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

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

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

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

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

1

A Theoretical Scheme of the Large-Scale Evolution by Generating New Genes from Gene Duplication

Jinya Otsuka
JO Institute of Biophysics, Tokyo, Japan

1. Introduction

In the famous book "The Origin of Species" by Darwin (1859), the gradual accumulation of selectively advantageous variants has been proposed qualitatively by obtaining a hint from the artificial selection of domestic animals and plants as well as from the observation of unique species in a geographically isolated region. The core of this proposal has become evident, after the re-discovery of Mendelian heredity, by the detection of hereditary variants, i. e., mutants, and extensive investigations have been carried out for the behavior of mutants especially in the *Drosophila* population (Dobzhansky, 1941; Mayer, 1942; Huxley, 1943; Simpson, 1944). In parallel, Darwinian evolution is mathematically formulated in population genetics to estimate the probability that a spontaneously generated mutant is fixed in, or eliminated from, the population according to the positive or negative value of a selective parameter (Fisher, 1930; Wright, 1949). Although the accumulation of such mutants as those found in the *Drosophila* was supposed to explain the whole process of evolution, the mutants detected at that time were mainly due to the point mutations in established genes, and most of them were defective. Thus, doubts remain about whether the gradual accumulation of such mutants gives rise to radically new organs such as wings and eyes. Another criticism against the survival of the fittest in Darwinian evolution is also raised by the ecological fact of diversity that different styles of organisms coexist in the same area (Nowak et al., 1994).

The gene and genome sequencing, which started in the latter half of the last century, has brought new information about the evolution of organisms. First, the amino acid sequence similarities of paralogous proteins strongly suggest that the repertoire of protein functions has been expanded by gene duplication, succeeding nucleotide base substitutions, partial insertion and deletion, and further by domain shuffling in some cases (Ingram, 1963; Gilbert, 1978; Ferris & White, 1979). Such examples are now increasing, proposing many protein families and superfamilies. Second, the clustering analysis of proteomes reveals a characteristic feature that the proteins functioning in the core part are essentially common to both prokaryotes and eukaryotes, and that the decisive difference in gene repertoire between the organisms is observed in the peripheral parts displaying different living styles (Kojima & Otsuka, 2000 a, b, c; Kojima & Otsuka, 2002). These sequence data are now compiled into databases (e. g., Wheeler et al., 2004; Birney et al., 2006).

Although the importance of gene duplication in evolution was already indicated in the last century (Ohno, 1970), this indication still remained describing the circumstantial evidence of gene duplication and the fossil record of vertebrate organs in a qualitative way. Theoretically, some new concept is needed to formulate the evolution by gene duplication, going beyond the narrow view of population genetics which only focuses on a mutated gene. For this purpose, the author has recently proposed the new concept of 'biological activity', which is determined by a whole genome, and explained the divergence of the original style of organisms and the new style of organisms having a new gene generated from the counterpart of duplicated genes (Otsuka, 2005; 2008). This evolution by gene duplication will be called the large-scale evolution, being distinguished from Darwinian evolution.

In this chapter, the explanatory remarks are first given for the concept of 'biological activity' and the large-scale of evolution will be then investigated in detail on the three types of organisms, which are different in their genome constitution and transmission. The genome is a single DNA molecule in most prokaryotes and it is a set of chromosomes in lower eukaryotes. These organisms will be tentatively called the monoploid organisms as the first type of organisms. Some lower eukaryotes exchange homologous chromosomes through the process of conjugation. These lower eukaryotes are treated as the second type of organisms. In higher animals and plants, each of the cells constituting the adult form carries the genome consisting of the plural number of homologous chromosome pairs, and the monoploid state only appears in the gametes (egg and sperm). These higher eukaryotes will be treated as the third type, being called the diploid organisms in the sense that the present study focuses on the evolution of the characters expressed in their diploid state. The main purpose of the present study is to elucidate the difference between the three types of organisms, especially in the probabilities that two or more kinds of new genes are generated from different origins of gene duplication. This study reveals that the second type organism is most suitable to generate many kinds of new genes and the third type organism is next in line. The cell differentiation is a representative character, which requires many kinds of genes for its expression, and the present result provides an explanation for the fact that the cell differentiation has started in the second type of organisms and then evolved to the higher hierarchy in the third type of organisms.

2. The concept of biological activity

Although the 'biological activity' is a macroscopic quantity generally characterizing various biological systems such as an ecological system, an organism, an individual cell of a multicelluar organism etc. (Otsuka, 2004, 2005, 2008), it will be explained focusing on an organism for the present purpose of considering the large-scale evolution of organisms by gene duplication. In general, an organism may be characterized by a set of two macro-variables, the genome size N and its systematization - S_N of genes and their products. The systematization corresponds to the negentropy, which should be measured for the specific arrangement of nucleotides in individual genes, the degree of accuracy in transmitting the genetic information to the amino acid sequences of proteins, the formation of metabolic pathways by enzyme protein functions, the regulation and control at various levels of biological processes, the cell structure constructed by the interaction of metabolic products, and for furthering the communication between differentiated cells in the case of multicellular organisms. The energy acquired by an organism depends not only on the

genome size N and systematization - S_N but also on the material and energy source M available from the environment. Thus, the energy acquired by the organism during its lifetime is expressed as $E_a(M; N, S_N)$, which may be an increasing function of N and S_N as well as of M. On the other hand, the organism utilizes the acquired energy and materials to construct the biomolecules for its growth and self-reproduction. The energy $E_s(N, S_N)$ stored in the form of biomolecules is also another increasing function of N and S_N. The difference between the acquired energy and the stored energy, $E_a(M; N, S_N) - E_s(N, S_N)$, is lost as heat. According to the second law of thermodynamics, the entropy production by the heat must compensate for the entropy reduction, i. e., - S_N, by the systematization. Thus, the following inequality must hold:

$$E_a(M;N,S_N) - E_s(N,S_N) - TS_N > 0 \tag{1}$$

where T is the temperature. In other words, this indicates the upper boundary of systematization (negentropy) by entropy production (Otsuka & Nozawa, 1998). However, organisms must have developed the systematization to increase the acquired energy through the evolutionary process of gene and genome duplication, nucleotide base substitutions and selection, and this is the main problem in the present study. The larger value of $E_a(M; N, S_N) - E_s(N, S_N) - TS_N$ gives a measure for the biological processes to proceed more smoothly. In this sense, the quantity of $E_a(M; N, S_N) - E_s(N, S_N) - TS_N$, which an organism produces during one generation, will be called the 'biological activity' of the organism. The 'biological activity' has thermodynamic connotation as a departure from equilibrium, but this is in a reverse relation to the free energy in thermodynamics, which decreases upon any change in a given system by the decrease in internal energy and/or by the increase in entropy. In an organism, the acquired energy is stored in ATP and NADH molecules as chemical energy, and it is gradually consumed in the syntheses of biomolecules under the guidance of the enzymes, without drastically raising the temperature. In such moderate reactions, the temperature is almost constant, and the quantity obtained from the 'biological activity' divided by the product of the Boltzmann constant k and temperature T is considered to be approximately proportional to the self-reproducing rate of an organism, which will be denoted by $R(M; N, S_N)$ hereafter.

This concept of 'biological activity' or self-reproducing rate is useful to formulate the large-scale of evolution arising from the gene duplication and succeeding generation of new genes. The essence of the present theory considers the following process of evolution in terms of 'biological activity'. First, the enlarged genome size $N + \Delta N$ due to gene duplication makes the stored energy $E_s(N+\Delta N, S_N)$ larger than $E_s(N, S_N)$, while the acquired energy $E_a(M; N+\Delta N, S_N)$ remains almost equal to $E_a(M; N, S_N)$. Thus, the 'biological activity' of a variant bearing duplicated genes becomes lower than that of the original style organism. Moreover, the biological activity of the variant further decreases by the increase in systematization from S_N to $S_{N+\Delta N}$, as a new gene generated from the counterpart of duplicated genes is incorporated into an extended system of regulation and control. However, such a variant with the lower activity is not necessarily extinct but has a chance to recover as a new style of organisms, if the new gene begins expressing a new biological function to raise the acquired energy from $E_a(M, N+\Delta N, S_N)$ to $E_a(M', N+\Delta N, S_{N+\Delta N})$ by utilizing the new material and energy source M' other than M, or by moving to a new living area or by utilizing M more efficiently in the case of $M' = M$. This process of the large-scale evolution will be mathematically formulated to estimate the probabilities of generating new genes, for the

first type of organisms in section 3, for the second type in section 4 and for the third type in section 5.

3. Prokaryotes and lower eukaryotes in the monoploid state

For the mathematical description, the set of variables (N_i, S_{Ni}) characterizing a variant i will be simply denoted as a single variable x_i, unless the description of changes in its content is necessary. In the population of monoploid organisms taking a common material and energy source M, the number $n(x_i;t)$ of variants, each characterized by the monoploid genome x_i, obeys the following time-change equation.

$$\frac{d}{dt}n(x_i;t) = \{Q_{xi}(t)R(M;x_i) - D(x_i)\}n(x_i;t) + \sum_{j(\neq i)} q_{xi,xj}(t)R(M;x_j)n(x_j;t) \tag{2}$$

where the self-reproducing rate and death rate of the variant x_i are denoted by $R(M;x_i)$ and $D(x_i)$, respectively. The apparent decrease factor $Q_{xi}(t)$ in the self-reproducing rate of the variant x_i is related with the mutation term $q_{xj,xi}(t)$ from the variant x_i to other kinds of variants x_j 's in the following way.

$$Q_{xi}(t) = 1 - \sum_{j(\neq i)} q_{xj,xi}(t) \tag{3}$$

If the quantity $q_{xi,xi}(t)$ defined by $Q_{xi}(t)$ *-1* is introduced, the restriction $j \neq i$ can be removed from the summation of the second term on the right side of Eq. (2). For investigating the population behavior, Eq. (2) is transformed into the following two types of equations; one concerning the total number of all kinds of variants defined by $B(t) = \Sigma_i\, n(x_i;t)$ and another concerning the fraction $f(x_i;t)$ of variants x_i defined by $n(x_i;t)/B(t)$.

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

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

where the increase rate $W(M; x_i)$ of variant x_i and the average increase rate $W_{av}(M;t)$ of organisms in the population are defined by the following forms, respectively.

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

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

Strictly, the nucleotide base change occurs due to the miss in repairing damaged bases, while the gene duplication occurs by the illegitimate crossing over of DNA strands upon replication. Although they are simply represented by the mutation term $q_{xi,xj}(t)$ in the above mathematical formulation, the point mutation due to nucleotide base change and the gene duplication are distinguished from each other in the following mathematical treatment.

Darwinian evolution corresponds to the evaluation of the time-change of variant fractions mainly by the first term on the right side of Eq. (5), as discussed by Eigen (1971). If the increase rate $W(M;x_i)$ of an occasionally generated mutant x_i is greater than the average increase rate, that is, $W(M;x_i) - W_{av}(M;t) > 0$, the fraction $f(x_i;t)$ increases with time according to the first term on the right side 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), although this increase is ultimately stopped by the decrease in available material M. On the other hand, the fraction $f(x_i;t)$ decreases when $W(M;x_i) - W_{av}(M;t) < 0$. 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, each characterized by x_{opt}. However, such Darwinian evolution may only hold for the point mutations in existing genes.

The large-scale evolutionary process of generating new gene(s) from gene duplication is obtained by evaluating the fraction of variants up to the first and higher orders of the mutation term. For this illustration, Eq. (5) will be formally integrated with respect to time t:

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

After the organisms x_{opt} have become dominant in the population, $W_{av}(M;t)$ is approximately equal to $W(M; x_{opt})$, the fractions of variants except for x_{opt} are neglected on the right side of Eq. (8), and the mutation term $q_{xi,xopt}(t)$ is replaced by the mutation rate $q_{xi,xopt}$ defined as an average of mutation terms during a sufficiently long time t, i. e.,

$$q_{xi,xopt} \equiv \frac{1}{t}\int_0^t q_{xi,xopt}(\tau)d\tau \tag{9}$$

Then, the fraction $f(x_i)$ of variants x_i is finally related with the fraction $f(x_{opt})$ of dominant organisms x_{opt} in the following form.

$$f(x_i) = \frac{q_{xi,xopt}R(M;x_{opt})}{W(M;x_{opt}) - W(M;x_i)}f(x_{opt}) \tag{10}$$

Among such satellite variants, the variant arising from the gene duplication is especially notable in the sense that it has the potential to generate a new gene from the counterpart of duplicated genes. If the probability of generating a new gene I from the duplicated part in x_i is denoted by $q_{xI,xi}$, a new style of the organism carrying the new gene I is generated from the original style of an organism with the following probability $P_{m1}(x_I \leftarrow x_i \leftarrow x_o)$.

$$P_{m1}(x_I \leftarrow x_i \leftarrow x_o) = \frac{q_{xI,xi}q_{xi,xo}R(M;x_o)}{W(M;x_o) - W(M;x_i)} \tag{11}$$

where x_{opt} is rewritten into x_o with the meaning of the original style of an organism. Here, x_i and x_I correspond to $(N+\Delta N, S_N)$ and $(N+\Delta N, S_{N+\Delta N})$, respectively, in terms of the set of variables characterizing an organism in section 2.

When a biologically meaningful character is newly exhibited by two new genes generated from different origins of gene duplication, the variant, which experienced gene duplication *i*, must successively experience further gene duplication *j* in the other part of the genome to exhibit such a new character. The fraction $f(x_{ij}; t)$ of such variants x_{ij} obeys the following equation as a special case of Eq. (5).

$$\frac{d}{dt}f(x_{ij};t)=\{W(M;x_{ij})-W_{av}(M;t)\}f(x_{ij};t)+q_{xij,xi}(t)R(M;x_i)f(x_i;t) \tag{12}$$

where $q_{xij,xi}(t)$ represents the mutation term from the variant x_i to the variant x_{ij} and the smaller terms including the mutation from the variant x_{ij} to other variants are neglected. By formally integrating Eq. (12), the fraction $f(x_{ij})$ of variants x_{ij} is finally expressed as

$$f(x_{ij})=\frac{q_{xij,xi}R(M;x_i)}{W(M;x_{opt})-W(M;x_{ij})}f(x_i) \tag{13}$$

where $W_{av}(M;t)$ is approximated to be $W(M;x_{opt})$ and the mutation term $q_{xij,xi}(t)$ is replaced by the mutation rate $q_{xij,xi}$, i. e.,

$$q_{xij,xi}\equiv\frac{1}{t}\int_0^t q_{xij,xi}(\tau)d\tau \tag{14}$$

By inserting the expression (10) of fraction $f(x_i)$ into the right side of Eq. (13), the fraction $f(x_{ij})$ of variants x_{ij} is related with the fraction $f(x_{opt})$ of dominant organisms x_{opt} by the second order of mutation rates in the following form.

$$f(x_{ij})=\frac{q_{xij,xi}R(M;x_i)}{W(M;x_{opt})-W(M;x_{ij})}\cdot\frac{q_{xi,xopt}R(M;x_{opt})}{W(M;x_{opt})-W(M;x_i)}f(x_{opt}) \tag{15}$$

Thus, a new style of the organism x_{IJ} carrying new genes *I* and *J* is generated from the original style of an organism x_o with the following probability $P_{m2}(x_{IJ} \leftarrow x_{ij} \leftarrow x_o)$.

$$P_{m2}(x_{IJ}\leftarrow x_{ij}\leftarrow x_o)=\frac{q_{xIJ,xIj}q_{xI,xi}q_{xij,xi}R(M;x_i)}{W(M;x_o)-W(M;x_{ij})}\cdot\frac{q_{xi,xo}R(M;x_o)}{W(M;x_o)-W(M;x_i)} \tag{16}$$

where $q_{xIJ,xIj}$ is the probability of generating the new gene *J* from the duplicated part in *j*. This procedure can be easily extended to the general case of successively generating three or more new genes.

Before describing the result of the general case, the expression of probabilities (11) and (16) will be simplified by assuming that the gene duplication only reduces the self-reproducing rate of the variant without any influence on the death rate. When the self-reproducing rate of the original style organism is simply denoted by *R* and that of the variant x_i is expressed as $R(1-s_1)$ with the reduction factor satisfying $0 < s_1 < 1$, the probability (11) is simply expressed as

$$P_{m1}(x_I\leftarrow x_i\leftarrow x_o)=\frac{Q_1}{s_1} \tag{17}$$

where $q_{xI,xi}q_{xi,xo}$ is denoted by Q_1. In the same way, the self-reproducing rate of the variant x_{ij} is denoted as $R(1- s_1- s_2)$ with the additional reduction factor s_2 under the condition of $0 < s_1 + s_2 < 1$ and $q_{xIJ,xIj}q_{xI,xi}q_{xij,xi}q_{xi,xo}$ is denoted by Q_2. The expression of the probability (16) then becomes

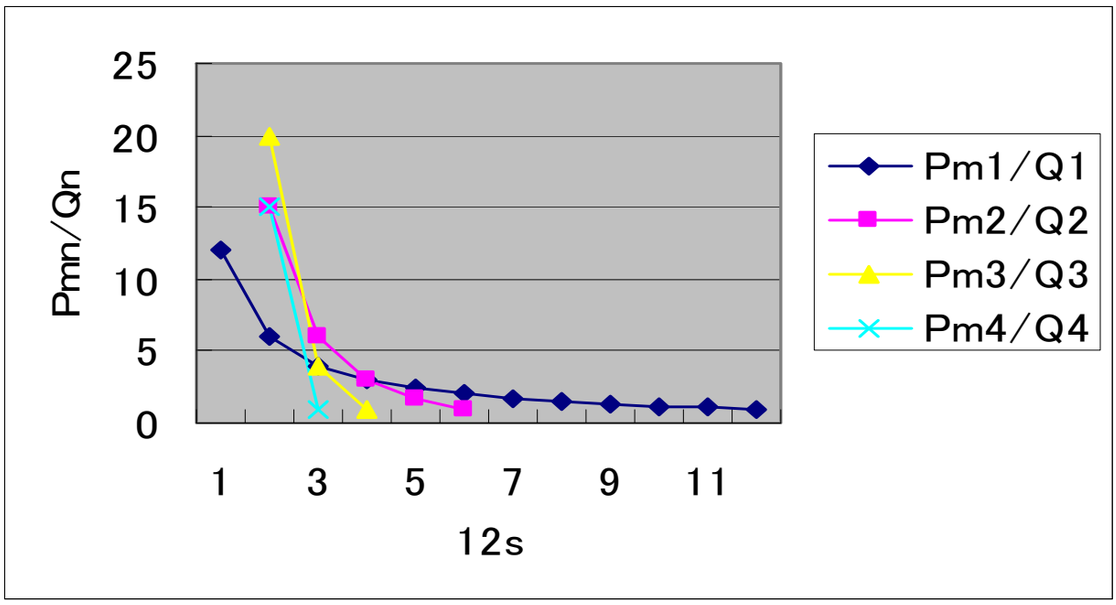


Fig. 1. The probabilities of generating new genes from gene duplication in the monoploid organism. On the basis of Eq. (20), the values of P_{mn}/Q_n are plotted against the twelve-fold reduction factor $12s$ for $n = 1, 2, 3$ and 4. Although the value of Q_n becomes smaller for a larger value of n, the plotting of the probability P_{mn} in the unit of Q_n makes the figure compact. The probability P_{m1} is present in a whole range of reduction factor $0 < s < 1$. As the number of n increases, however, the range of reduction factor s, where the probability P_{mn} is present, is narrowed to $0 < s < 1/n$.

$$P_{m2}(x_{IJ} \leftarrow x_{ij} \leftarrow x_o) = \frac{(1-s_1)}{s_1(s_1+s_2)}Q_2 \tag{18}$$

This expression of probabilities (17) and (18) is easily extended to express the probability of successively generating n kinds of new genes in the following way.

$$P_{mn} = \frac{(1-s_1)(1-s_1-s_2)\cdots\cdots(1-s_1-s_2-s_3-\cdots\cdots-s_{n-1})}{s_1(s_1+s_2)\cdots\cdots(s_1+s_2+s_3+\cdots\cdots+s_n)}Q_n \tag{19}$$

The reduction factors s_i's in Eq. (19) are in the relations of $0 < s_1 + s_2 +\cdots\cdots\cdots+ s_n < 1$ and $0 < s_1, s_2, \cdots\cdots\cdots, s_n < 1$. Strictly, the values of s_i's are different depending on the length of duplicated sequences and on the order of gene duplication events. For the simple investigation of the n dependence of P_{mn}, however, these reduction factors are assumed to be commonly equal to one variable s. Then, the first relation becomes $0 < s < 1/n$, and Eq. (19) is reduced to

$$P_{mn} = \frac{(1-s)(1-2s)(1-3s)\cdots\cdots\cdots\{1-(n-1)s\}}{n!s^n}Q_n \tag{20}$$

On the basis of this expression (20), the probabilities P_{mn}'s for several values of n are plotted against the reduction factor s in Fig. 1. In the case of $n = 1$, the reduction factor s is permitted in a whole range of $0 < s < 1$ and the probability P_{m1} of generating a new gene is present in this range. This means that the monoploid organism is suitable to create a new gene step by step, testing the biological function of the new gene product, even if the gene size is large. As the value of n increases, however, the reduction factor s is restricted to the narrower range of $0 < s < 1/n$. When the monoploid organism creates simultaneously multiple kinds of new genes from different origins of gene duplication, therefore, these genes are obliged to be of a smaller size. Moreover, the probability P_{mn} is also decreased as the value of n increases. This is because Q_n becomes smaller for the larger value of n. Thus, it is difficult for the monoploid organism to evolve a new character which requires the expression of many kinds of new and large genes. This result is common to the prokaryote with a single DNA molecule and the lower eukaryote with the plural number of chromosomes, if the latter does not conjugate to exchange homologous chromosomes.

4. The monoploid eukaryotes that exchange homologous chromosomes through conjugation

Some monoploid eukaryotes with the plural number of chromosomes conjugate to form a zygote during their life cycle, and the zygote produces monoploid descendants by exchanging homologous chromosomes upon the meiosis. Although the conjugation also occurs in prokaryotes, it only takes place to exchange plasmids and partial genes. Originally, the conjugation would have evolved to avoid the accumulation of disadvantageous mutations in a special lineage and to maintain the stability of a population by weakening the influence of such mutations. However, the conjugation in the eukaryote with the plural number of chromosomes makes it possible to produce the descendant receiving two or more new genes, even if these new genes are relatively large. Thus, the conjugation of such eukaryotes is considered to be the strategy to overcome the difficulty of generating many and large new genes from the successive gene duplication in a single lineage of monoploid organisms. For this illustration, several examples will be first listed in the following subsections 4.1 to 4.3, and they are used to estimate the probabilities of producing the descendant received more new genes by the conjugation of variants, each carrying a smaller number of new genes.

4.1 The probability of producing the descendant received two new genes

Such a descendant is produced from the conjugation of two types of variants, one carrying a new gene I on a chromosome C_1 and another carrying a new gene J on another kind of chromosome C_2. The genome of the variant carrying the new gene I is denoted by (C_{1I}, C_{20}) and the genome of another variant carrying the new gene J is by (C_{10}, C_{2J}). The conjugation of these two types of variants yields the zygote, whose genome constitution is represented by $(C_{1I}, C_{10}; C_{2J}, C_{20})$. If the homologous chromosomes are randomly partitioned into two daughter cells, the probability P_{c2} of producing the new monoploid descendant received the genome (C_{1I}, C_{2J}) is calculated to be $P_{m1}^2/2$.

4.2 The probability of producing the descendant received three new genes

The descendant received three new genes I, J and K can be produced from the conjugation of variants, one carrying one new gene I and another carrying two new genes J and K. Two cases are considerable for this production.

One is the case that the new gene *I* is encoded on the chromosome C_1 and both new genes *J* and *K* are encoded on another kind of chromosome C_2. Then, the genome of the variant carrying the new gene *I* is denoted by (C_{1I}, C_{20}) and the genome of another variant carrying the new genes *J* and *K* is denoted by (C_{10}, C_{2JK}). The conjugation of these two variants forms the zygote $(C_{1I}, C_{10}; C_{2JK}, C_{20})$, which can produce four types of monoploid descendants, (C_{1I}, C_{2JK}), (C_{1I}, C_{20}), (C_{10}, C_{2JK}) and (C_{10}, C_{20}). If the homologous chromosomes are equivalently partitioned into two daughter cells, regardless of carrying new genes or not, the new monoploid descendant (C_{1I}, C_{2JK}) is produced with the probability of $P_{m1}P_{m2}/2$.
In the second case, the new genes *J* and *K* are encoded on separate chromosomes. If the chromosome carrying the new gene *K* is denoted by C_3, the genome of the variant carrying new genes *J* and *K* is represented by (C_{10}, C_{2J}, C_{3K}). The conjugation of this variant and the variant (C_{1I}, C_{20}, C_{30}) forms the zygote $(C_{1I}, C_{10}; C_{2J}, C_{20}; C_{3K}, C_{30})$. Under the random partition of homologous chromosomes, this zygote yields a new monoploid descendant (C_{1I}, C_{2J}, C_{3K}) with the probability of $P_{m1}P_{m2}/4$.
As a whole, $3P_{m1}P_{m2}/4$ is obtained for the probability P_{c3} of producing a new monoploid organism received three new genes by conjugation.

4.3 The probability of producing the descendant received four new genes

The highest probability of producing the descendant received four new genes is obtained by the conjugation of two variants, one carrying two new genes *I* and *J*, and another carrying other two new genes *K* and *L*. The following three cases (i) ~ (iii) are considerable. (i) The new genes *I* and *J* are encoded on the chromosome C_1 in one variant, while the new genes *K* and *L* are encoded on the chromosome C_2 in another variant. The conjugation of these two variants forms the zygote $(C_{1IJ}, C_{10}; C_{2KL}, C_{20})$, which yields four types of monoploid descendants, (C_{1IJ}, C_{2KL}), (C_{1IJ}, C_{20}), (C_{10}, C_{2KL}) and (C_{10}, C_{20}). If the homologous chromosomes are randomly partitioned into two descendants, the probability of producing the monoploid descendant (C_{1IJ}, C_{2KL}) is calculated to be $P_{m2}^2/2$. (ii) The new genes *I* and *J* are encoded on the chromosome C_1 in one variant but the new genes *K* and *L* are encoded on the chromosomes C_2 and C_3, respectively, in another variant. The conjugation of these two variants forms the zygote $(C_{1IJ}, C_{10}; C_{2K}, C_{20}; C_{3L}, C_{30})$. If the homologous chromosomes in each kind of *1*, *2* and *3* are randomly partitioned into two daughter cells, the probability of producing the monoploid descendant $(C_{1IJ}, C_{2K}, C_{3L})$ is calculated to be $P_{m2}^2/4$. (iii) The new genes *I* and *J* are encoded on the chromosomes C_1 and C_2, respectively, in one variant, while the new genes *K* and *L* are encoded on the chromosomes C_3 and C_4, respectively, in another variant. The conjugation of these two variants forms the zygote $(C_{1I}, C_{10}; C_{2J}, C_{20}; C_{3K}, C_{30}; C_{4L}, C_{40})$, and yields the monoploid descendant $(C_{1I}, C_{2J}, C_{3k}, C_{4L})$ with the probability $P_{m2}^2/8$.
The monoploid organism receiving four new genes can be also produced by the conjugation of a variant with one new gene *I* on the chromosome C_1 and another variant with three new genes *J*, *K* and *L*. The following three cases (iv) ~ (vi) are considerable for the location of the three new genes *J*, *K* and *L*. (iv) The three new genes are encoded on the same chromosome C_2. In this case, the conjugation of the two variants forms the zygote $(C_{1I}, C_{10}; C_{2JKL}, C_{20})$ and yields the monoploid descendant (C_{1I}, C_{2JKL}) with the probability of $P_{m1}P_{m3}/2$. (v) The new gene *J* is encoded on the chromosome C_2 and the other two new genes *K* and *L* are encoded on the chromosome C_3. The conjugation of these variants forms the zygote $(C_{1I}, C_{10}; C_{2J}, C_{20}; C_{3KL}, C_{30})$ and yields the descendant monoploid $(C_{1I}, C_{2J}, C_{3KL})$ with the probability of

$P_{m1}P_{m3}/4$. (vi) The new genes *J*, *K* and *L* are encoded on the chromosomes C_2, C_3 and C_4, respectively. In this case, the probability of producing the monoploid descendant $(C_{1l}, C_{2J}, C_{3K}, C_{4L})$ is further decreased to be $P_{m1}P_{m3}/8$.

As illustrated in the above examples in subsections 4.1 to 4.3, the probability P_{c2n} of producing the monoploid descendant received the even number $2n$ of new genes through one time of conjugation is generally expressed as

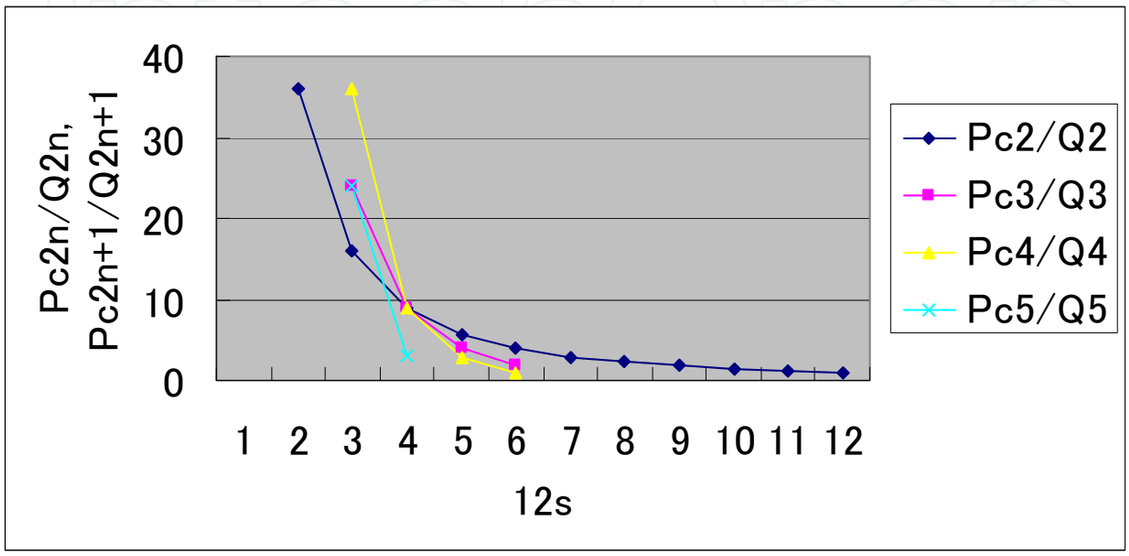


Fig. 2. The probabilities of producing the descendants received multiple kinds of new genes by the conjugation of monoploid organisms. The probability P_{c2n} of producing the descendant received $2n$ kinds of new genes is simply expressed as the square of the probability P_{mn}, i. e., $P_{c2n} = P_{mn}^2$. In the same way, the probability P_{c2n+1} of producing the descendant received $(2n+1)$ kinds of new genes is expressed as the product of the probabilities P_{mn+1} and P_{mn}, i. e., $P_{c2n+1} = P_{mn+1}P_{mn}$. Using the relations of $Q_{2n} = Q_n^2$ and $Q_{2n+1} = Q_nQ_{n+1}$, P_{c2n}/Q_{2n} and P_{c2n+1}/Q_{2n+1} are plotted against the twelve- fold reduction factor $12s$ for $n = 1$ and 2. It should be noted that the probabilities P_{c2n} and P_{c2n+1} are present in the wider range of reduction factor than the probabilities P_{m2n} and P_{m2n+1} shown in Fig. 1, respectively.

$$P_{c2n} = a_{n,n}P_{mn}^{\ 2} + b_{n+1,n-1}P_{mn+1}P_{mn-1} + \ldots\ldots\ldots\ldots \quad (21)$$

and the probability P_{c2n+1} of producing the monoploid descendant received the odd number $(2n+1)$ of new genes is expressed as

$$P_{c2n+1} = a_{n,n+1}P_{mn}P_{mn+1} + b_{n+2,n-1}P_{mn+2}P_{mn-1} + \ldots\ldots\ldots \quad (22)$$

Although the coefficients $a_{n,n}$, $a_{n,n+1}$, $b_{n+1,n-1}$, $b_{n+2,\ n-1}$ *etc.* depend not only on the number of new genes but also on the distribution of new genes over chromosomes in a complex way, the first terms are most important on the right sides of Eqs. (21) and (22), respectively. This is because the probabilities P_{mn} and P_{mn+1} in these terms are present in the wider range of reduction factor than those in other terms, as indicated in the preceding section. Thus, $P_{c2} \sim P_{mn}^2$ and $P_{c2n+1} \sim P_{mn}P_{mn+1}$, without the coefficients $a_{n,n}$ and $a_{n,n+1}$, are plotted against the

reduction factor s for some values of n in Fig. 2. The probability P_{c2n} is present in the same range of reduction factor as the probability P_{mn} is present and the probability P_{c2n+1} is present in the same range of reduction factor as the probability P_{mn+1} is. This indicates that the larger size of new genes not generated from the successive gene duplication in a single lineage of monoploid organisms can be assembled into an organism through conjugation. Although the values of P_{c2n} and P_{c2n+1} are smaller than those of P_{mn} and P_{mn+1}, respectively, due to the relations of $Q_{2n} < Q_n$ and $Q_{2n+1} < Q_{n+1}$, the smaller value of the probability only means the longer time for the monoploid organism to receive $2n$ or $(2n+1)$ new genes through the conjugation of variants than the time for a single lineage of monoploid organisms to generate n or $(n+1)$ new genes from gene duplication. If these larger new genes assembled by conjugation endow the descendant with a superior new character, such descendants increase their fraction as a new style of organisms. In this sense, it should be also noted that the descendant receiving $n(= 3, 4, 5, \cdots\cdots)$ kinds of new genes can be produced with the lower probability of $(1/2)^{n(n-1)/2}P_{m1}{}^n$ by the successive conjugation of variants having experienced gene duplication on different kinds of chromosomes. Such successive hybridization of different variants, each of them carrying one new gene, may become the main course to yield a new style of organisms carrying three or more new genes, if the homologous chromosomes different in carrying two or more new genes, such as those appeared in the first case of subsection 4.2 and in (i), (ii), (iv) and (v) of subsection 4.3, are severely incompatible upon the meiosis in the zygote.

At any rate, the eukaryote with the plural number of chromosomes is suitable to create new characters each expressed by many kinds of new genes, through the conjugation exchanging homologous chromosomes. This explains the diversity of various living styles of eukaryotes, ranging from the unicellular organisms called the *Protoctista* evolving various intracellular organs to the multicellular organisms evolving cell differentiation. As will be discussed in the last section, it is evident from the phylogeny of eukaryotes that the multicellularity and cell differentiation have also started in the monoploid eukaryotes, although the higher hierarchy of cell differentiation has developed in the diploid eukaryotes.

5. Higher eukaryotes in the diploid state

The higher eukaryote in the diploid state is characterized by the pairs of homologous chromosomes, and its large-scale evolution contains the process to establish the homozygote of new genes as well as their generation from gene duplication. Although the number of homologous chromosome pairs is different depending on the species of diploid organisms, a specific pair of homologous chromosomes (x_i, x_k) will be first focused for simplicity, where the suffixes i and k denote different mutations on the respective chromosomes. The number $n(x_i,x_k;t)$ of variants carrying such a pair (x_i,x_k) obeys the following time-change equation in the population of organisms exchanging the homologous chromosomes upon reproduction.

$$\frac{d}{dt}n(x_i,x_k;t) = \sum_{j,l} Q(x_i,x_k;t)_{ijxkl} R(M;x_i,x_k)_{ijxkl} n(x_i,x_j;t)n(x_k,x_l;t) - D(x_i,x_k)n(x_i,x_k;t)$$
$$+ \sum_{i',k'(\neq i,k)} \sum_{j,l} q(x_i,x_k \leftarrow x_{i'},x_{k'};t)_{i'j\times k'l} R(M;x_{i'},x_{k'})_{i'j\times k'l} n(x_{i'},x_j;t)n(x_{k'},x_l;t) \quad (23)$$

where $R(M; x_i, x_k)_{ijxkl}$ is the rate of producing the children (x_i, x_k) from the mating of a variant (x_i, x_j) with another variant (x_k, x_l) under a common material and energy source M, and $D(x_i, x_k)$ is the death rate of the organism (x_i, x_k). The apparent decrease factor $Q(x_i, x_k; t)_{ijxkl}$ is related with the mutation term $q(x_{i'}, x_{k'} \leftarrow x_i, x_k; t)_{ijxkl}$ in the following way.

$$Q(x_i, x_k; t)_{ij\times kl} = 1 - \sum_{i',k'(\neq i,k)} q(x_{i'}, x_{k'} \leftarrow x_i, x_k; t)_{ij\times kl} \tag{24}$$

Although Eq. (23) makes no distinction between the male and the female for simplicity, this distinction does not essentially alter the following process of evolution.
In the same way as for monoploid organisms, the population behavior of diploid organisms becomes transparent by transforming Eq. (23) into the equation concerning the total number of organisms given by $B(t) = \Sigma_i \Sigma_k n(x_i, x_k; t)$ and that concerning the fraction of variants (x_i, x_k) defined by $f(x_i, x_k; t) = n(x_i, x_k; t)/B(t)$. These equations are expressed in the following forms, respectively.

$$\frac{d}{dt} B(t) = \overline{W}(t) B(t) \tag{25}$$

$$\begin{aligned} \frac{d}{dt} f(x_i, x_k; t) &= \{W(x_i, x_k; t) - \overline{W}(t)\} f(x_i, x_k; t) \\ &+ \sum_{i',k'} \sum_{j,l} q(x_i, x_k \leftarrow x_{i'}, x_{k'}; t)_{i'jxk'l} R(M; x_{i'}, x_{k'})_{i'j\times k'l} f(x_{i'}, x_j; t) f(x_{k'}, x_l; t) B(t) \end{aligned} \tag{26}$$

where the increase rate $W(x_i, x_k; t)$ of the variant (x_i, x_k) is defined by

$$W(x_i, x_k; t) \equiv \sum_j \sum_l R(M; x_i, x_k)_{ij\times kl} f(x_i, x_j; t) f(x_k, x_l; t) B(t) / f(x_i, x_k; t) - D(x_i, x_k) \tag{27}$$

the average increase rate $\overline{W}(t)$ is by

$$\overline{W}(t) \equiv \sum_i \sum_k W(x_i, x_k; t) f(x_i, x_k; t) \tag{28}$$

and $q(x_i, x_k \leftarrow x_i, x_k; t)_{ijxkl}$ is defined by $Q(x_i, x_k; t)_{ijxkl} - 1$. If the suffixes i, j, k and l denote the point mutations in existing genes, Eq. (26) represents Darwinian evolution gradually leading to the organisms with an optimal increase rate, each characterized by (x_{opt}, x_{opt}).
Because the gene duplication occurs only rarely, it is natural to consider that the large-scale evolution due to gene duplication starts after the organisms (x_{opt}, x_{opt}) have been dominant in the population. If the chromosome having experienced gene duplication is newly denoted by x_i and the point mutation is neglected, the fraction $f(x_i, x_{opt}; t)$ of variants (x_i, x_{opt}) obeys the following equation as a special case of Eq. (26).

$$\begin{aligned} \frac{d}{dt} f(x_i, x_{opt}; t) &= \{W(x_i, x_{opt}; t) - \overline{W}(t)\} f(x_i, x_{opt}; t) \\ &+ q(x_i, x_{opt} \leftarrow x_{opt}, x_{opt}; t)_{optoptxoptopt} R(M; x_{opt}, x_{opt})_{optoptxoptopt} f^2(x_{opt}, x_{opt}; t) B(t) \end{aligned} \tag{29}$$

where the increase rate $W(x_i, x_{opt}; t)$ of the variant (x_i, x_{opt}) is given by

$$W(x_i, x_{opt}; t) = R(M; x_i, x_{opt})_{iopt\times optopt} f(x_{opt}, x_{opt}; t) B(t) - D(x_i, x_{opt}) \tag{30}$$

and the average increase rate $\overline{W}(t)$ is by

$$\overline{W}(t) = W(x_i, x_{opt}; t) f(x_i, x_{opt}; t) + W(x_{opt}, x_{opt}; t) f(x_{opt}, x_{opt}; t) \tag{31}$$

The probability of generating a new style of organisms carrying a new gene is derived from Eq. (29). In the population where the organisms (x_{opt}, x_{opt}) are dominant, $\overline{W}(t)$ is approximately equal to $W(x_{opt}, x_{opt})$, and both $f(x_{opt}, x_{opt}; t)$ and $B(t)$ are hardly dependent on time. Eq. (29) is then integrated to give the following relation between the fraction of variants $f(x_i, x_{opt})$ and that of dominant organisms $f(x_{opt}, x_{opt})$.

$$f(x_i, x_{opt}) = \frac{q(x_i, x_{opt} \leftarrow x_{opt}, x_{opt})_{optoptxoptopt} R(M; x_{opt}, x_{opt})_{optopt \times optopt}}{W(x_{opt}, x_{opt}) - W(x_i, x_{opt})} f^2(x_{opt}, x_{opt}) B \tag{32}$$

where the rate of generating the gene duplication i is defined for a sufficiently long time t by

$$q(x_i, x_{opt} \leftarrow x_{opt}, x_{opt})_{optoptxoptopt} \equiv \frac{1}{t} \int_0^t q(x_i, x_{opt} \leftarrow x_{opt}, x_{opt}; \tau)_{optoptxoptopt} d\tau \tag{33}$$

Although this relation (32) seems to be different in including the population size B from Eq. (10) of monoploid organisms at first glance, the denominator on the right side of Eq. (32) also contains the population size B as seen in Eqs. (27) and (30). If the population size is large enough to neglect the difference in death rate between the variant (x_i, x_{opt}) and the dominant organism (x_{opt}, x_{opt}), therefore, the difference in the increase rate $W(x_{opt}, x_{opt}) - W(x_i, x_{opt})$ is approximately equal to $\{R(M; x_{opt}, x_{opt})_{optoptxoptopt} - R(M; x_i, x_{opt})_{ioptxoptopt}\} B f(x_{opt}, x_{opt})$, and Eq. (32) is reduced to be

$$f(x_i, x_{opt}) = \frac{q(x_i, x_{opt} \leftarrow x_{opt}, x_{opt})_{optoptxoptopt} R(M; x_{opt}, x_{opt})_{optopt \times optopt}}{R(M; x_{opt}, x_{opt})_{optopt \times optopt} - R(M; x_i, x_{opt})_{iopt \times optopt}} f(x_{opt}, x_{opt}) \tag{34}$$

This is essentially the same form as Eq. (10) of the monoploid organism in the case when the gene duplication hardly changes the death rate, i. e., $D(x_i) \approx D(x_{opt})$. Denoting the probability of generating a new gene I from the gene duplicated part i by $q(x_I \leftarrow x_i)$, the probability $P_{d1}(x_I, x_o \leftarrow x_o, x_o)$ that a new style of the organism (x_I, x_o) carrying the new gene I heterogeneously is generated from the original style of an organism (x_o, x_o) is expressed as

$$P_{d1}(x_I, x_o \leftarrow x_i, x_o) = \frac{q(x_I \leftarrow x_i) q(x_i, x_o \leftarrow x_o, x_o)_{ooxoo} R(M; x_o, x_o)_{oo \times oo}}{R(M; x_o, x_o)_{oo \times oo} - R(M; x_i, x_o)_{io \times oo}} \tag{35}$$

where x_{opt} in Eq. (34) is rewritten into x_o with the meaning of the original type chromosome. Thus, a new style diploid organism also arises from the minor members in the population just like the case of monoploid organisms.

However, the content of the above probability in diploid organisms is different from the case of monoploid organisms in the following points. First of all, the reproducing rate $R(M; x_i, x_o)_{ioxoo}$ is only the half of $R(M; x_o, x_o)_{ooxoo}$ even in the random partition of homologous chromosomes, and the former may be further decreased by the lowering of the biological activity of the variant (x_i, x_o). Second, the further gene duplication to produce two or more new genes is hardly expected in the homologous chromosomes (x_i, x_o), because the fraction

of such variants experienced successive gene duplication becomes much lower, not only due to the severer lowering of biological activity but also by the severer incompatibility of homologous chromosomes or by the separation of the chromosomes carrying different origins of duplicated genes in the descendants. That is, if the further gene duplication j occurs on the chromosome x_i to yield x_{ij}, for example, the incompatibility of chromosomes x_{ij} and x_o becomes severer upon the mitosis and/or the meiosis. If the gene duplication j occurs on the chromosome x_o to yield x_j, on the contrary, the chromosome x_j is separated from the chromosome x_i in the descendants.

In spite of such conservative property, the diploid organism with the plural number of homologous chromosome pairs can give rise to a new style of an organism getting together two or more new genes, through the successive hybridization among the satellite variants having experienced gene duplication on different kinds of chromosomes. As the first example, the appearance of a new style organism received two kinds of new genes I and J will be considered by this mechanism of hybridization. In this case, two pairs of homologous chromosomes $(x_0, x_0; y_0, y_0)$ are focused, and the probability of generating the heterozygote $(x_I,x_o;y_J,y_o)$ from the original style of organisms $(x_o,x_o;y_o,y_o)$ is considered through the hybridization of two types of variants $(x_i,x_o;y_o,y_o)$ and $(x_o,x_o;y_j,y_o)$. According to Eq. (35), this probability $P_{d2}(x_I,x_o;y_J,y_o \leftarrow x_o,x_o;y_o,y_o)$ is given by

$$\begin{aligned} &P_{d2}(x_{I,}x_o;y_J,y_o \leftarrow x_o,x_o;y_o,y_o) \\ &= \frac{q(x_I \leftarrow x_i)q(x_i,x_o \leftarrow x_o,x_o)_{oooox oooo} R(M;x_o,x_o;y_o,y_o)_{oooo\times oooo}}{R(M;x_o,x_o;y_o,y_o)_{oooo\times oooo} - R(M;x_i,x_0;y_o,y_o)_{iooo\times oooo}} \\ &\cdot \frac{q(y_J \leftarrow y_j)q(y_j,y_o \leftarrow y_o,y_o)_{oooox oooo} R(M;x_o,x_o;y_o,y_o)_{oooo\times oooo}}{R(M;x_o,x_o;y_o,y_o)_{oooo\times oooo} - R(M;x_o,x_o;y_j,y_o)_{oojo\times oooo}} r_2 \end{aligned} \tag{36}$$

where r_2 is the ratio of the children received two kinds of new genes I and J, taking the value of $(1/2)^2$ in the case of random partition of homologous chromosomes. In order to show the result of further hybridization process, Eqs. (35) and (36) will be simplified in their expression at this stage. The probabilities $q(x_I \leftarrow x_i)$ and $q(x_J \leftarrow x_j)$ of generating new genes I and J from duplicated parts i and j in Eqs. (35) and (36) may be equal to the corresponding probabilities $q_{xI,xi}$ and $q_{xIJ,xIj}$ in Eqs. (11) and (16), respectively, because the nucleotide base substitution rate is almost common to both eukaryotes and prokaryotes (Kimura, 1980; Otsuka et al., 1997). Although it is still difficult to estimate the occurrence frequency of gene duplication, this frequency is also assumed to be common to both monoploid and diploid organisms, i. e., $q_{xi,xo} \sim q(x_i,x_o \leftarrow x_o,x_o)_{oooox ooo}$ and $q_{xij,xi} \sim q(y_j,y_o \leftarrow y_o,y_o)_{oooox oooo}$, for simplicity. The reproducing rates $R(M; x_o,x_o; y_o,y_o)_{oooox oooo}$, $R(M; x_i,x_o; y_o,y_o)_{iooox oooo}$ and $R(M;x_o,x_o;y_j,y_o)_{oojox oooo}$ are simply denoted by R, $R(1 - S_1)$ and $R(1 - S_2)$, respectively, with the reduction factors S_1 and S_2, where both S_1 and S_2 satisfy the relation $1/2 < S_1, S_2 < 1$ as noted already. Eqs. (35) and (36) are then rewritten into

$$P_{d1}(x_I,x_o \leftarrow x_o,x_o) = \frac{Q_1}{S_1} \tag{37}$$

and

$$P_{d2}(x_I,x_o;y_J,y_o \leftarrow x_o,x_o;y_o,y_o) = \frac{Q_2}{S_1 S_2} r_2 \tag{38}$$

respectively. Here, Q_1 and Q_2 represent the terms $q(x_I \leftarrow x_i)q(x_i,x_o \leftarrow x_o,x_o)_{ooxoo}$ and $q(x_I \leftarrow x_i)q(x_i,x_o \leftarrow x_o,x_o)_{ooooxoooo}q(y_J \leftarrow y_j)q(y_j,y_o \leftarrow y_o,y_o)_{ooooxoooo}$, respectively. As the extension, the probability P_{dn}, with which a new style diploid organism carrying n kinds of new genes heterogeneously is generated from the successive hybridization of variants, is expressed in the following form.

$$P_{dn} = \frac{Q_n}{S_1 S_2 \cdots\cdots S_n} r_n \tag{39}$$

where S_i $(i=1, 2, \cdots\cdots, n)$ is the reduction factor in the producing rate of the variant carrying duplicated genes on the chromosome i, Q_n is the product of the probabilities of generating n kinds of new genes from gene duplication on the respective chromosomes and r_n is the ratio of the children received these new genes. Although reduction factors S_i's in Eq. (39) independently take values in the range of $1/2 < S_i < 1$, they are tentatively represented by a common variable S for a simple illustration of n dependence of P_{dn} in a figure. Then, the probability P_{dn} is simply expressed by

$$P_{dn} = \frac{Q_n}{S^n} r_n \tag{40}$$

These probabilities P_{dn}'s in Eq. (40) are plotted against the reduction factor S in Fig. 3 for several values of n. As noted already, the reduction factor S is restricted to the

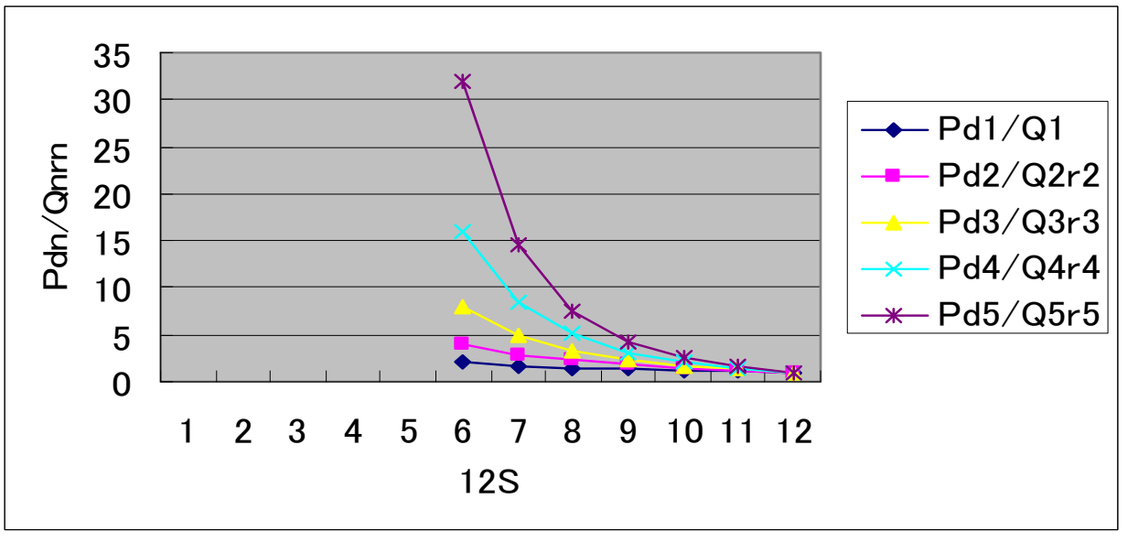


Fig. 3. The probabilities of generating new genes from gene duplication and successive hybridization in diploid organisms. On the basis of Eq. (40), the values of $P_{dn}/Q_n r_n$ are plotted against the twelve-fold reduction factor $12S$ for $n = 1, 2, 3, 4$ and 5. The value of r_1 is equal to one, and the curve of P_{d1} *vs* S is consistent with the curve of P_{m1} *vs* s in Fig. 1, but the range of reduction factor S is restricted to the range of $1/2 < S < 1$. For a larger value of n, however, the probability P_{dn} is still present in the range of $1/2 < S < 1$. Although $P_{dn+1}/Q_{n+1}r_{n+1}$ is larger than $P_{dn}/Q_n r_n$ in the figure, P_{dn+1} is smaller than P_{dn}. This is because $Q_{n+1}r_{n+1}$ is smaller than $Q_n r_n$, as discussed in the text.

range of $1/2 < S < 1$, and the probability P_{dn} is within the range of $2^n Q_n r_n > P_{dn} > Q_n r_n$. If the homologous chromosomes are randomly partitioned into the children regardless of carrying a new gene or not, r_n takes the value of $(1/2)^n$ and the above relation of P_{dn} becomes $Q_n > P_{dn} > Q_n/2^n$. Moreover, the value of Q_n becomes smaller for the larger value of n, and the probability P_{dn} becomes lower as the number of new genes assembled by hybridization is increased. The lower probability means the longer time or more generations for a new style organism carrying more kinds of new genes to appear. Thus, the diploid organism has a chance to acquire many kinds of new genes by hybridization, but it takes a longer time to realize this chance.

Moreover, the process to establish the homozygote is further continued after the new style organism carrying n kinds of new genes heterogeneously is generated with the probability P_{dn}. Although it is laborious to follow this process completely, the essence of this process can be elucidated by investigating the ratio of children that receive these new genes homogeneously and heterogeneously from the mating between the organisms each carrying n kinds of new genes heterogeneously. If the chromosomes in each homologous pair are randomly partitioned into the children regardless of carrying a new gene or not, the ratio of children receiving $(n-k)$ kinds of new genes is calculated to be ${}_nC_k 3^{n-k}/4^n$ with the normalization factor 4^n, where k takes a value ranging from zero to n. This indicates that more than half of the children receive all new genes $(k = 0)$ for $n = 1, 2$. If the one or two new genes exhibit an excellent character, therefore, the descendants increase their fraction monotonously as a new style of organisms. However, the ratio of children receiving a full set of new genes becomes smaller for a larger value of n. In the case of $n = 5$, for example, the ratio of the children that receive five kinds of new genes $(k = 0)$ decreases to $(3/4)^5$, while other five types of children each appear with the ratio of $(3/4)^4/4$ by receiving four kinds of new genes $(k = 1)$ in different ways. When a biologically meaningful character is expressed by five kinds of new genes, therefore, only $(3/4)^5$ of the children succeed in expressing this character but other five types of children are reserved as those carrying 'hidden genes' for producing other characters by further hybridization with other types of variants. Such divergence of characters becomes more outstanding when a larger number of new genes are required for the expression of a character. This divergent property in the process to establish many kinds of new genes as the homozygote explains the explosive divergence of body plans that has occasionally occurred in diploid organisms, because the cell differentiation is a representative character expressed by many kinds of genes and its hierarchical evolution constructs body plans, as will be discussed in the next section. Until the new style organisms are established as the homozygote, the mating between the variants of heterozygote also regenerates the original style of organisms. The phenomenon called the "reversion" or "atavism" in classical biology may be the vestige of this evolutionary process to establish the homozygote.

If the influence of transposons is explicitly considered, it makes the above process more complicated in such a way that duplicated genes are separately transferred to different kinds of chromosomes. When various origins of duplicated genes or new genes are concentrated on one chromosome, however, the descendants received such a chromosome may be extinct due to the incompatibility of this chromosome with its partner chromosome not carrying any new gene. Thus, many kinds of new genes for expressing a new character may be scattered over different kinds of chromosomes in survivors just like the result of the present model scheme.

6. Conclusions and discussion

The variants, which experienced gene duplication, first decline to be minor members in a population by the load of carrying extra gene(s), but some of them revives as a new style of organisms by the generation of new gene(s) from the counterpart of duplicated genes. After the new gene(s) appear, the new style organisms increase their fraction being further elaborated by Darwinian evolution. This course of the large-scale evolution is essentially the same in any type of organisms, and this is a necessary condition for the new style of organisms and the original style of organisms to be able to coexist utilizing different material and energy sources or to live in separate areas, showing a striking contrast to the survival of the fittest in Darwinian evolution. This evolutionary pattern also gives an explanation to the punctuated mode of evolution, which has been proposed from paleontology against the gradual accumulation of variants in Darwinian evolution (Eldredge & Gould, 1972).

However, the detailed processes of this large-scale evolution are different depending on the types of genome constitution and transmission. The monoploid organism is suitable to generate one new gene step by step testing its biological function, but hardly generates many kinds of new genes simultaneously. The lower eukaryote, whose genome consists of the plural number of chromosomes, resolves this difficulty to produce a new style of organisms receiving many kinds of new genes by the conjugation of variants carrying different origins of new genes. The diploid organism can also produce a new character responsible for multiple kinds of new genes by the successive hybridization of different variants but its conservative property requires the succeeding process to establish the homozygote of these genes. This process becomes longer for a larger number of new genes to be established. During this long process, the further hybridization with other variants also occurs, occasionally yielding the explosive divergence of new characters depending on the combinatorial sets of new genes. This conclusion of the present study explains the recently revealed evolutionary patterns of prokaryotes and eukaryotes to a great extent, getting an insight into the problems how and why the monoploid eukaryotes have evolved to the diploid eukaryotes.

According to the analyses of base-pair changes in ribosomal RNAs, the main lineages of present-day prokaryotes diverged $3.0x10^9$ years ago, developing various chemical syntheses, O_2-releasing photosynthesis and O_2 respiration, respectively (Otsuka et al., 1999), after the earlier divergence of archaebacteria, eubacteria and eukaryotes (Sugaya & Otsuka, 2002). Several stages from simple electron transport pathways to O_2 respiration and O_2-releasing photosynthesis are still observed in the present-day eubacteria and the elongation of the pathways has taken place stepwise by gene duplication, as can be traced from the amino acid sequence similarities between their component proteins and the ubiquitous permeases (Otsuka, 2002; Otsuka & Kawai, 2006), although such similarity search of amino acid sequences is not systematically carried out yet for chemical syntheses. However, the excellent abilities of O_2 respiration and O_2-releasing photosynthesis cannot be fully exhibited in the simple cell structure of prokaryotes (Otsuka, 2005), and the genome size of the eubacteria having these abilities is also limited to the order of 10^6 *bp* compactly encoding *3,000 ~ 4,000* genes like the other prokaryotes (Wheeler et al., 2004).

On the other hand, the eukaryotes have experienced much more evolutionary events until some of them establish the diploid state. The ancestral eukaryote probably became the predator of eubacteria by developing the intracellular structure, endocytosis and exocytosis

as well as the signal transduction network. Such cell structure would have been suitable to acquire the mitochondria as the endosymbionts of O_2-respiratory eubacteria, which is estimated to have occurred $2.0x10^9$ years ago (Margulis, 1981; Yang et al., 1985; Otsuka et al., 1999). Under the supply of abundant ATP molecules efficiently synthesized by the mitochondria, the yeast *Sacchromyces serevisiae*, which appeared $1.8\ x10^9$ years ago (Otsuka et al., 1997), has expanded its genome to $1.2x10^7$ *bp* encoding *6, 300* genes (Wheeler et al., 2004) and can take the diploid state under nutrient conditions, although it usually takes the monoploid state. The enlarged genome consisting of the plural number of chromosomes also requires the special apparatus for the faithful segregation of sister DNAs upon cell division, in contrast to the prokaryotes where the membrane attachment mechanism of DNA only operates (Jacob et al., 1963; Ogden et al., 1988; de Boer, 1993). Multiple kinds of gene products such as the primitive spindle pole and kinetochore and polar microtubules are already present in the yeast (Alberts et al., 1994) while several bundles of microtubules only pass through tunnels in the typical *Dinoflagellates* (Kubai, 1975; Hearth, 1980; Wise, 1988). Thus, the components of this auxiliary apparatus for cell division may have evolved step by step at the stage of unicellular eukaryotes. Although the molecular mechanism underlying the switching from the monoploid to diploid states and vice versa is not fully clarified yet, the example of yeast indicates that this mechanism itself has also evolved at the stage of unicellular eukaryotes.

However, the evolution from the monoploid eukaryote to the diploid eukaryote has taken place considerably gradually via several stages. This is reasonable because the diploid state is an extreme case of gene duplication. If the genome size jumps from N to $2N$, this means the increase in the stored energy and systematization from $E_s(N,S_N)$ and S_N to $E_s(2N, S_{2N})$ and S_{2N}. Thus, the acquired energy must be also increased to maintain the biological activity. As indicated already (Otsuka, 2008), this increase in acquired energy is possibly attained by the cooperative action of differentiated cells. However, the evolution of cell differentiation cannot occur suddenly. On this problem, the present result throws light, in the point that the conjugation of lower eukaryotes with the plural number of chromosomes is suitable to assemble many kinds of new genes necessary for cell differentiation. In fact, the recently revealed phylogeny of eukaryotes strongly suggests at least the following five stages in the evolution from the monoploid to diploid eukaryotes. (a) First, the monoploid eukaryote evolves the conjugation to exchange the homologous chromosomes. (b) Second, this eukaryote then develops multicellularity and cell differentiation in the monoploid state by assembling many kinds of new genes. (c) Third, the cell differentiation also advances to the cells in the diploid state. (d) Fourth, the eukaryote evolves to alternate the monoploid generation and the diploid generation. (e) Finally, the eukaryote evolves to the diploid organism with the higher hierarchy of cell differentiation.

As far as the present knowledge of the phylogeny of eukaryotes and their genome constitution (Otsuka et al., 1997) is concerned, the first lineages having evolved multicellularity and cell differentiation are some of the fungi that appear after yeast and the sea algae, which have further acquired photosynthetic eubacteria as the endosymbionts in the lineage of fungi (Van den Eynde et al., 1988). However, the most advanced one of them still remains at the stage (d), alternating the monoploid generation and the diploid generation. Apart from the lineages of fungi and algae, the evolution of advancing the cell differentiation to the diploid state has taken place in the animals and the green plants, whose divergence is estimated to have occurred $1.2x10^9$ years ago (Dickerson, 1971). Among

them, the green plants, which have also acquired the chloroplasts as the endosymbionts of photosynthetic eubacteria independently of sea algae, provide a representative example of the above five stages of evolution from the monoploid organisms to the diploid organisms. The *Cojugatae* such as *Roya* and *Spirogyra* are at the stage (a), the *Chara* of *Charophyta* is at the stage (b), the *Bryophyta* is at the stage (c), where the fertilized egg on female gametophyte grows into sporangium, the *Pterophyta* is at the stage (d), and the seed plants are at the stage (e). According to the recent analysis of neutral nucleotide base substitutions in *rbcL* genes on the chloroplast genomes (Kawai & Otsuka, 2004), the divergence of *Charophyta* and *Bryophyta* occurred more than 10^9 years ago, the divergence of *Bryophyta* and *Pterophyta* occurred around $4.7x10^8$ years ago, and the divergence of *Pterophyta* and seed plants occurred about $3.8x10^8$ years ago.

The molecular mechanism underlying the cell differentiation is not fully clarified yet, but it is probably based on a set of receptors, the corresponding ligands, signal transduction proteins, transcriptional regulators as well as the proteins exhibiting the respective cell-type specific functions. Moreover, the amino acid sequences of these proteins under the control of signal transduction network become longer by the attachment of special amino acid residue repeats such as serines and threonines. Thus, the assembly of so many kinds of large genes into a genome must have first progressed under the scheme of the conjugation of monoploid eukaryotes with the plural number of chromosomes. After a set of genes responsible for cell differentiation are established in the monoploid state, the increase in the repertoire of the respective members would have occurred relatively easily. In particular, a small number of nucleotide base substitutions could bring about the expansion of such protein families as transcriptional regulators, receptors and kinases associated with the signal transduction network, although these kinds of proteins have their origins at the stage of unicellular eukaryotes. The increase in acquired energy by the cell differentiation in the monoploid state makes it possible to realize the cell differentiation in the diploid state. The example of green plants suggests that the cell differentiation in the diploid state has started from the zygote and gradually spread to form other organs of diploid cells, resulting in the alternation of the monoploid generation and the diploid generation. The diploid state is suitable to protect the differentiated cells from the point mutations, as will be discussed in the last part of this section, but it takes a longer time or many generations to establish a set of many genes for advancing the further cell differentiation in the diploid state as the homozygote. Although this is the barrier lying between the stage (d) and the stage (e), the diploid organisms having gone over this barrier receive a good chance to produce various combinatorial sets of new genes leading to the explosive divergence of morphological characters. Such explosive divergence has the merit of testing simultaneously various characters for survival.

Although any example of animals at the stages (b) and (c) is hardly found at the present time, the *Cnidaria* still alternates the monoploid generation and the diploid one. The divergence of *Cnidaria* and the common ancestor of other animals occurred immediately after the animal-plant divergence (Otsuka & Sugaya, 2003). The famous explosion of body plans giving rise to *Annelida, Mollusca, Arthropoda, Echinodermata* and *Chordata*, which is first found by the fossil record of Ediacara and Avalon faunas (Mathews & Missarzhersky, 1975; Rozanov & Zhuravlev, 1992) and of Cambrian Burgess Shale (Gould, 1989) and then estimated to have occurred successively during the period of $9\sim6x10^8$ years ago by the analysis of neutral nucleotide base substitutions (Otsuka & Sugaya, 2003), is probably based on the evolutionary scheme of diploid organisms described in sections 5, because these

animals show the living style defined as the diploid organism in the present chapter. Such divergence of body plans occasionally occurred afterwards in each of the above phyla. The examples well investigated in paleontology are the divergence of *Placodermi*, cartilaginous fish and bony fish, the divergence of amphibians, reptiles and mammals, and the divergence of dinosaurs and birds, which occurred in the *Chordata* within the recent $4x10^8$ years (Carroll, 1988). The seed plants also show the similar tendency in the successive divergence of *Coniferophyta*, *Anthophyta* and their relatives (Fairon-Demaret & Scheckler, 1987; Rothwell et al., 1989; Rowe, 1992; Stewart & Rothwell, 1993; Kawai & Otsuka, 2004), although many of these seed plants can also self-reproduce by the parthenogenesis and their explosive feature seems mild. Although the explosive divergence of body plans can be also explained by the biological activity expressed in terms of the interaction between differentiated cells (Otsuka, 2008), the present study derives this divergence from the aspect of the generation of new genes from gene duplication in diploid organisms.

The fossil record of these examples indicates that the original style of organisms prospered over a wide region when new styles of organisms diverged, being consistent with the present theory. The prosperity of the original style of organisms means that their biological activity is high, and this is necessary to permit the existence of variants carrying duplicated genes in the population and further to enhance the chance of assembling many kinds of new genes into a genome by hybridization. This is in contrast with Darwinian evolution generating new species adapted to the special environment of a geographically isolated district by accumulating point mutations.

Finally, some discussions will be given to the problem why the cell differentiation has been shifted from the monoploid state to the diploid state. This problem arises from the present result that the diploid organism is not necessarily superior to the monoploid organism with the ability of exchanging homologous chromosomes in assembling many kinds of new genes for cell differentiation. The main reason of this shifting may be the protection of differentiated cells from the point mutations due to the miss in repairing damaged nucleotide bases. First of all, many more genes are needed to develop the higher hierarchy of cell differentiation. In fact, the genome size of higher eukaryotes is expanded to the order of $10^8 \sim 10^9$ *bp*, e. g., $1.2x10^8$ *bp* encoding *24,000* genes in *Arabidopsis thaliana*, $1.4x10^8$ *bp* encoding *13,000* genes in *Drosophila melanogaster* and $3.1x10^9$ *bp* encoding *30,000* genes in *Homo sapiens* (Wheeler et al., 2004). Second, it takes a longer time, one or more years, to develop the higher hierarchy of cell differentiation to form an adult form in the higher eukaryotes, although the growth rate and the lifetime seem to be further regulated differently depending on species. On the other hand, the mutation rate due to the miss in repair is 10^{-9} per site per year in eukaryotes as well as in prokaryotes (Kimura, 1980; Otsuka et al., 1997). As the evidence for the above discussion, the males of some species of ants and bees are born by the haploid parthenogenesis, showing that the monoploid state is sufficient for the high hierarchy of cell differentiation during their short lifetime. Although the accuracy in repairing damaged DNAs can be raised by the additional energy for proofreading (Hopfield, 1974), the evolution of organisms has not been directed to use such additional energy. On the contrary, the nucleotide base substitution rate becomes about tenfold faster in animal mitochondrial genome than in the host cell genome, as is used to resolve the phylogeny of recently diverged animals (Hasegawa et al., 1985; Pesole et al., 1999; Otsuka et al., 2001). This faster mutation rate strongly suggests that the energy to proofread the small genome of mitochondria is diminished and instead the saved energy is used to raise the biological activity of the host cell. For the

same sequence length of gene duplication, therefore, the reduction factor may take a smaller value in animals than in lower eukaryotes and prokaryotes. Thus, the fraction of variants carrying the 'hidden genes' generated from gene duplication may be high enough to hybridize between them in higher eukaryotes, especially in animals. Such 'hidden genes' belong to the category of 'genetic polymorphism', which has been first proposed by Ford (1965) and is subsequently disclosed by electrophoretic studies, although the 'genetic polymorphism' was only regarded as the result of random fixation of selectively neutral or nearly neutral mutations by the neutralist (Kimura, 1977).

It is still somewhat mysterious that the introns and spacers are more expanded in animal genomes than in the genomes of other eukaryotes. Such expansion can be seen from the ratio of the genome size to the number of encoded genes described above. It is conceivable that the introns are necessary for messenger RNAs to pass through the nuclear membrane and the spacers enhance the crossing over of homologous chromosomes without injuring established genes, but the expansion of introns and spacers in the higher eukaryotes might imply any other biological role of their nucleotide sequences.

7. References

Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K. & Watson, J. D. (1994). *Molecular Biology of the Cell.* 3rd Ed. Garland Publishing Inc., New York & London. pp. 941-943

Birney, E. & Other fifty persons (Ensembl 2006). Database Issue. *Nucleic Acids Res.* Vol. 1, D556-D561

Carroll, R. L. (1988). *Vertebrate Paleontology and Evolution.* W. H. Freeman, New York

Darwin, C. (1859). *The Origin of Species*. John Murry, London

de Boer, R. A. J. (1993). Chromosome Segregation and Cytokinesis in Bacteria. *Curr. Opin. Cell Biol*., Vol 5, pp. 232-237

Dickerson, R. E. (1971). The Structure of Cytochrome c and the Rate of Evolution. *J. Mol. Evol*., Vol. 1, pp. 26-45

Dobzhansky, T. (1941). *Genetics and the Origin of Species.* 2nd Ed. Columbia University Press, New York

Eigen, E. (1971). Selforganization of Matter and Evolution of Biological Macromolecules. *Die Natürwissenshaften,* Vol. 58, pp. 465-523

Eldredge, N. & Gould, S. J. (1972). Punctuated Equilibria: An Alternative to Phyletic Gradualism. In: *Models in Paleobiology,* T. J. M. Schopf, (Ed.). p. 82, Freeman and Cooper, San Francisco

Fairon-Demaret, M. & Scheckler, E. S. (1987). Typification and Redescription of *Moresnetia zalesskyi* Stockmans, 1948, an Early Seed Plant from the Upper Famennian of Belgium. *Bull Inst Roy Sci Nat Belg Sci Terre*, Vol. 57, pp. 193-199

Ferris, S. D. & Whitt, G. S. (1979). Evolution of the Differential Regulation of Duplicated Gene after Polyploidization. *J. Mol. Evol.,* Vol. 12, pp. 267-317

Fisher, R. A. (1930). *The General Theory of Natural Selection.* Oxford Univ. Press, London and New York

Ford, E. B. (1965). *Genetic Polymorphism*. Faber & Faber, London

Gilbert, W. (1978). Why Genes in Pieces ? *Nature*, Vol. 271, p. 501

Gould, S. J. (1989). *Wonderful Life. The Burgess Shale and the Nature of History.* W. W. Norton & Company Inc., New York

Hasegawa, M.; Kishino, H. & Yano, T. (1985). Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA. *J. Mol. Evol.*, Vol. 22, pp. 160-174

Hearth, I. B. (1980). Variant Mitosis in Lower Eukaryotes: Indicators of the Evolution of Mitosis? *Int. Rev. Cytol.*, Vol. 64, pp. 1-80

Hopfield, J. J. (1974). Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity. *Proc. Nat. Acad. Sci. USA,* Vol. 71, pp. 4135-4139

Huxley, J. (1943). *Evolution: The Modern Synthesis. The Grandson of T. H. Huxley Explores Mendelism, Evolutionary Trends, and Genetic Systems.* Harper & Row, New York

Ingram, V. M. (1963). *The Hemoglobin in Genetics and Evolution.* Columbia Press, New York

Jacob, F.; Brenner, S. & Cuzin, F. (1963). On the Regulation of DNA Replication in Bacteria. *Cold Spr. Harb. Symp. Quant. Biol.,* Vol. 28, pp. 329-348

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.,* Vol. 58, pp. 479-489

Kimura, M. (1977). Causes of Evolution and Polymorphism at the Molecular Level. *Proceedings of the Second Taniguchi International Symposium on Biophysics*, pp. 1-28, Mishima, Japan

Kimura, M. (1980). A Simple Method for Estimating Evolutionary Rate of Base Substitutions through Comparative Studies of Nucleotide Sequences. *J. Mol. Evol.,* Vol. 16, pp. 111-120

Kojima, S. & Otsuka, J. (2000a). Characterization of Organisms by the Paralogous Relationships of Proteins. Part I. *Escherichia coli*. *Res. Commun. in Biochemi Cell & Molec. Biology*, Vol. 4, pp. 59-82

Kojima, S. & Otsuka, J. (2000b). Characterization of Organisms by the Paralogous Relationships of Proteins. Part II. *Methanococcus jannaschii*. *Res. Commun. in Biochemi. Cell & Molec. Biology*, Vol. 4, pp. 83-100

Kojima, S. & Otsuka, J. (2000c). Characterization of Organisms by the Paralogous Relationships of Proteins. Part III. *Saccharomyces cerevisiae*. *Res. Commun. in Biochemi. Cell & Molec. Biology*, Vol. 4, pp. 101-138

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. in Biochemi. Cell & Molec. Biology*, Vol. 6, pp. 72-102

Kubai, D. F. (1975). The Evolution of the Mitotic Spindle. *Int. Rev. Cytol.*, Vol. 43, pp. 167-227

Margulis, L. (1981). *Symbiosis in Cell Evolution: Life and its Environment on the Early Earth.* W. H. Freeman, San Francisco

Mathews, S. C. & Missarzhevsky, V. (1975). Small Shelly Fossils of Late Precambrian and Early Cambrian Age: A Review of Recent Work. *Q. J. Geol. Soc. London,* Vol. 131, pp. 289-304

Mayer, E. (1942). *Systematics and the Origin of Species. A Correlation of the Evidence and Points of View of Systematics WithThose of Other Biological Disciplines, Particularly Genetics and Ecology*. Columbia University Press, New York

Nowak, M. A.; Bonhoeffer, S. & May, R. M. (1994). Spacial Games and the Maintenance of Cooperation. *Proc. Natl. Acad. Sci. USA,* Vol. 91, pp. 4877-4881

Ogden, G. B.; Pratt, M. J. & Schaechter, M. (1988). The Replication Origin of the *Escherichia coli* Chromosome Binds to Cell Membrane only When Hemimethylated. *Cell*, Vol. 54, pp. 127-135

Ohno, S. (1970). *Evolution by Gene Duplication*. Spring-Verlag, Berlin.

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.*, Vol. 184, pp. 171-186

Otsuka, J. & Nozawa, Y. (1998). Self-Reproducing System can Behave as Maxwell's Demon: Theoretical Illustration under Prebiotic Conditions. *J. Theor. Biol.*, Vol. 194, pp. 205-221

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.*, Vol. 48, pp. 218-235

Otsuka, J.; Kawai, Y. & Sugaya, N. (2001). The Influence of Selection on the Evolutionary Distance Estimated from the Base Changes between Homologous Nucleotide Sequences. *J. Theor. Biol.*, Vol. 213, pp. 129-144

Otsuka, J. (2002). An Inquiry into the Evolutionary History of Photosynthetic and Respiratory Systems from the Similarity Relationships of Member Proteins. In: *Recent Research Developments in Proteins,* Transworld Research Network, (Ed.), Vol. 1, pp. 229-256, Kerala, India

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.*, Vol. 222, pp. 447-460

Otsuka, J. (2004). A Theoretical Characterization of Ecological System by Circular Flow of Materials. *Ecological Complexity*, Vol. 1, pp. 237-252

Otsuka, J. (2005). A Theoretical Scheme for the Large-Scale Evolution of Organisms towards a Higher Order of Organization and Diversity. In: *Recent Research Developments in Experimental & Theoretical Biology*, Transworld Research Network, (Ed.), Vol. 1, pp. 93-122, Kerala, India

Otsuka, J. & Kawai, Y. (2006). Phylogenetical Relationships among Permeases and the Membrane Proteins in Photosynthetic and Respiratory Systems. *Trends in Photochemistry and Photobiology*, Vol. 11, pp. 1-22

Otsuka, J. (2008). A Theoretical Approach to the Large-Scale Evolution of Multicellularity and Cell Differentiation. *J. Theor. Biol.*, Vol. 255, pp.129-136

Pesole, G.; Gissi, C.; Chirico, A. D. & Saccone, C. (1999). Nucleotide Substitution Rate of Mammalian Mitochondrial Genomes. *J. Mol. Evol.*, Vol. 44, pp. 427-434

Rothwell, G. W.; Scheckler, S. E. & Gillespie, W. H. (1989). *Elkinsia* gen. nov., a Late Devonian Gymnosperm with Cupulate Ovules. *Bot Gazette*, Vol. 150, pp. 170-189

Rowe, N. P. (1992). Winged Late Devonian Seeds. *Nature,* Vol. 359, p. 682

Rozanov, A. Y. & Zhuravlev, A. Y. (1992). The Lower Cambrian Fossil Record of the Soviet Union, In: *Origin and Early Evolution of the Metazoa*. J. H. Lipps & P. W. Signor, (Eds.), pp. 205-282, Plenum Press, New York and London

Simpson, G. G. (1944). *Tempo and Mode in Evolution: A Synthesis of Paleontology and Genetics,* Columbia University Press, New York

Stewart, N. S. & Rothwell, G. W. (1993). *Paleobotany and the Evolution of Plants.* pp. 438-467, Cambridge University Press, Cambridge

Sugaya, N. & Otsuka, J. (2002). 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.*, Vol. 55, pp. 584-594

Van den Eynde, H.; De Baere, R.; De Roeck, E.; Van de Peer, Y.; Vandenberghe, A.; Willekens, P. & De Wachter, R. (1988). The 5S Ribosomal RNA Sequences of a Red Algal Rhodoplast and Gymnosperm Chloroplast: Implication for the Evolution of Plastids and Cyanobacteria. *J. Mol. Evol.*, Vol. 27, pp. 126-132

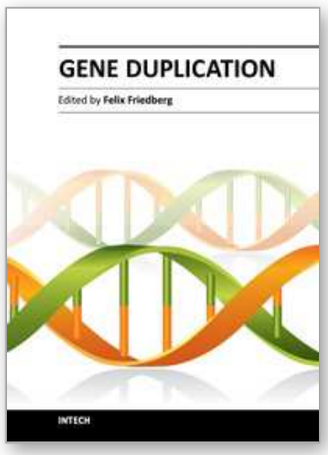
Wheeler, D. L.; Church, D. M.; Edgar, R.; Federhen, S.; Helmberg, W.; Madden, T. L.; Pontius, J. U.; Schuler, G. D.; Schriml, L. M.; Sequeira, E.; Suzek, T. O.; Tatusova, T. A. & Wagner, L. (2004). National Center of Biotechnology Information. *Nucl. Acid Res.*, Vol. 32, Database Issue D35

Wise, D. M. (1988). The Diversity of Mitosis: the Value of Evolutionary Experiments. *Biochem. Cell Biol.*, Vol. 66, pp. 515-529.

Wright, S. (1949). Adaptation and Selection. In: *Genetics, Paleontology and Evolution.* G. L. Jepson; G. G. Simpson & E. Mayer, (Eds.), pp. 365-389, Princeton Univ. Press, Princeton, New Jersey

Yang, D.; Oyaizu, Y.; Oyaizu, H.; Olsen, G. J. & Woese, C. R. (1985). Mitochondrial Origin. *Proc. Natl. Acad. Sci. USA*, Vol. 82, pp. 4443-4447

IntechOpen



Gene Duplication

Edited by Prof. Felix Friedberg

ISBN 978-953-307-387-3
Hard cover, 400 pages
Publisher InTech
Published online 17, October, 2011
Published in print edition October, 2011

The book Gene Duplication consists of 21 chapters divided in 3 parts: General Aspects, A Look at Some Gene Families and Examining Bundles of Genes. The importance of the study of Gene Duplication stems from the realization that the dynamic process of duplication is the "sine qua non" underlying the evolution of all living matter. Genes may be altered before or after the duplication process thereby undergoing neofunctionalization, thus creating in time new organisms which populate the Earth.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jinya Otsuka (2011). A Theoretical Scheme of the Large-Scale Evolution by Generating New Genes from Gene Duplication, Gene Duplication, Prof. Felix Friedberg (Ed.), ISBN: 978-953-307-387-3, InTech, Available from: http://www.intechopen.com/books/gene-duplication/a-theoretical-scheme-of-the-large-scale-evolution-by-generating-new-genes-from-gene-duplication

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen