversion of RNA genes to DNA genes, duplication of DNA genes or endosymbiosis, the biological activity or increase rate of an organism is first lowered by the increase in stored energy but it gradually recovered by advancement of systematization to increase the acquired energy overcoming the increased stored energy and the entropy reduction due to systematization itself. During the recovery of biological activity, new style(s) of organisms appear to utilize new material and energy sources by generating new genes from the base changes in extra polynucleotides. The Darwinian evolution corresponds to the process of selecting the nucleotide bases changed in a definite set of genes to the optimal increase rate of an organism. It depends on the environment which bases are selected and the region-specific species is generated by the geographical isolation.

While the organisms in the DNA-RNA-protein world have evolved to the higher organization in this way, the original organisms in the RNA-protein world would have survived as RNA phages and viruses, utilizing the materials and energy in the host prokaryote and eukaryote. This is another example showing that the positive value of biological activity can be retained by decreasing the stored energy and systematization.

From the fact that the eukaryotes are similar to the archaebacteria in DNA dependent DNA polymerases while they are similar to eubacteria in the cytoplasmic membrane, the following controversial hypotheses are also proposed for the origin of eukaryote. (*a*) Eukaryotes resulted from the complete fusion of two or more cells, wherein the cytoplasm formed from eubacteria and nucleus from archaeon [37], from a virus [38,39] or from a precell [40]. *(b)* Eukaryote developed from an archaebacterium and acquired their eubacterial characteristics through endosymbiosis of a proto-mitochondrion of eubacterial origin [41]. Although these hypotheses seem to be under the influence of the early estimation by Woese [9] that the divergence of archaebacteria and eubacteria first occurred and then the divergence of archaebacteria and eukaryote occurred, the similarity of eukaryotes to archaebacteria in DNA dependent polymerases strongly suggests the possibility that these DNA polymerases remain almost unchanged since the conversion from RNA genes to DNA genes, according to the more careful estimation [12,13]. Moreover, this possibility also provides a straightforward explanation for the origin of plural number of linear DNA chains in the eukaryotes, distinctive from the single circular DNA genome in both archaebacterial and eubacteria. The similarity of eukaryote to eubacteria in the cytoplasmic membrane is also reasonable because the eukaryotes have also adapted themselves to the cooling surface of the Earth, preying on the eubacteria.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Darwin C. The origin of species. John Murry, London; 1859.
2. Dobzhansky T. Genetics and the origin of species. 2nd edn. Columbia University Press, New York; 1941.
3. Mayer E. Systematics and the origin of species. A correlation of the evidence and points of view of systematics with those of other biological disciplines, Particularly Genetics and Ecology, Columbia University Press, New York; 1942.

4. Simpson GG. Tempo and mode in evolution, A synthesis of Paleontology and Genetics. Columbia University Press, New York; 1944.
5. Fisher RA. The General Theory of Natural Selection. Oxford University Press, London and New York; 1930.
6. Wright S. Adaptation and selection. In: Jepson, GL, Simpson GG, Mayer E. (eds.) Genetics, Paleontology and Evolution. Princeton University Press, Princeton, New Jersey. 1949:365-389.
7. Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217:624-626.
8. Kimura M. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol.1980;16:111-120.
9. Woese CR. Bacterial evolution. Microbiol. Rev. 1987;51:221-271.
10. Yang D, Oyaizu Y, Oyaizu H, Olson GJ, Woese CR. Mitochondrial origin. Proc. Natl. Sci. USA. 1985;82:4443-4447.
11. Van den Eynde H, De Baere R, De Roeck E, Van de Poer Y, Vandenberghe A,Willekens P, De Wacher R. The 5S ribosomal RNA sequences of a red algal rhodoplast and a gymnosperm chloroplast: Implication for the evolution of plastids and cyanobacteria. J. Mol. Evol. 1988;27:126-132.
12. Otsuka J, Terai G, Nakano T. Phylogeny of organisms investigated by the base-pair changes in the stem regions of small and large ribosomal subunit RNAs. J. Mol. Evol. 1999;48:218-235.
13. Sugaya N, Otsuka J. The lineage-specific base-pair contents in the stem regions of ribosomal RNAs and their influence on the estimation of evolutionary distances. J. Mol. Evol. 2002;55:584-594.
14. Ingram VM. The Hemoglobin in Genetics and Evolution. Columbia University Press, New York; 1963.
15. Ohno S. Evolution by Gene Duplication. Spring-Verlag, Berlin; 1970.
16. Gilbert W. Why genes in pieces ? Nature. 1978;271:501.
17. Ferris SD, Whitt GS. Evolution of the differential regulation of duplicated gene after polyploidization. J. Mol. Evol. 1979; 12:267-317.
18. Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM. Molecular Biology of the Gene.; 4th edn. Volume II. Benjamin/Cummings Publishing Company, Inc., New York. 1987;1098-1161.
19. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. Molecular Biology of the Cell. 1994; 3rd edn. Garland Publishing, Inc. New York.
20. Otsuka J. A mathematical formulation of evolution and innovation I. Unicellular organisms. Phy. Sci. & Biophy. J. 2017;1(1):1-10.
21. Otsuka J. The concept of biological activity and its application to biological phenomena. J. Phys. Chem. & Biophys. 2017;7:235-240.
22. Otsuka J. The negative entropy in organisms: Its maintenance and extension. Journal of Modern Physics. 2018;7:2156-2169.
23. Otsuka J. The large-scale evolution by generating new genes from gene duplication; similarity and difference between monoploid and diploid organisms.J. Theor. Biol. 2011;278:120-126.
24. Dickerson RE. The structure of cytochrome c and the rate of molecular evolution. J. Mol. Evol. 1971;1:26-45.
25. Otsuka J, Nakano T, Terai J. A theoretical study on the nucleotide changes under a definite functional constraint of forming stable base-pairs in the stem regions of ribosomal RNAs; its application to the phylogeny of eukaryotes. J. Theor. Biol. 1997;184:171-186.
26. Cech TR. A model for the RNA-catalyzed replication of RNA. Proc. Natl. Acad. Sci. USA. 1986;83:4360-4363.
27. Otsuka J. On the origin of genes. Phy. Sci. & Biophy. J. 2019;3(2):1-9.
28. Fukuchi S, Otsuka J. Evolution of the self-reproducing system to the biosynthesis of membrane: An approach from the amino acid sequence similarity in proteins. J. Theor. Biol. 1996;182:117-136.
29. Otsuka J, Kikuchi N, Kojima S. Similarity relations of DNA and RNA polymerases investigated by the principal component analysis of amino acid sequences. Biochimica et Biophysca Acta. 1999;1434: 221-247.
30. Jacob P, Monod Genetic regulatory mechanism in the synthesis of proteins. J. Mol. Biol. 1961;3:318-356.
31. Otsuka J. An inquiry into the evolutionary history of photosynthetic and respiratory systems from the similarity relationships of member proteins.Recent Res. Devel. Proteins. 2002;1:229-256.

32. Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, Madden TL, Pantinus JU, Shular GD, Schriml LM, Sequeira E, SuzehTO, Tatusova TA, Wagner L. Nucl. Acid Res. 2004;32: Database issue D35.
33. Holland HD. The oxygenation of the atmosphere and oceans. Philosophical Transactions of Royal Society B. B iological Science. 2006;361(1470):903-915.
34. Kleine T, Maier UG, Leister D. DNA transfer from organelles to the nucleus: The idiosyneratic genetics of endosymbiosis. Ann. Rev. Plant Biol. 2009;60:115-138.
35. Otsuka J. A theoretical scheme for the large-scale evolution of organisms towards a higher order of organization and diversity. Recent Research Developments in Experimental and Theoretical Biology. 2005;1:93-122.
36. Otsuka J. A theoretical characterization of ecological systems by circular flow of materials. Ecological Complexity. 2004;1:237-252.
37. Martin W. Archaebacteria and the origin of the eukaryotic nucleus. Current Opinion in Microbiology. 2005;8(6):630-637.
38. Takemura M. Poxviruses and the origin of the eukaryotic nucleus. J. Mol. Evol. 2001; 52(5):419-427.
39. Bell PJ. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus ? J. Mol. Evol. 2001;53(3):251-256.
40. Wächtershäuser G. From volcanic origin of chemoautotrophic life to bacteria, archaea and eukarya. Philosophical transactions of royal society of London, Series B, Biological Science. 2006;361 (1474):1787-1806.
41. Lane N. The vital question: Why is the life the way it is? (paperback edn.) Profile Books. 2016;157-191.
ISBN:978-1-781-25037-2.