MEDWIN PUBLISHERS
Committed to create value for Researchers

Physical Science & Biophysics Journal

A Mathematical Formulation of Evolution and Innovation II. From Unicellular Monoploid Eukaryotes to Multicellular Diploid Eukaryotes

Otsuka J*

JO Institute of Biophysics, Tokyo, Japan

***Corresponding author:** Jinya Otsuka, JO Institute of Biophysics, Higashi-Ohizumi 6-49-18, Nerima-ku, Tokyo 178-0063, Japan, Tel: +813 3921 1466; Email: jin.otsuka@kyj.biglobe.ne.jp

Research Article

Volume 2 Issue 2

Received Date: November 01, 2018

Published Date: November 21, 2018

Abstract

Among the evolutionary lines of eukaryotes after the acquirement of the mitochondria, the present study focuses on the evolution and innovation from unicellular monoploids to multicellular diploids. It is first investigated mathematically how the conjugation of monoploid eukaryotes to exchange homologous chromosomes and the hybridization of diploid eukaryotes are effective on the accumulation of new genes generated from gene duplication. In the monoploid eukaryote containing multiple kinds of chromosomes, the exchange of homologous chromosomes enhances the chance to yield new style monoploids receiving many kinds of new genes sufficient for causing multicellularity and cell differentiation. Although the multicellular diploid eukaryote is the next in the line of fixing a full set of new genes homologously, various variants carrying partial sets of new genes are generated on the way to establish the new style diploids homologously and these variants successively hybridize with other latent variants to yield the next stage of new style diploids. This explains the punctuated mode of explosive divergence of body plans suggested from paleontology. Second, this innovation from the monoploids to diploids is theoretically investigated from the physiological aspect of cell differentiation. Although the cooperative action of differentiated cells is an excellent strategy to acquire the energy and material sources from the outside, the material and energy are also required for the development of cell differentiation and their amount becomes larger in the diploid state than in the monoploid state. On the other hand, the diploid state is suitable for elongating the duration time of differentiated cells against nucleotide base changes. To attain this purpose overcoming the first physiological problem, the eukaryotes have advanced their organization to multicellular diploids through the intermediate stages of alternating the monoploid generation differentiated into female and male types with the diploid one. This innovation process is illustrated in green plants and animals.

Keywords: Cell differentiation; Conjugation; Gene duplication; Hybridization; Organization; Sex

Introduction

Succeeding to the previous paper of unicellular organisms [1], the present study investigates the evolution and innovation from the unicellular monoploid eukaryotes to the multicellular diploid eukarypotes. The biological studies of evolution have started from the observation of multicellular diploid eukaryotes and led to the proposal of the gradual accumulation of selectively advantageous variants for the generation of new species by Darwin [2]. Then, Darwinian evolution is formulated mathematically in population genetics to estimate the probability that a spontaneously arisen mutant is fixed in a population according to its degree of selective advantage [3,4]. This study also reveals that a selectively neutral mutant is fixed with the probability equal to the mutation rate, independently of the population size.

The gene and genome sequencing has brought new information about the evolution of organisms. First, the comparison of orthologous genes from different species finds the selectively neutral nucleotide base changes in the third codon positions and other regions under weak functional constraint [5,6]. Using the neutral base changes, the phylogeny of organisms is reconstructed and its comparison with the fossil record finds that the change rate is almost constant, $2 \sim 3x10^{-9}$ per site per year, independently of the life times of organisms [7,8]. The reconstruction of phylogeny is further extended to a wider range of organisms, using the base-pair changes in the stem regions of ribosomal RNAs (rRNAs) and reveals that proteobacteria, eubacteria and eukaryotes first diverged before $4x10^{9}$ years ago but the divergence of protoctista, fungi, sea algae, green plants and animals occurred after their ancestral eukaryote acquired the mitochondria around $1.8x10^{9}$ years ago as the endosymbionts of O_2-respiratory eubacteria [8-10]. Second, the amino acid sequence similarities of paralogous proteins strongly suggest that the repertoire of protein functions has been expanded by gene duplication and by the succeeding changes in the counterpart of duplicated genes due to the nucleotide base changes, partial deletion and/or insertion, and domain shuffling [11-14]. The clustering analysis of proteome further confirms that the multicellular eukaryotes have especially expanded the families and super families of the proteins responsible for cell differentiation such as cell adhesion, cell-cell communication, intracellular signal transduction and transcription regulation, in comparison with unicellular organisms [15,16].

The main purpose of the present study is to elucidate the mechanism by which so many kinds of member protein genes necessary for multicellularity and cell differentiation are gotten together and the reason why the diploid state is realized. It is first shown mathematically that the accumulation of many kinds of new genes generated from gene duplication is enhanced by the innovation of monoploid eukaryote to exchange homologous chromosomes and this enhancement is continued in the hybridization of diploid eukaryotes. Then, the theoretical investigation is carried out for physiological problems concerning the supply of material and energy to the development of cell differentiation, sexual differentiation and the elongation of duration time of the genome expanded for the higher hierarchy of cell differentiation. This study is based on the concept of biological activity proposed previously [17,18].

Innovation of Monoploid Eukaryotes by Exchanging Homologous Chromosomes through Conjugation

As indicated already [1], the population of unicellular monoploid eukaryotes taking the material and energy source M from the outside is characterized by the following set of two equations; one concerning the total number $B(t)$ of all kinds of variants

$$\frac{d}{dt}B(t) = W_{av}(M;t)B(t) \qquad (1)$$

and another concerning the fraction $f_{xi}(t)$ of variants with the internal variable x_i of genome size and systematization

$$\frac{d}{dt}f_{xi}(t) = \{W(M;x_i) - W_{av}(M;t)\}f_{xi}(t) + \sum_j q_{xi,xj}(t)R(M;x_j)f_{xj}(t) \qquad (2)$$

The increase rate $W(M;x_i)$ of the variant x_i is defined by the self-reproducing rate $R(M;x_i)$ minus death rate $D(x_i)$and the average increase rate of the organisms $W_{av}(M;t)$ is defined by

$$W_{av}(M;t) \equiv \sum_i W(M;x_i)f_{xi}(t) \qquad (3)$$

The mutation term $q_{xi,xi}(t)$ means the mutation of the variant x_i to other kinds of variants, i. e., $-\sum_{j(\neq i)} q_{xj,xi}(t)$.

Otsuka J. A Mathematical Formulation of Evolution and Innovation II. From Unicellular Monoploid Eukaryotes to Multicellular Diploid Eukaryotes. Phy Sci & Biophy J 2018, 2(2): 000112.

Darwinian evolution corresponds to the approximate solution of Equation (2) by considering only the first order of mutation term mainly due to the nucleotide base changes in the genome. By this evolution, organisms are elaborated by the mutation and selection, and most of them reach the ones x_o with the optimum increase rate $W(M;x_o)$. Because the gene duplication occurs less frequently than the nucleotide base changes, the fraction of variants with the biological activity lowered by gene duplication is evaluated after the optimum organisms x_o become dominant in the population. For this purpose, Equation (2) will be solved up to the higher orders of mutation terms by averaging the mutation term $q_{x\mu,x\mu-1}(t)$ from $(\mu-1)$ kinds to μ kinds of gene duplication over a sufficiently long time to be regarded as the rate of gene duplication.

$$q_{x\mu,x\mu-1} = \frac{1}{t}\int_0^t q_{x\mu,x\mu-1}(\tau)d\tau \qquad (4)$$

In this large time scale, Equation (2) gives the following relation between the fraction $f_{x\nu}$ of variants x_ν having experienced ν kinds of gene duplication and the fraction f_{xo} of dominant organisms x_o in a stationary state [1].

$$f_{x\nu} = \prod_{\mu=1}^{\nu} \frac{q_{x\mu,x\mu-1}R(M,x_{\mu-1})}{W(M;x_o)-W(M;x_\mu)} f_{xo} \qquad (5)$$

To express numerically the fraction $f_{x\nu}$ of variants x_ν, the self-reproducing rate of an optimal organism is simply assumed to decrease with a reduction factor r by every step of gene duplication, and the death rate of an organism is assumed to be hardly influenced by gene duplication. Then, Equation (5) is expressed in the following form.

$$f_{x\nu} = \frac{(1-r)(1-2r).....\{1-(\nu-1)r\}}{\nu! r^\nu} Q_\nu f_{xo} \qquad (6)$$

where Q_ν is denoted by

$$Q_\nu = \prod_{\mu=1}^{\nu} q_{x\mu,x\mu-1} \qquad (7)$$

The values of fractions $f_{x\nu}$'s relative to $Q_\nu f_{xo}$ are plotted against the values of ν in Figure 1. As seen in this figure, the fraction $f_{x\nu}$ of variants x_ν decreases as the number ν increases, although the reduction factor r may become smaller than that in the prokaryote by the supply of ATP molecules from the mitochondria.

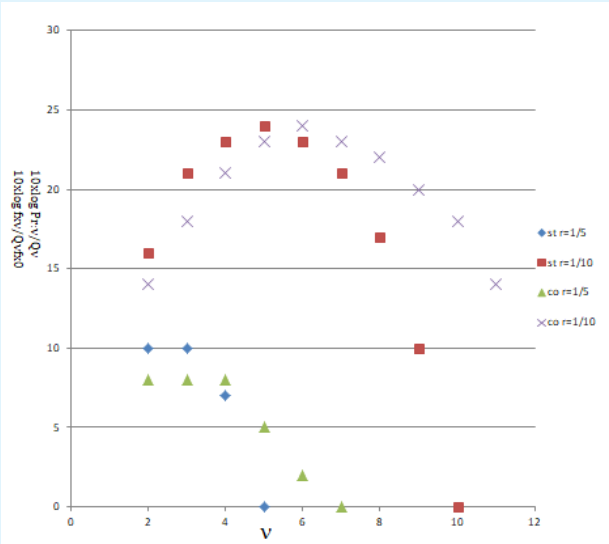

Figure 1: Comparison in accumulating new genes generated from gene duplication between the multiple steps of gene duplication and the exchange homologous chromosomes through conjugation. According to Equation (6), the values of $f_{x\nu}/Q_\nu f_{xo}$ are plotted against the values of ν for two cases of r =1/5 and r = 1/10, which are denoted as *st* r =1/5 and *st* r = 1/10, respectively. Although each of these plotted curves takes a maximum value on the way to $f_{x\nu}$ = 0, the values of $f_{x\nu}$'s are monotonously decreased because the product Q_ν of mutation rates is more decreased as the number ν is increased. The values of $P_{r:\nu}/Q_\nu$ calculated on the basis of Equation (8) for the cases of r =1/5 and r = 1/10 are also plotted as *co* r =1/5 and *co* r = 1/10, respectively, where $P_{r:\nu}$ of even number ν is calculated from $\nu_1=\nu_2=\nu/2$ and that of odd number ν is calculated from $\nu_1=\nu_2+1=(\nu+1)/2$. Although $P_{r:\nu}/Q_\nu$ is slightly smaller than $f_{x\nu}/Q_\nu f_{xo}$ in the smaller number of ν, the former becomes certainly larger than the latter in the larger number of ν for each case of reduction factor r and moreover the number ν, which gives non-zero probability $P_{r:\nu}$ extends to the region where $f_{x\nu}$ is zero. In the case of r = 1/10, for example, $P_{r:\nu}$'s are not zero even in the region of ν from 12 to19, although they are not shown in the figure for simplicity.

However, the exchange of homologous chromosomes through conjugation yields the monoploid variant receiving more kinds of duplicated genes with the higher probability than that expected from the increase rate. This probability depends on the number of chromosomes

carrying duplicated genes, and two extreme cases will be considered.

In the case when the variant carrying ν_1 kinds of duplicated genes separately on ν_1 kinds of chromosomes conjugates with another variant carrying ν_2 kinds of duplicated genes separately on other ν_2 kinds of chromosomes, the zygote produces the daughter monoploids receiving $(\nu_1+\nu_2)$ kinds of duplicated genes with the following probability $P_{r:(\nu 1+\nu 2)}$ by the random partition of homologous chromosomes in each pair.

$$P_{r:(\nu 1+\nu 2)} = (\frac{1}{2})^{\nu 1+\nu 2} \frac{(1-r)(1-2r).....\{1-(\nu_1-1)r\}}{\nu_1! r^{\nu 1}} \frac{(1-r)(1-2r).....\{1-(\nu_2-1)r\}}{\nu_2! r^{\nu 2}} Q_{\nu 1+\nu 2} \quad (8)$$

Among these probabilities, $P_{r:\nu}$'s $(\nu/2=\nu_1=\nu_2$and $(\nu+1)/2=\nu_1=\nu_2+1)$ are also plotted in Figure 1. As seen in this figure, the value of $P_{r:\nu}/Q_\nu$ becomes larger than that of $f_{x\nu}/Q_\nu f_{x0}$ in the region of large number ν and, moreover, non-zero values of $P_{r:\nu}$ extend over the region where $f_{x\nu}$ is zero.

In the case when the variant monoploid carrying ν_1 kinds of duplicated genes in one kind of chromosome conjugates with other variant carrying ν_2 kinds of duplicated genes in another kind of chromosome, the zygote produces the monoploids receiving $(\nu_1+\nu_2)$ kinds of duplicated genes with the probability $P'_{r:(\nu 1+\nu 2)}$ containing the coefficient $(1/2)^2$ instead of $(1/2)^{\nu_1+\nu_2}$ in $P_{r:(\nu 1+\nu 2)}$. This probability $P'_{r:(\nu 1+\nu 2)}$ is much higher than $P_{r:(\nu 1+\nu 2)}$ in Equation (8) especially when ν_1 and ν_2 are large values.

At the stage when the monoploid eukaryotes began the conjugation to exchange homologous chromosomes, they would have carried only several kinds of chromosomes at most and the crossing over between homologous chromosomes would have also occurred. Thus, the probability, with which the monoploid eukaryote received different kinds of duplicated genes through conjugation, would have been intermediate between the first and second cases. At any rate, the conjugation enhances the chance to produce the monoploids receiving many kinds of new genes sufficient for causing multicellularity and cell differentiation, although the monoploid variants not expressing a new character ultimately return to the fraction $f_{x\nu}$.

Evolution of Diploid Eukaryotes by Gene Duplication

For simplicity, we consider the case when the diploid eukaryotes are monoecism. Then, the number $N(x_i,x_j;t)$ of diploid variants characterized by two sets of genetic information carriers x_i and x_j obeys the following time change equation in the population of eukaryotes taking a material and energy source M and exchanging homologous chromosomes by hybridization.

$$\frac{d}{dt}N(x_i,x_j;t) = \sum_{k,l} R(M;x_i,x_j)_{ik,jl} N(x_i,x_k;t)N(x_j,x_l;t) - D(x_i,x_j;t)N(x_i,x_j;t) + \sum_{i',j'}\sum_{k,l} q(x_i,x_j \leftarrow x_{i'},x_{j'};t)_{i'k,j'l} R(M;x_{i'},x_{j'})_{i'k,j'l} N(x_{i''},x_k;t)N(x_{j'},x_l;t) \quad (9)$$

Here, $R(M;x_i,x_j)_{ik,jl}$ is the rate of producing the children (x_i,x_j) from the hybridization of a variant (x_i, x_k) with another variant (x_j,x_l) and $D(x_i,x_j)$ is the death rate of the variant (x_i,x_j). The mutation term $q(x_i,x_j \leftarrow x_i,x_j;t)_{ik,jl}$ is defined by $-\sum_{i',j'(\neq i,j)} q(x_{i''},x_{j''} \leftarrow x_i,x_j;t)_{ik,jl}$.

The population behaviour of diploid eukaryotes also becomes transparent by transforming Equation (9) into the equation concerning the total number of diploid eukaryotes $B_d(t) \equiv \sum_{i,j} N(x_i,x_j;t)$, which includes $N(x_i,x_l;t)$, $N(x_k,x_j;t)$ and $N(x_k,x_l;t)$ produced from the hybridization of $N(x_i,x_k;t)$ with $N(x_j,x_l;t)$, and into the equation concerning the fraction of variants (x_i,x_j) defined as $F(x_i,x_j;t) \equiv N(x_i,x_j;t)/B_d(t)$. These equations are expressed in the following forms, respectively.

$$\frac{d}{dt}B_d(t) = \overline{W}(M;t)B_d(t) \quad (10)$$

$$\frac{d}{dt}F(x_i,x_j;t)=\{W(M;x_ix_j;t)-\overline{W}(M;t)\}F(x_i,x_j;t)$$
$$+\sum_{i',j`}\sum_{k,l}q(x_i,x_j\leftarrow x_{i'},x_{j`};t)_{i'k,j'l}R(M;x_{i`},x_{j`})_{i`k,j`l}F(x_{i'},x_k;t)F(x_{j'},x_lt)B_d(t) \quad (11)$$

Here, the increase rate $W(M;x_ix_j;t)$ of the variants (x_i, x_j) and the average increase rate $\overline{W}(M;t)$ are defined by

$$W(M;x_i,x_j;t)$$
$$\equiv\sum_{k,l}R(M;x_i,x_j)_{ik,jl}F(x_i,x_k;t)F(x_j,x_l;t)B_d(t)/F(x_i,x_j;t)-D(x_i,x_j) \quad (12)$$

and by

$$\overline{W}(M;t)\equiv\sum_{i,j}W(M;x_i,x_j;t)F(x_i,x_j;t) \quad (13)$$

respectively.

When the suffixes i, j, k and l denote the mutation arising mainly from the nucleotide base changes, most of the diploid eukaryotes in this population gradually become the ones (x_o,x_o) with the optimum increase rate $W(M;x_o,x_o)$, making the selectively advantageous bases homologous, by Darwinian evolution. Under this situation, the fractions of variants having experienced gene duplication are derived from Equation (11) by the procedure similar to the derivation of Equation (5) from Equation (2). The fraction $F(x_\nu,x_o)$ of variants carrying ν kinds of duplicated genes is related with the fraction $F(x_{\nu-1},x_o)$ of variants carrying $(\nu\text{-}1)$ kinds of duplicated genes in the following way.

$$F(x_\nu,x_o)=\frac{q(x_\nu,x_o\leftarrow x_{\nu-1},x_o)R(M;x_{\nu-1},x_o)}{W(M;x_o,x_o)-W(M;x_\nu,x_o)}F(x_{\nu-1},x_o)F(x_o,x_o)B_d \quad (14)$$

where $q(x_\nu,x_o\leftarrow x_{\nu-1},x_o)$ is the mutation rate from $(\nu\text{-}1)$ kinds of gene duplication to ν kinds of gene duplication. If the death rate of a diploid eukaryote hardly depends on the gene duplication, i. e., $D(x_\nu,x_o)\approx D(x_o,x_o)$, the denominator on the right side of Equation (14) becomes

$$W(M;x_o,x_o)-W(M;x_\nu,x_o)\approx\{R(M;x_o,x_o)-R(M;x_\nu,x_o)\}F(x_o,x_o)B_b \quad (15)$$

and Equation (14) becomes

$$F(x_\nu,x_o)=\frac{q(x_\nu,x_o\leftarrow x_{\nu-1},x_o)R(M;x_{\nu-1},x_o)}{R(M;x_o,x_o)-R(M;x_\nu,x_o)}F(x_{\nu-1},x_o) \quad (16)$$

This is the same form as for the monoploid organisms. Thus, the relative fraction $F(x_\nu,x_o)$ of variants carrying ν kinds of duplicated genes to the fraction $F(x_o,x_o)$of dominant diploids is also expressed in the same form as Equation (6), if the reproducing rate $R(M;x_\mu,x_o)$ is reduced to $R(1\text{-}\mu r)$ in comparison with the reproducing rate $R(M;x_o,x_o)\equiv R$.

The hybridization of diploid variants carrying different kinds of duplicated genes also yields the children receiving more kinds of duplicated genes than those expected from the decrease in increase rate, but its probability is near to $P_{r:(\nu1+\nu2)}$ in Equation (8) because the diploid eukaryote carries more than ten kinds of chromosomes and most of duplicated genes are distributed separately on different kinds of chromosomes. A remarkable difference in evolutionary pattern between diploid and monoploid eukaryotes comes from the process to fix these new genes arising from duplicated genes. Although a set of new genes suitable for expressing a new style character is immediately fixed in the monoploid eukaryote to form a new population, the

breeding of new style of diploid eukaryotes each carrying a set of new genes heterogeneously still produces various variants. If the homo, hetero, and vacant state concerning a new gene of J^{th} pair of homologous chromosomes are denoted by

$$a_J^2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}_J, \quad a_J b_J = \begin{pmatrix} 1 \\ 0 \end{pmatrix}_J = \begin{pmatrix} 0 \\ 1 \end{pmatrix}_J, \quad \text{and} \quad b_J^2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}_J, \tag{17}$$

respectively, the ratios of different types of children born from the parents each carrying ν kinds of new genes heterogeneously on separate pairs of homologous chromosomes are expressed by the following way.

$$\prod_{J=1}^{\nu}(a_J^2 + 2a_J b_J + b_J^2) = \prod_{J=1}^{\nu}(a_J^2 + 2a_J b_J) + \sum_{J=1}^{\nu} b_J^2 \prod_{I(\neq J)=1}^{\nu}(a_I^2 + 2a_I b_I) + \sum_{J(\neq K)=1}^{\nu} b_J^2 b_K^2 \prod_{I(\neq J,K)=1}^{\nu}(a_I^2 + 2a_I b_I) + \cdots\cdots\cdots + \prod_{I=1}^{\nu} b_I^2 \tag{18}$$

The first term on the right side of Equation (18) corresponds to the children receiving a full set of new genes homologously and heterogeneously, and such children amount to 3^{ν} when the total number of children on the left side is set to be 4^{ν}. The second term on the right side corresponds to the children lacking one kind of new gene, and the number of such children is ${}_{\nu}C_1 3^{\nu-1}$. In the same way, the number of children lacking κ kinds of new genes amounts to ${}_{\nu}C_k 3^{\nu-\kappa}$ and the last term corresponds to the children completely lacking new genes. These children lacking some of new genes return to the population of the original style diploid eukaryotes, not expressing a new morphological character, but serve to produce second stage of new morphological characters by the hybridization with the latent variants carrying other kinds of new genes.

To estimate the probability that the appearance of the first new style diploids induces the second stage of new style diploids, we will follow the descendants of the first new style diploids carrying ν kinds of new genes heterogeneously. Among the children that exhibit the new morphological character in Equation (18), the ratio of children carrying a full set of new genes heterogeneously amounts to $(3/4)^{\nu}(2/3)^{\nu} = (1/2)^{\nu}$, while the ratio of children carrying them homologously is only $(1/4)^{\nu}$ and the remaining children carry some of new genes homologously and others heterogeneously. The number of such heterogeneous children amounts to $2^m/2^{\nu}$, when a pair of parents each exhibiting the new character produces 2^m children. The total number of such heterogeneous diploids becomes $(2^m/2^{\nu})^h$ after h generations, if the new style diploids continue to increase by 2^m times per generation. At this stage, the breading between the heterogeneous parents also produces the children lacking κ kinds of new genes with the amount of $({}_{\nu}C_{\kappa} 3^{\nu-k}/4^{\nu})(2^m/2^{\nu})^h$, according to Equation (18). The hybridization of these children with other latent variant carrying other k kinds of new genes on separate chromosomes yields the children, $1/2^{\nu}$ of which receives heterogeneously $(\nu-\kappa)$ kinds of the first lineage of new genes and κ kinds of other new genes. If this mosaic set of new genes exhibits the second stage of a new morphological character, the second stage of new diploids are yielded with the higher probability than the first new style diploids. When one example of ${}_{\nu}C_{\kappa}$ is focused, the following inequality relation gives such condition.

$$\frac{1}{2^{\nu}} P_{r:\kappa} \frac{3^{\nu-\kappa}}{4^{\nu}} \left(\frac{2^m}{2^{\nu}}\right)^h P_{r:\nu}^2 > P_{r:\nu} \tag{19}$$

This relation is rewritten to the following form.

$$(m-\nu)h > 3\nu - (\nu-\kappa)\log 3/\log 2 - (\log P_{r:\kappa} + \log P_{r:\nu})/\log 2 \tag{20}$$

To satisfy this inequality (20), the coefficient of h, m-ν, on the left side must be positive. This subsidiary condition indicates that the prolificacy is necessary for the first new style diploids to produce the second stage of new style diploids with higher probability. For example, more than $10^3 (2^m > 2^{10})$ children are necessary in the case of $\nu = 10$, although more than four $(2^m > 2^2)$ children are sufficient for the case of $\nu=2$. Under this subsidiary condition $m > \nu$, it will be then investigated how many generations are needed to realize the inequality (20) for a given set of ν and κ ($<\nu$). For this purpose, the values of the right hand side of inequality (20) are plotted against the values of ν for three cases of κ values in Figure 2. As seen in this figure, the value of $(m$-$\nu)h$ to satisfy the inequality (20)

becomes larger as the values of ν and κ are increased. In the case of $\nu=10$ and $\kappa=5$, for example, more than one hundred generations are needed to satisfy the inequality (20), when $(m-\nu)$ is nearly equal to one. Nevertheless, this period is short in the scale of geological time, and it is reasonable to say that the second stage of new style diploids appear soon after the appearance of the first new diploids. If different morphological characters are expressed depending on the mosaic sets of $(\nu-\kappa)$ and κ kinds of new genes, the second stage of new style diploids show the explosive divergence of body plans because the number of ${}_{\nu}C_{\kappa}$ becomes enormously large for large values of ν and κ.

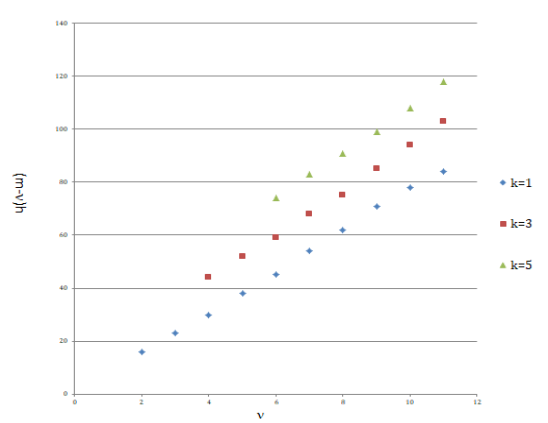

Figure 2: The condition for the second stage of new style diploids to be generated with the higher probability than the first new style diploids. The values of the right hand side of inequality (20) are plotted against the values of ν in the three cases of κ values. In this plotting, the values of $P_{r:\nu}/Q_{\nu}$ in Figure 1 are used for the values of $P_{r:\nu}$ and $P_{r:\kappa}$ by tentatively assuming Q_{ν} and Q_{κ} to be $10^{-2\nu}$, and $10^{-2\kappa}$, respectively. The inequality (20) holds in the region where the values of $(m-\nu)h$ (longitudinal coordinate) are larger than the values on the plotted curve. For example, more than *16* of $(m-\nu)h$ value is sufficient for the satisfaction of the inequality in the case of $\nu=2$ and $\kappa=1$, but the value of $(m-\nu)h$ amounts to more than *108* in the case of $\nu=10$ and $\kappa=5$.

Moreover, the next stage of divergence can be also induced on the way to establish the second stage of new style diploids homologously, if the latent variants carrying new genes are still present in the population. After the latent variants are decreased, the divergence is ceased until the new genes generated from gene duplication are accumulated in the survived lineages. This evolutionary pattern of diploid eukaryotes explains the punctuated mode of explosive divergence of body plans suggested from paleontology [19].

In the case when the new genes are concentrated on a smaller number of chromosomes, the degree of divergence is decreased to the simpler one. This is also seen from the above scheme if ν and κ are assigned to the smaller number of chromosomes carrying multiple kinds of new genes and a_J^2 and $a_J b_J$ denote the homologous and heterogeneous states concerning multiple kinds of new genes, respectively.

The above result is essentially the same for the dioecism, although the distinction between male and female types makes the mathematical description somewhat complicated.

Physiological Problems in the Innovation to Multicellular Diploids and Divergence Pattern of Organisms to Resolve These Problems

The new genes accumulated by the mechanisms described in the second and third sections are used for resolving the following physiological problems as well as for advancing the cell differentiation to the higher hierarchy; (I) the material and energy have to be supplied to the development of cell differentiation until the cooperative action of differentiated cells begins to acquire them from the outside and the amount of supplied material and energy becomes larger in the diploid state than in the monoploid state [17,18] and (II) the duration time of the genome expanded for the cell differentiation has to be elongated to acquire the material and energy from the outside during the longer time.

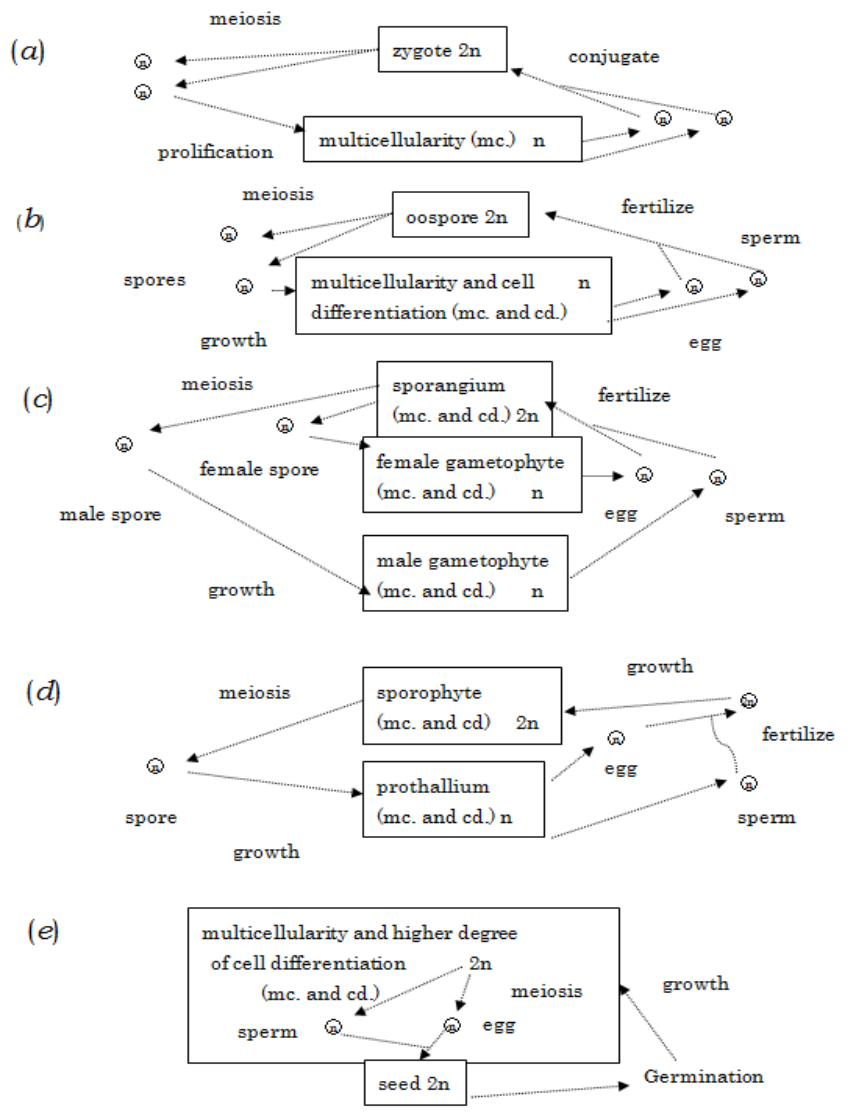

Figure 3: The representative examples of green plants showing the intermediate stages in the innovation from monoploids to diploids along the advancement in cell differentiation. (*a*) The *Conjugatae* such as *Roya* and *Spirogyra* repeats a cycle of unicellular monoploids, multicellular monoploids, conjugation of two monoploid cells to form a zygote and returning to the unicellular monoploids by meiosis. (*b*) The *Chara* of *Charophyta* develops the multicellularity and cell differentiation in the monoploid state. This adult form produces both eggs and sperms, whose fertilization yields the oospore. By the meiosis, the oospore produces the spores, each of which grows to the adult form. (*c*) In the *Bryophyta*, the monoploid generation differentiates into female and male types of gametophytes, and the fertilized egg on the female gametophyte grows to the sporangium which produces both female and male types of spores. (*d*) The *Pterophyta* develops the cell differentiation in the diploid generation called the sporophyte to acquire the energy and materials by itself and the monoploid generation called the *prothallium* is specialized to produce only eggs and sperms. In some species of *Pterophyta*, the *prothallium* also differentiates into female and male types. (*e*) The seed plants, in which the monoploid generation (egg and sperm) is produced in the reproductive organ incorporated into the diploid body, take the energy and materials solely by the cooperative action of differentiated cells in the diploid state and develop the albumen to supply the energy and materials with the seed for its germination.

Abbreviations: mc; multicellularity; cd: cell differentiation; n: monoploid state; 2n: diploid state

The green plants illustrate the evolutionary steps to resolve these physiological problems, as shown in Figure 3*(a)*~*(e)*. First, the cell differentiation occurs in the monoploid state through the meiosis in zygote, then the monoploid generation differentiates into female and male types, and the female parent endows the eggs with the material and energy necessary for the development of cell differentiation in the diploid state under the evolution of mitosis. The cell differentiation in the diploid generation succeeds in evolving the vascular bundles as well as leaves and roots, and further evolves the reproductive organ which not only produces the eggs and sperms but also supplies the fertilized eggs with material and energy source for their growth to seeds.

It is easily ascertained that the diploid state elongates the duration time of differentiated cells against the nucleotide base changes. The base changes are considered to occur by the mis-repair of damaged bases in DNAs [10], and they must occur at all sites including functionally important ones in the genome. When the base change rate is denoted by u, the diploid eukaryote consisting of Z cells retains the following number of cells not suffering base change at any pairs of homologous s sites in their genome during the time t; $Z\{1-(ut)^2\}^s \approx Z\{1-s(ut)^2\}$. This number is calculated to be $Z(1-10^{-5})$ even after one hundred years for the genome size $s = 10^9 bp$, using $u=10^{-9}$per site per year. In the multicellular monoploids, on the contrary, the number of cells that do not suffer the base change is calculated to be $Z(1-ut)^s \approx Z(1-sut)$ after the time t. If the genome size s of a monoploid eukaryote were$10^8 bp$, the above number would be reduced to $Z(1-10^{-1})$ after one year. In fact, the genome size s is expanded to the order of $10^8 bp$ in *Arabidopsis* and *Drosophila*, $3x10^9 bp$ in *Homo sapiens* and further to $10^{10} bp$ in *Taxodials*, while *Saccharomyces* only carries the genome of $s = 10^7 bp$ [20].

According to the analyses on neutral amino acid replacement [21] and nucleotide base substitutions [8], the divergence of green plants and animals occurred $1.2x10^9$ years ago, the divergence of stages *(b)* and *(c)* of green plants occurred before 10^9 years ago, and the divergence of stages *(c)*, *(d)* and *(e)* of green plants successively occurred after $5x10^8$years ago [22]. Although the early stages of animals are hardly found at the present time, the *Cnidaria* alternates the monoploid generation differentiated into the female and male types with the asexual diploid generation. The hermaphroditism is common to the lower diploid animals such as *Pulmonata* and *Oligochaeta*, but the dioecism becomes prevailing in the higher diploid animals, in contrast to the monoecism in most diploid green plants. Such difference in sexual differentiation may be due to the difference in living style between predators and autotrophs. The seed plants and animals having realized the diploid state show the explosive divergence of morphological characters as the genetical mechanism formulated in the preceding section. In particular, the animals show the punctuated mode of explosive divergence of *Mullusca*, *Annelida*, *Arthropoda*, *Echinodermata* and *Chordata* that is first found by fossil records [23-26] and then ascertained to have occurred during the period of $8x10^8$~$5x10^8$ years ago by the analysis on base-pair changes in mitochondrial rRNAs [27]. The comparison of genome between these phyla is expected to identify the mosaic set of genes responsible for their divergence. Such mosaic sets of genes could be also found between lower taxonomical categories of multicellular eukaryotes, e. g., between different classes in each phylum.

Conclusions and Discussion

Darwinian evolution under the nucleotide base changes is the fundamental process to maintain the negative entropy of an organism, resolving the paradox of Maxwell's demon [18,28], but the generation of new species only by this evolution necessitates the geographical isolation and/or climate change. This is also the case for multicellular diploid eukaryotes, although this evolution becomes slower than that of monoploids by the process of fixing selectively advantageous bases homologously. On the contrary, the generation of new genes from gene duplication extends the range of negative entropy (systematization) to yield the divergence of new and old styles of organisms [17,18,28]. In particular, the multicellularity and cell differentiation of eukaryotes after the acquirement of the mitochondria is drastic. It starts from the accumulation of many kinds of new genes generated from gene duplication in the monoploid eukaryote by the exchange of homologous chromosomes through conjugation and advances to the diploid state through the intermediate stages of alternating the monoploid generation with the diploid one. The new genes generated from gene duplication are gotten together in the intermediate stages by the superposition of the schemes formulated in the second and third sections. These new genes have caused the differentiation of gametes into female type (egg) and male type (sperm) as well as the differentiation of somatic cells. The material and energy source endowed with the egg is indispensable for the development of cell differentiation in the next generation especially of diploid state. The differentiation of gametes into female and male types is succeeded in the diploid organisms, but the sexual differentiation of diploid bodies is different between green plants and animals. While the diploid bodies of green plants are mostly

monoecism, the diploid bodies of higher animals are differentiated into female and male types. In particular, the females of mammals evolve the corpus mammae to supply the baby with milk and the females of the *Eutheria* further evolve the placenta for the growth of fertilized egg to the embryo and fetus. This evolution of sexual differentiation is closely related with the development of higher hierarchy of somatic cell differentiation especially in brain.

Multicellularity and cell differentiation are also recognized in fungi and sea algae. In the *Myxomycota*, the unicellular monoploid forms usually self-reproduce individually, but they conjugate and the zygotes further aggregate to form an apocyte called the *plasmodium* under the dried condition of environment. However, the cells in polyploid state are more difficult to overcome the physiological problem (I) than in the diploid state. In the *Basidiomycota*, the monoploid forms called the *hypha* conjugate, and the zygotes aggregate and grow to a fruit body (mushroom) in which the spores are produced by the meiosis. However, this multicellular form only scatters the mature spores far way. On contrary, the sea algae, which have acquired the rhodoplasts as the endosymbionts of cyanobacteria independently of green plants [10,29], have advanced the cell differentiation. Some of them such as *Laminaria* reaches the stage corresponding to *(d)* of green plants. These facts imply that the initiation of cell differentiation is aided by the material and energy supplied from photosynthetic plastids as well as from mitochondria. This might be also the case for the ancestor of animals. In addition to the close relation of phylogeny with the green plants [8,10], the lower animals such as *Prorifera* and *Cnidaria* still carry symbiotic algae.

References

1. Otsuka J (2017) A mathematical formulation of evolution and innovation I. Unicellular organisms. Phys Sci & Biophys Journal 1(1): 1-10.

2. Darwin C (1859) The origin of species. John Murry, London.

3. Fisher RA (1930) The general theory of natural selection. Dover Publications, Oxford University Press, London, New York.

4. Wright S (1949) Adaptation and selection. *In:* JepsonGL, SimpsonGG, Mayer E (Eds.), Genetics, paelontology and evolution. Princeton University Press, Princeton, New Jersey, pp: 365-389.

5. Kimura M (1968) Evolutionary rate at the molecular level. Nature 217: 624-626.

6. King JL, Jukes TH (1969) Non-Darwinian evolution. Science 164(3881): 788-798.

7. Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2): 111-120.

8. Otsuka J, Nakano T, Terai G (1997) 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 184(2): 171-186.

9. Woese CR (1987) Bacterial evolution. Microbiol Rev 51(2): 221-271.

10. Otsuka J, Terai G, Nakano T (1999) Phylogeny of organisms investigated by the base-pair changes in the stem regions of small and large ribosomal subunit RNAs. J Mol Evol 48(2): 218-235.

11. Ingram VM (1963) The haemoglobin in genetics and evolution. Columbia Press, New York, pp: 165.

12. Ohno S (1970) Evolution by gene duplication.Spring-Verlag, Berlin.

13. Gilbert W (1978) Why genes in pieces? Nature 271: 501.

14. Ferris SD, Whitt GS (1979) Evolution of the differential regulation of duplicated gene after polyploidization. J Mol Evol 12(4): 267-317.

15. Kojima S, Otsuka J (2000) Characterization of organisms by paralogous relationships of proteins. Part III. *Saccharomyces cerevisiae*. ResCommun Biochem Cell Mol Biol 4: 101-138.

16. Kojima S, Otsuka J (2002) Characterization of proteome by similarity linkages of paralogous functional domains and special amino acid-rich regions. Part IV. *Drosophila melanogaster*. Res Commun Biochem Cell Mol Biol 6: 72-102.

17. Otsuka J (2008) A theoretical approach to the large-scale evolution of multicellularity and cell differentiation. J Theor Biol 255(1): 129-136.

18. Otsuka J (2017) The concept of biological activity and its application to biological phenomena. J Phy Chem & Biophys 7: 235-241.

19. Eldredge N, Gould SJ (1972) Punctuated equilibria: an alternative to phyletic gradualism. *In:* Schopf TJM (Ed.), Models in paleobiology. San Francisco, Freeman and Cooper, pp: 82-115.

20. Wheeler DL, Church DM, Edger R, Federhen S, Helmberg W, et al. (2004) Database resources of the National Center for Biotechnology Information: update. Nucleic Acids Res 32 (Database issue): D35-D40.

21. Dickerson RE (1971) The structure of cytochrome *c* and the rate of molecular evolution. J Mol Evol 1(1): 26-45.

22. Kawai Y, Otsuka J (2004) The deep phylogeny of land plants inferred from a full analysis of nucleotide base changes in terms of mutation and selection. J Mol Evol 58(4): 479-489.

23. Matthews SC, Missarzhevsky V (1975) Small shelly fossils of late precambrian and early cambrian age: a review of recent work. J Geol Soc 131(3): 289-303.

24. Gould SJ (1989) Wonderful Life. The burgess shale and the nature of history. WW Norton & Company Inc., New York.

25. Rozanov AY, Zhuravlev AY (1992) The lower cambrian fossil record of the Soviet Union. In: Lipps JH, Signor PW (Eds.), Origin and early evolution of the metazoa. Plenum Press, New York, London, pp: 205-282.

26. Rasmussen B, Bengston S, Fletcher IR, McNaughton NJ (2002) Discoidal impressions and trace-like fossils more than 1200 million years old. Science 296(5570): 1112-1115.

27. Otsuka J, Sugaya N (2003) Advanced formulation of base pair changes in the stem regions of ribosomal RNAs; its application to mitochondrial rRNAs for resolving the phylogeny of animals. J Theor Biol 222(4): 447-460.

28. Otsuka J (2018) The negative entropy in organisms; its maintenance and extension. J Mod Phys 9(12): 2156-2169.

29. Van den Eynde H, De Roeck E, Van de Peer Y, Vandenberghe A, Willekens P, et al. (1988) 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 27(2): 126-132.