A differential equation, deduced from a DNA-type genetic algorithm with the lagging-strand-biased mutagenesis

Ichiro Fujihara a,*, Mitsuru Furusawa b

a College of General Education, Osaka Sangyo University, Daito-cho, Osaka 574-8530, Japan
b Chitoise Laboratory Corp., Biotechnology Research Center, 2-13-3 Nagawa-honcho, Miyamae-ku, Kawasaki, Kanagawa 216-0041, Japan

ABSTRACT

Recent evidence indicated that the significant fidelity difference existed between the leading and lagging-strand. Using a 2-D DNA-type genetic algorithm (GA), we have shown that the lagging-stand-biased mutagenesis (disparity mutagenesis) has a significant advantage in the promotion of evolution, compared with the traditional parity mutagenesis model. Our aim of the present study was to deduce a differential equation which well reflected the result of simulations. The analytical solution of the differential equation obtained was in good agreement with the results of the simulation experiment. Comparing the results of the disparity mutagenesis with those of the parity one, the characteristics of the disparity mutagenesis were discussed in terms of relative mutation rates between the lagging and leading strands. Conditions of the extinction of a species were also discussed.

1. Introduction

When any two species are arbitrarily chosen, it is believed that they have evolved from a common ancestral species. The ancestral species, however, had been extinct ages ago in most cases. The memory of the mutations might be vanished away during long evolutionary time. Accordingly, it would be more difficult to ascertain the key mutations that caused this divergence.

Thirty years ago, one of the present authors, MF, thought of the idea that experimental acceleration of evolution using a living organism might reveal the process of species differentiation. To realize the acceleration of evolution, we proposed a specific mutant, “disparity mutator”, in which the lagging-strand-biased mutagenesis was achieved [1, 2]. Using disparity mutators, we could experimentally accelerate evolution in bacteria [3, 4], yeasts [5], malaria parasites [6] and mice [7]. By the way, as a plausible way to simulate the evolutionary process and to approach general optimization problems, J. Holland proposed a genetic algorithm (GA) in 1975 [8]. Subsequently, GA was mainly used for the solution of complex optimization problems. Unexpectedly, however, the contribution of GA to evolutionary genetics was very small. For instance, to our knowledge there has been no review or report on the relationship between GA and biological evolution in the last five years.

Living things evolve as a population (species). The population is formed by the proliferation of each individual member, which is definitely controlled by DNA. However, studies of conventional population genetics have been carried out in the past without paying attention to the mechanism of DNA replication. Instead, average mutation rates per DNA (or cell) replication, and the cell death by deleterious mutations have mostly been considered. As the result, a specific value, “error-threshold”, was worked out. Namely, it was stated that a cell dies when mutation number/cell/replication > 1 [9, 10].

On the contrary, we firstly developed a DNA-type genetic algorithm (GA), which mimicked the molecular structure of DNA and its replication machinery. We simulated the process of evolution using this DNA-type GA [11]. The distinctive characteristic of our study is that the fidelity difference between the leading and lagging strand was considered, based on the evidence of the fidelity difference in living things [12, 13, 14]. Our GA consisted of two complemental strand algorithms as if DNA was replicated. From the results of simulations, here we obtained a differential equation for our disparity mutagenesis model. In the present study, the lagging-stand-biased mutagenesis was applied as a matter of practical convenience.

The effects of the disparity mutagenesis can be summarized as follows. 1) When the fidelity of leading DNA strand is enough high, the genotype of a parent DNA might be transmitted to either daughter DNAs as it is. Therefore, the ancestral genotype is forever transmitted in terms
of a replicon, even when the mutation rates of the lagging strand are high enough. 2) As a result, the error-threshold of the DNA population rises beyond 1.

The GA used in the present study was basically the same one as previously published, which was the 2-D DNA-type GA [15]. This digital DNA takes into consideration gene-to-gene interactions and semi-conservative replication using a leading and lagging strand according to the DNA of living things. The aim of the present study was to deduce a differential equation from the results of simulations using this GA.

Two differential equations were deduced from the conventional parity mutagenesis and from the disparity mutagenesis, respectively. The solutions of these equations agree well with the results of the simulations. We examine the mathematical characteristics of these equations and discuss the biological meanings of the disparity mutagenesis model.

Here we summarize the characteristics of the disparity systems derived from the results of the previous paper.

1. Most importantly, the disparity mutagenesis significantly increased the error-threshold. This evidence would clearly explain the increased adaptability shown by the living disparity mutants.

2. Furthermore, the increase of the error-threshold in the disparity mutagenesis would introduce an evolutionary jump when a drastic environmental change occurs. The parity mutagenesis also could produce an evolutionary jump, but only within a very limited range of mutation rates.

3. In a long stable state of environments, evolution eventually reached a plateau in the disparity mutagenesis due to the high fidelity of the leading strand, even when the average mutation rates were sufficiently high. It might provide us the logic why living fossils exist.

4. In short, the leading strand with a high fidelity guarantees the stability of species. On the other hand, the lagging strand with a low fidelity makes it possible to explore a much broader fitness landscape. After all, although it seems to be contradictory, the disparity mutagenesis realizes the expansion of diversity and the increase of stability of species at the same time.

2. Methods

2.1. Genetic algorithm

In the disparity system, the lagging strand introduces more mutations to promote evolution, but it is also more likely to cause death of a cell. Therefore, the leading strand introduces as few mutations as possible to preserve the species. The rate of evolution and the stability of the species seem to depend on the balance of the number of mutations entering two strands. On the other hand, in the parity system mutations are equally distributed in both strands.

For the simulation of evolution processes, the DNA-type 2-D GA was used, the genome of which consisted of double-stranded DNA, and semi-conservatively replicated using the leading and lagging strand. The genome consisted of fifty functional genes (FC). A single FC gene had three unique fitness landscapes. The two hundred non-functional junk genes (JK) were distributed evenly in number between FC. Two neighbor genes including JK are interacting, that means all genes are interacting. This model GNA is expressed as (FC = 50, JK = 200). The details of the contents of the GA and conditions for simulations were described in the previous study [15].

2.2. Conditions of simulation

In our previous report [15] we regarded DNA as a sequence of connected genes on a two-dimensional information lattice and assumed that neighboring genes interact with each other.

At the initial state of the simulation, the population consists of 100 cells (or genomes) and the coordinates of the FS (fitness score) of each gene are set to “0” on the horizontal axis in every individual cell. This means that from the start of the simulation, the early homogeneous cell population had to survive in a completely new environment.

Fig. 1 shows an example of how genes are arranged on a two-dimensional lattice and how the cell’s alive/dead was determined. In this model, genetic variation is indicated by a shift in the coordinates of the gene. It is assumed that the survival of a cell requires the interaction of all neighboring genes, whether JK or FC genes. If the distance between adjacent genes is less than or equal one, they are regarded interacting. When the gene interaction was interrupted at any position by mutations, the chromosome immediately died. Under these assumptions, we carried out Monte Carlo simulations to investigate how the DNA sequence changes during cell division by repeatedly introducing mutations in the leading and lagging strands of DNA. The main purpose of the calculations was to clarify the difference of FS between the parity and disparity mutagenesis.

For the disparity systems, even if the fidelity of the lagging chain is low, the robustness is sufficient if the fidelity of the leading chain is high. Acceleration or deceleration of evolution can only be brought about by environmental changes.

Fig. 1 illustrates the survival conditions when a gene undergoes mutation on the information lattice and how the gene gets FS scores. The vertical axis represents the order of the genes, and the horizontal axis represents the FS scores for each gene.

In the previous report [15], the simulation was performed under the following conditions.

1) The mutation rate of the lagging strand was an integer and was normally distributed around Pe (mutation rate of lagging strand).
2) The mutation rate of the leading strand was real and 0 ≤ Pe < 1 (Pe; mutation rate of leading strand).
3) Since most mutations are unfavorable to organisms in nature, 80% of mutations moved in the unfavorable direction.
4) If the cell population increases too much, selective pressure (truncation selection) is applied, and the total number of cells was maintained at about 2000.

See the previous report [15] for more details on the model.

In the previous study, it was shown that the effects of PU and Pe on cell growth in the early stages of the simulation were rapid and rather
complicated. Therefore, in the present study, we focused on the early stage of simulation and changed the simulation conditions as follows:

1) The mutation rate of the lagging strand is simply \( P_a \).

3') There is no bias in favor of or against mutation. Left and right movements were made with equal probability.

3. Results and discussion

3.1. Simulations of fitness scores

Fig. 2 shows the generation dependence of FS for the disparity system from \( P_a = 2 \) to 7, when \( P_e = 0 \). First, the initial value of FS is 35 and reaches a plateau value depending on \( P_a \). It is easy to see that the smaller the \( P_a \), the faster the FS rises. As in the previous report, the rate of increase of FS and the highest value of FS were obtained at \( P_a = 2 \). Both the values decreased with increase in \( P_a \). When \( P_e = 0 \) and \( P_a = 2, 3, 4 \) and 5, their final values of FS attained were about the same. However, the plateau value of FS is smaller with higher \( P_a \), as is the rate of increase of FS. Similar observations were obtained in the previous calculations [15].

Figs. 3(a) and 3(b) show the \( P_e \) dependence of FS for \( P_a = 2 \), and \( P_a = 3 \), respectively. As seen in these figures, in the initial stage of generation, if \( P_e \) is small, FS rises very quickly. The final FS becomes highest when \( P_e = 0 \) for \( P_a = 2 \) and \( P_a = 3 \). It has also been shown that as \( P_e \) increases, the final FS and the rate of increase of FS become lower and finally extinct. The cell extinction points, \( P_{e_{ext}} \), were observed at \( P_{e_{ext}} = 0.622 \) for \( P_a = 2 \) and \( P_{e_{ext}} = 0.337 \) for \( P_a = 3 \).

Figs. 4(a) and 4(b) show the \( P_e \) dependence of FS for \( P_a = 4 \), and \( P_a = 5 \), respectively. Comparing Figs. 3(a) and 3(b), it can be seen the rates of increase of FS’s for \( P_a = 4 \) and \( P_a = 5 \) become lower than those for \( P_a = 2 \) and \( P_a = 3 \). The final FS’s also seem lower than those for \( P_a = 2 \) and \( P_a = 3 \). The final FS and the rate of increase in FS are expected to be highest at \( P_e = 0 \), as well as at \( P_a = 2 \) and \( P_a = 3 \). However, as seen in Fig. 4(a), the final FS and the rate of increase in FS at \( P_a = 0 \) were smaller than those at \( P_e = 0.025 \).

In general, the final placement of genes and the speed of rise depends on random numbers employed. Simulations were performed using several different random number sequences under the same conditions of \( P_a = 4 \) and \( P_e = 0 \). As a result, it was found that FS fluctuates about \( \pm 5 \) with different random number sequences. This may be the reason the final FS and the rate of rise in FS at \( P_e = 0 \) is smaller than those at \( P_e = 0.025 \). \( P_{e_{ext}} \), were observed at \( P_{e_{ext}} = 0.191 \) for \( P_a = 4 \) and \( P_{e_{ext}} = 0.060 \) for \( P_a = 5 \). These simulation results suggest the high fidelity of the leading strand is an essential condition for a fast rise and high score of FS.

3.2. Fitness scores of parity systems

Fig. 5 shows the dependence of FS for the parity system, when \( P_e (= P_a) \). As shown in this figure, the rate of increase of FS and the final height of FS became higher as \( P_e \) increased from 0.2. It was in a wide range of \( P_e (= P_a) = 0.3-0.7 \) that the FS was able to rise faster and reach a higher point. In a previous paper [15], we evaluated the stability of cells as a species with Shannon-Wiener ‘H’ The stability of the cells as a species was evaluated in the same way as before. The cells were stable as a species only in a small region where the mutation rate was about \( P_e (= P_a) = 0.7 \). The speed of rise and the highest value of FS began to decrease rapidly when \( P_e (= P_a) \) was beyond 0.8.

4. Derivation of the equation

To quantitatively determine the magnitude of FS and its generational change, the following two factors appear to be particularly important. They are 1) The change in the cell number before and after cell division, \( \Delta P_{s_{ext}} \), and 2) the number of mutations introduced into
the surviving cells after cell division, \( N_m \). To calculate these quantities in parity and disparity systems, it is needed to evaluate the probability of survival of cells from the leading and lagging strands after mutation and the expected values of mutations entering the surviving cells.

Let \( P_d \) be the probability of a cell’s death when a single mutation is introduced. The probability that a cell survives after the insertion of a single mutation is \( (1 - P_d) \). The leading strand of both parity and disparity systems receives a \( P_e(\leq 1.0) \) mutation per cell division. Let \( pr_{lead}^{g_{lead}} \) and \( pr_{lag}^{g_{lag}} \) be the probability that a progeny born from the leading strand will be dead or alive, respectively. They are also expected values of dead and alive cells from a single cell. They are expressed as follows:

\[
pr_{lead}^{g_{lead}} = P_e \cdot P_d
\]

\[
pr_{lag}^{g_{lag}} = 1 - P_e \cdot P_d
\]

In the disparity system, a single cell division introduces \( P_a \) mutations in the lagging strand and \( P_e \) mutations in the leading strand. In the case of leading strands, the probability that a progeny born from the leading strand will be alive or dead, is the same with those for the parity system, as expressed by the equations (1) and (2).

In a single cell division, \( P_a \) mutations are inserted in the lagging strand. Let \( pr_{lag}^{g_{lag}} \) and \( pr_{lead}^{g_{lead}} \) be the probability that a progeny born from the lagging strand will be alive or dead, respectively. They are expressed as follows:

\[
pr_{lag}^{g_{lag}} = (1 - P_d)^{P_a}
\]

\[
pr_{lead}^{g_{lead}} = 1 - (1 - P_d)^{P_a}
\]

### 4.1. Changes in cell number due to cell division, \( \Delta P_{g_{pari}} \), for parity system

In the parity system, a cell’s DNA can be considered to replicate using leading and lagging strands with the same probability of mutation \( P_e(=P_a) \). The lagging strands can, then, be regarded as leading strands in parity system. The increase in the number of cells due to cell division, \( \Delta P_{g_{pari}}^{-} \), is obtained by subtracting 1 from the expected value of the number of living cells from the two leading strands.

\[
\Delta P_{g_{pari}}^{-} = 2pr_{lead}^{g_{lead}} - 1
\]

\[
= 2(1 - P_e \cdot P_d) - 1
\]

\[
= 1 - 2P_e \cdot P_d
\]

After cell division, if two cells are born alive from two leading strands (parity system), then \( \Delta P_{g_{pari}}^{-} = 1 \). Conversely, if no cells remain, then \( \Delta P_{g_{pari}}^{-} = -1 \).

### 4.2. Changes in cell number due to cell division, \( \Delta P_{g_{disip}} \), for disparity system

For the disparity system, the expected number of surviving cells after division is the sum of \( pr_{lag}^{g_{lag}} \) and \( pr_{lead}^{g_{lead}} \), which are represented by equations (2) and (3), respectively. The increase in the number of cells due to cell division, \( \Delta P_{g_{disip}} \), is obtained by subtracting 1 from the expected value of living cells from the leading and lagging strands. It is represented as follows:

\[
\Delta P_{g_{disip}} = pr_{lag}^{g_{lag}} + pr_{lead}^{g_{lead}} - 1
\]

\[
= (1 - P_e \cdot P_d) + (1 - P_d)^{P_a} - 1
\]

\[
= (1 - P_d)^{P_a} - P_e \cdot P_d
\]

The equation (6) indicates that as \( P_a \) increases, the difference in the number of cells before and after cell division, \( \Delta P_{g_{disip}} \), decreases almost exponentially. However, it also indicates that the number of cells is never reduced if \( P_e = 0 \). In the disparity system, the fidelity of the leading strand is particularly important to maintain the cell number.
4.3. Number of mutations introduced during cell division, \( N_m^{pari} \), for parity system

In the leading strand, the probability of mutation is \( Pe \) and the probability of survival is \( (1 - Pd) \). Thus, the average number of mutations introduced into the surviving leading strand, \( N_{m,lead} \), is given by:

\[
N_{m,lead} = Pe \cdot (1 - Pd) 
\]

(7)

Then, in the parity system, the expected number of mutations inserted into the two leading strands of the next generation, \( N_m^{pari} \), is written as follows:

\[
N_m^{pari} = 2N_{m,lead} 
\]

\[
= 2Pe \cdot (1 - Pd) 
\]

(8)

4.4. Number of mutations introduced during cell division, \( N_m^{disp} \), for disparity system

In the disparity system, \( Pa \) mutations are introduced into the lagging strand and the surviving probability is \( (1 - Pd)^2Pe \). Thus, the number of mutations remaining in the surviving lagging strand, \( N_{m,lagg} \), is given by:

\[
N_{m,lagg} = Pd \cdot (1 - Pd)^2Pe 
\]

(9)

Then, in the disparity system, the expected number \( N_m^{disp} \) of mutations inserted into the leading and lagging strands of the next generation can be written by using equations (7) and (9) as follows:

\[
N_m^{disp} = N_{m,lead} + N_{m,lagg} 
\]

\[
= Pe \cdot (1 - Pd) + Pd \cdot (1 - Pd)^2Pe 
\]

(10)

As with \( \Delta P_g \), when selection pressure is applied to the number of cells according to the \( FS \) value, the larger the \( N_m \), the more favorable the increase in \( FS \).

5. Generational change of \( FS \) – through \( \Delta P_g \) and \( N_m \)

5.1. Evolutionary activity coefficient \( A(Pa, Pe) \)

As described in Section 4, the larger \( \Delta P_g \), the more daughter cells there are after cell division and the larger \( N_m \), the more mutations are inserted in the surviving daughter cells. The product of \( \Delta P_g \) and \( N_m \) is an excellent indicator of the rate of evolution of a species and the final \( FS \) value that can be reached.

Then, we assumed that the evolutionary activity coefficients, \( A(\Delta P_g, N_m) \), which expresses the speed of rise of \( FS \), can be expressed as follows:

\[
A(\Delta P_g, N_m) = \Delta P_g \times N_m 
\]

(11)

\( \Delta P_g \) and \( N_m \) for the parity systems are expressed by equations (5) and (6), respectively, and those for the disparity systems are expressed by equation (6) and (10). In the present study, \( Pd \) is a constant which is related to the pattern of the fitness landscape shown in Fig. 1. However, the dependence of \( Pd \) on the landscape pattern has not been tested at present. \( A(\Delta P_g, N_m) \) for the disparity system can be regarded as a function of \( Pe(d = Pa) \). While \( A(\Delta P_g, N_m) \) for the disparity system can be regarded as a function of \( Pa \) and \( Pe \). From here on, instead of \( A(\Delta P_g, N_m) \), we use the notation \( A_{pari}(Pe) \) for parity systems and \( A_{disp}(Pa, Pe) \) for disparity systems. The evolutionary activity coefficient for parity system, \( A_{pari}(Pe) \), is obtained using equations (5) and (6), and is expressed as follows:

\[
A_{pari}(Pe) = (1 - 2Pe \cdot Pd) \cdot [2Pe \cdot (1 - Pd)] 
\]

(12)

And the evolutionary activity coefficient for disparity system, \( A_{disp}(Pa, Pe) \), is obtained using equations (5) and (10), and is expressed as follows:

\[
A_{disp}(Pa, Pe) = [(1 - Pd)^2Pe - Pe \cdot Pd] \cdot [Pe \cdot (1 - Pd)^2Pe + Pe \cdot (1 - Pd)] 
\]

(13)

In a disparity system, the smaller both \( Pa \) and \( Pe \) are, the larger the value of \( A_{disp}(Pa, Pe) \) will be. For example, the sequence of the magnitudes of \( A_{disp}(Pa, Pe) \) is \( A_{disp}(2,0) > A_{disp}(3,0) > A_{disp}(4,0) > A_{disp}(5,0) \) and so on. Fig. 2 shows that the lower the \( Pa \), the faster the \( FS \) rise in a disparity system. Then, we assumed the following relations hold between \( \partial FS/\partial G \) and \( A_{pari}(Pe) \) for the parity systems and \( \partial FS/\partial G \) and \( A_{disp}(Pa, Pe) \) for the disparity systems.

\[
\partial FS/\partial G \propto f(A_{pari}(Pe)) \quad \text{(parity system)} 
\]

\[
\partial FS/\partial G \propto f(A_{disp}(Pa, Pe)) \quad \text{(disparity system)} 
\]

(14)

(15)

where, \( G \) is generation and the functions \( f(A_{pari}(Pe)) \) and \( f(A_{disp}(Pa, Pe)) \) are a monotonically increasing function of \( A_{pari}(Pe) \) and \( A_{disp}(Pa, Pe) \). Equations (14) and (15) imply that the larger \( A_{pari}(Pe) \) and \( A_{disp}(Pa, Pe) \) are, the faster the rise of \( FS \) rises.

\( FS \) is the sum of the fitness scores obtained when 50 functional genes in DNA come to a scoring point on the information lattice as shown in Fig. 1. As can be seen in Fig. 1, the adjacent genes must be linked to each other, so \( FS \) cannot be larger than a certain \( F_{max} \). Then, in addition to equation (14) or equation (15), we must also consider that the final value of \( FS \) will be less than \( F_{max} \).

As shown in Figs. 2–5, when evolutionary stage was early and \( FS \) was low, \( FS \) seemed to move rapidly to higher position. The rate of increase in \( FS \) was assumed to be proportional to the difference between the highest value and the current \( FS \). This situation can be expressed by a differential equation as follows:

\[
\partial FS/\partial G \propto (H_p - FS) 
\]

(16)

where, \( H_p \) is the height of the plateau that the cell population can reach and is dependent on \( Pe \) for parity systems and dependent on \( Pa \) and \( Pe \) for disparity systems.

\[
H_p = H(p(Pa, Pa)) \quad \text{(disparity system)} 
\]

(17)

\[
H_p = H(p(Pe)) \quad \text{(parity system)} 
\]

(18)

However, as shown in Fig. 2, when \( Pa = 2 \) to 5 in the disparity system, the final plateau height was almost the same, if \( Pe = 0 \). Even at \( Pa = 6 \) and \( Pa = 7 \), the \( FS \) still seems to be rising after 150,000 generations. Then, \( H_p \) of the disparity system is assumed to be \( F_{max} \) when \( Pe = 0 \), and is considered to be zero in the extinction limit \( Pe, \text{ext} \).

\( H(p(Pa, Pe)) \) should be closely related to the evolutionary activity coefficient as well as the rate of increase in \( FS \). We assume here that the dependence of \( H(p(Pa, Pe)) \) on \( Pa \) and \( Pe \) is as follows:

\[
H(p(Pa, Pe)) = F_{max} \cdot a_{disp}(Pa, Pe) 
\]

(19)

Similarly, we assume that the dependence of \( H(p(Pe)) \) on \( Pe \) for parity system is assumed as follows:

\[
H(p(Pe)) = F_{max} \cdot a_{pari}(Pe) 
\]

(20)

where \( a_{disp}(Pa, Pe) \) and \( a_{pari}(Pe) \) are the normalized evolutionary activity coefficients expressed as follows:

\[
0 \leq a_{disp}(Pa, Pe) = A_{disp}(Pa, Pe)/\max\{A_{disp}(Pa, Pe)\} \quad \text{for} \quad 0 \leq Pe \leq Pe,\text{ext} \quad \leq 1 
\]

(21)

\[
0 \leq a_{pari}(Pe) = A_{pari}(Pe)/\max\{A_{pari}(Pe)\} \quad \text{for} \quad 0 \leq Pe \leq Pe,\text{ext} \quad \leq 1 
\]

(22)

Fig. 6 shows the dependences of \( a_{pari}(Pe) \) on \( Pe \) for the parity and the dependence of \( a_{disp}(Pa, Pe) \) on \( Pa \) and \( Pe \) for the disparity systems (\( Pa = 2 \) to 5). In the disparity systems, \( a(Pa, Pe) \) is 1 at \( Pe = 0 \).
and decreases monotonously with increase in $P_e$. The negative value of $A_{disp}(P_a, P_e)$ indicates $P_e$ is over the extinction limit, $P_{e, ext}$. On the other hand, in the parity system, $a_{par}(P_e)$ draws like a parabola with respect to $P_e$, and is highest when $P_e$ is between 0.5 and 0.6. These data suggest that the plateau is highest when $P_e$ is zero for the disparity system, and when $P_e$ is between 0.5 to 0.6 for the parity system. If $A_{disp}(P_a, P_e)$ or $a_{par}(P_e)$ is less than zero, the cell population becomes extinct.

5.2. Derivation of differential equation

Considering the equations (15), (16) and (17), the generation dependence of $FS$ of the disparity system can be expressed by the following differential equation.

$$\frac{\partial FS}{\partial G} = C \cdot f(A_{disp}(P_a, P_e)) \cdot [H(P_a, P_e) - FS]$$  \hspace{1cm} (23)

By transforming this equation, we obtained the following separable differential equation.

$$\frac{\partial FS}{\partial G} \left[ \ln[F(H(P_a, P_e)) - \ln(1 - FS)] \right] = C \cdot f(A_{disp}(P_a, P_e))$$  \hspace{1cm} (24)

Solving the equation (24), we obtain the following equation.

$$FS = H(P_a, P_e) \cdot \left[ 1 - \exp\left(-C \cdot f(A_{disp}(P_a, P_e)) \cdot G \right) \right]$$  \hspace{1cm} (25)

where $H(P_a, P_e)$ is given by equation (19). Similarly, the generation dependence of $FS$ of the parity system can be expressed using equations (14), (16) and (18), as follows:

$$FS = H(P_a, P_e) \cdot \left[ 1 - \exp\left(-C \cdot f(A_{par}(P_e)) \cdot G \right) \right]$$  \hspace{1cm} (26)

Next, we consider the specific expressions of $f(A_{disp}(P_a, P_e))$ for the disparity systems and $f(A_{par}(P_e))$ for the parity systems. As mentioned earlier $f(A_{disp}(P_a, P_e))$ and $f(A_{par}(P_e))$ are monotonically increasing functions of $A_{disp}(P_a, P_e)$ and $A_{par}(P_e)$, respectively. First, for simplicity, we assumed the relationships $f(A_{par}(P_e)) = A_{par}(P_e)$ and $f(A_{disp}(P_a, P_e)) = A_{disp}(P_a, P_e)$. Then, for the parity system, using equations (12) and (20), equation (26) becomes as follows:

$$FS = F_{mx} \cdot a_{par}(P_e) \cdot \left( 1 - \exp\left[ -C \cdot \left( 1 - 2P_e - Pd \right) \right] \right) \cdot \left( 2 \cdot P_e \cdot (1 - Pd) \right) \cdot G$$  \hspace{1cm} (27)

For the disparity system, using equations (13) and (19), equation (25) becomes as follows:

$$FS = F_{mx} \cdot a_{disp}(P_a, P_e) \cdot \left( 1 - \exp\left[ -C \cdot \left( 1 - Pd \right) \right] \right) \cdot \left( 2 \cdot P_e \cdot (1 - Pd) \right) \cdot G$$  \hspace{1cm} (28)

6. Comparison of theoretical results and simulation data

As mentioned earlier, the initial value of $FS$ in the simulations was 35 and increased to about 140 in the highest case. In the present calculations using equations (27) and (28), the initial values were set to 0 and the highest plateau values were set to 100. To draw following figures from equation (28) and equation (27), we needed the probability of the cell dying due to a single insertion of mutation, $P_d$. It was set at 0.45 which was evaluated from the relation between $\ln(P_e, ext)$ and $P_a$ which will be explained in the following Section 7. The integral constant $C$ was determined based on the simulation results of the disparity system ($P_a = 2, P_e = 0$) to satisfy that the calculated value of $FS$ is zero at 0th generation and reaches the plateau value of 100 ($F_{mx}$) after 20,000 generations. In this paper, the same integration constant $C$ was used for all the $FS$ calculations for both the disparity and parity systems.

Fig. 7 shows the generation dependence of $FS$ for $P_a$ when $P_e = 0$, calculated using equation (28). Comparing with Fig. 2, the $P_a$ dependence of $FS$ in the disparity system with $P_e = 0$ is well represented by equation (28). When $P_e = 0$, the smaller the value of $P_a$, the faster the $FS$ rises. This is consistent with the results presented in Fig. 2. As we can estimate from equations (6) and (10), the number of cells that survives and the number of mutations introduced are higher, when $P_a = 2$. Both quantities became smaller as $P_a$ increased.

Figs. 8 and 9 show $P_e$ dependences of $FS$ of disparity systems for $P_a = 2$ to $P_a = 5$ calculated from equation (28). As shown in these figures, the $P_e$ dependence of $FS$ in the disparity system also seems to be well represented by equation (28). In the early stages, $FS$ rises rapidly like a logarithmic function: the smaller the $P_e$, the faster the rise in $FS$. In the disparity system, $P_a$ of about 2 would be sufficient for performing rapid evolution under the present conditions. As equation (6) indicates, $\Delta P_e$ decreases rapidly as $P_e$ increases. If the number of cells does not increase sufficiently, the truncation selection pressure does not work effectively and consequently the increase in $FS$ is reduced. When $P_e$ increases at any $P_a$, the cell growth decreases and eventually reaches zero. The negative $\Delta P_e$ in equation (6) indicate the cells become extinct. The graphs in Figs. 8 and 9 illustrate well the characteristics of the graphs in Figs. 3 and 4, respectively.

Fig. 10 shows the generation dependence of $FS$ for the parity system calculated using equation (27). As stated above, in the early stages of the disparity system, both the rate of increase of $FS$ and the magnitude
of \( FS \) decreased as \( Pe \) increased from zero at any \( Pa \). However, in the parity system, as \( Pe = Pa \) increases, the rate of rise in \( FS \) and the height of the plateau increases. Though the curves for \( Pe = 0.5 \) and \( Pe = 0.6 \) are too close to distinguish, both the rate of increase and the magnitude of \( FS \) are maximum around \( Pe = 0.5-0.6 \). Both the rate of increase and the magnitude of \( FS \) begin to decrease beyond \( Pe > 0.6 \). This is because in the parity system, when \( Pe \) is small, \( \Delta P_{g \text{mut}} \) is large but \( N_{m \text{mut}} \) is small. Conversely, when \( Pe \) is large, \( \Delta P_{g \text{mut}} \) is small but \( N_{m \text{mut}} \) is large. Therefore, the behavior of \( FS \) depends on the balance between \( \Delta P_{g \text{mut}} \) and \( N_{m \text{mut}} \). The fast increase rate and large value of \( FS \) were observed as well as in the disparity system. However, they are limited in a narrow region around \( Pe = 0.5 \) to 0.6. This behavior is well represented by the \( Pe \) dependence of \( a_{\text{mut}}(Pe) \) shown in Fig. 6. Comparing Fig. 10 with Fig. 5, we can see that the equation (27) is able to represent the \( Pe \) dependence of \( FS \) of the parity systems as well as the disparity systems. However, for \( Pe = 0.2 \), the calculated values seem to be slightly lower than the simulated values.

7. Cell extinction limit of \( Pe \) of disparity system

Here we also consider the relationship between the extinction of population and the mutation parameters; \( Pe \) and \( Pa \). The probability

\[ FS = \frac{1}{1 + \frac{Pe}{Pa}} \]

where \( Pe \) is the mutation rate of the parents and \( Pa \) is the mutation rate of the children. The relationship between generation and fitness score is shown in Fig. 8(a). The generation dependence of \( FS \) for \( Pe=2 \) is shown in Fig. 8(b). The generation dependence of \( FS \) for \( Pa=5 \) is shown in Fig. 9(a). The generation dependence of \( FS \) for \( Pa=5 \) is shown in Fig. 9(b). The generation dependence of \( FS \) for \( Pa=5 \) is shown in Fig. 10.
other for all ratios of FC and JK. In addition, as shown at the bottom of Table 1, Pd becomes slightly larger when the number of JK genes increases, but the difference is small. This suggests that the number of JK genes between FC genes does not have much influence on the results.

The present treatment shows that Nm, the number of mutations introduced into a cell, and ΔFblock, the changes in cell number due to cell division, play an important role in the evolution of the present genetic algorithm. The probability that a cell will die from the introduction of a single mutation, Pd, plays an important role in this process. If Pe = 0, cells from the leading strands never die. In such case of disparity system, cells can just barely maintain their numbers in the next generation, even if many mutations are introduced at one time into the lagging strand. However, in a parity system, the rate of increase and magnitude in F5 can only be expected within a narrow range of Pe (= Pd).

We conclude that the disparity system is more favorable to evolution than the parity system. Concerning the general implications of asymmetric replication of DNA, see the review [13, 14].

8. Concluding remarks

We proposed here two differential equations for evolution (the parity model and disparity model), based on the molecular mechanism of semiconservative and asymmetrical DNA replication. The biological implications of equation (27) for the parity model and equation (28) for the disparity model were summarized as follows:

1) Evolution is a function of the number of generations: Judging from the present results, it might appear that evolution could be the function of time. This is because a synchronized growth system is used. However, one thing is for sure – evolution is the function of the number of generations.

2) Increase of error threshold: Equation (27) (conventional parity model) led to catastrophe when the error rate of the leading and the lagging strands (Pe = Pa) was beyond a bit 1. However, equation (28) (disparity model) did not lead to catastrophe even when Pd = 15, if Pe was significantly small; meaning that the evolution ability of this population was still maintained.

3) Evolutionary jump: At the start of evolution experiments in the present study, a genetically homogeneous population was exposed to an absolutely novel environment. The fast rise of F5 (fitness score) observed in the parity and disparity models may mean that the environmental shock induces an evolutionary jump. We showed in the previous study that drastic environmental changes artificially introduced in a stable state of evolution induced an evolutionary jump without increasing mutation rates [15]. This jump was followed by the explosive evolution that was reminiscent of the Cambrian explosion. The quick response to the transmissible cancer observed in Tasmanian devil might be explained by assuming a disparity mutagenesis [16].

4) Stable evolution: To maintain an appropriate F5 and to keep low genetic diversity would be indispensable conditions for a stable species. These conditions are satisfied only at very narrow mutation

---

Table 1. The Pe of the cell extinction limit, Pe,ext, and lnPe,ext with different Pa in the disparity systems, which consist of 50 FC genes and 50, 100, 250 and 450 JK genes, respectively. The bottom column of the table shows Pd for each system, calculated from Fig. 11 and equation (30).

| Pa | (FC = 50, JK = 50) | (FC = 50, JK = 100) | (FC = 50, JK = 200) | (FC = 50, JK = 450) |
|----|-------------------|-------------------|-------------------|-------------------|
|    | Pe,ext            | ln(PE,ext)        | Pe,ext            | ln(PE,ext)        |
| 2  | 0.690             | -0.3711           | 0.665             | -0.4080           | 0.622             | -0.4748           | 0.590             | -0.5276           |
| 3  | 0.391             | -0.5990           | 0.370             | -0.9943           | 0.357             | -1.0877           | 0.318             | -1.1457           |
| 4  | 0.224             | -1.4961           | 0.212             | -1.5512           | 0.191             | -1.6555           | 0.174             | -1.7487           |
| 5  | 0.129             | -2.0479           | 0.119             | -2.1286           | 0.102             | -2.2828           | 0.094             | -2.3645           |
| 6  | 0.074             | -2.6037           | 0.071             | -2.6451           | 0.060             | -2.8134           | 0.054             | -2.9188           |
| 7  | 0.046             | -3.0791           | 0.040             | -3.2189           | 0.036             | -3.3242           | 0.031             | -3.4738           |
| 8  | 0.028             | -3.5756           | 0.024             | -3.7297           | 0.021             | -3.8538           | 0.019             | -3.9633           |
| 10 | 0.0102            | -4.5854           | 0.0091            | -4.6995           | 0.0075            | -4.8929           | 0.0072            | -4.9337           |
| Pd | 0.428             | 0.436             |                   |                   |                   |                   |                   | 0.451             | 0.457

Fig. 11. The logarithm of Pe,ext against Pa in disparity systems (FC = 50, JK = 50, 100, 250 and 450). Figure legend: Numerical data shown in Table 1 are plotted. Pe,ext: filled black circles (FC = 50, JK = 50), filled red triangles (FC = 50, JK = 100), filled green squares (FC = 50, JK = 200) and filled purple diamonds (FC = 50, JK = 450). The linear approximation line for each system is represented by the same color.

Pd that a cell will die when a single mutation is introduced is treated as a constant. The effect of differences in the fitness landscape on the FS values has not yet been clarified. The distribution of the fitness points in each FC gene and the number of JK genes between FC genes may affect the speed of F5 rise and the height of the plateau through Pd. In this paper, we confirmed the effect of the number of JK genes on the FS values. The Pe,ext (the extinction limit) in the disparity systems, which consisted of 50 FC genes and different number of JK genes, is shown in Table 1. The logarithm of Pe,ext against Pa is plotted in Fig. 11.

By the way, we have already dealt with this relationship in equation (6). From equation (6), if ΔFblock < 0, the number of cells in the disparity system goes down with increasing generations, so the Pe,ext becomes

Pe,ext · Pd = (1 − Pd)Pe

Taking the logarithm of both sides of equation (29) gives the following equation:

lnPe,ext = Pa · ln(1 − Pd) − ln Pd

From equation (30), we see that the slope of the graph is ln(1 − Pd). The probability of surviving by the insertion of a single mutation into a cell, (1 − Pd), was estimated from the slope of the line with Pa ranging from 2 to 5. In the present system (FC = 50, JK = 200), the value of (1 − Pd) was about 0.55. Therefore, Pd was set to 0.45 in this study. As shown in Fig. 11, the Pa dependence of ln(Pe,ext) is similar to each
5) Pandemic of COVID-19 and disparity mutagenesis: Genome of SARS-CoV-2 is not double-stranded DNA but single-stranded RNA. The all-RNA molecules appearing in an infected cell are + or – stranded RNA. These RNA molecules act as a template when replicating. In other words, a single-stranded RNA is “conservatively” replicated. Therefore, the template molecule must be error-less, but the newly-synthesizing strand is error-prone due to the low fidelity of the viral RNA polymerase. Thus, the mode of the RNA replication can be regarded as a typical disparity mutagenesis. The equation (28), $Pe = 0$, may provide a simple explanation of the strong infectivity of single-stranded RNA viruses including SARS-CoV-2.

Acknowledgements

The authors thank Dr. F. Rueker for his critical reading of the manuscript, and Dr. N. Gouda for his helpful advice and comments.

References

[1] M. Furusawa, H. Doi, Promotion of evolution: disparity in the frequency of strand-specific misreading between the lagging and leading strands enhances dispropor- tionate accumulation of mutations, J. Theor. Biol. 157 (1992) 127–133.
[2] M. Furusawa, H. Doi, Asymmetrical DNA replication promotes evolution: disparity theory of evolution, Genetica 102–103 (1998) 333–347.
[3] T. Iwaki, A. Kawamura, Y. Ishino, K. Kohno, Y. Kano, N. Goshima, M. Yara, M. Furusawa, H. Doi, F. Imamoto, Preferential replication-dependent mutagenesis in the lagging DNA strand in Escherichia coli, Mol. Gen. Genet. 251 (1996) 657–664.
[4] K. Tanabe, T. Kondo, Y. Onodera, M. Furusawa, A conspicuous adaptability to antibiotics in the Escherichia coli mutator strain dnaQ49, FEMS Microbiol. Lett. 176 (1999) 191–196.
[5] C. Shimoda, A. Itadani, A. Sugino, M. Furusawa, Isolation of thermos-tolerant mu-tants by using proofreading-deficient DNA polymerase $\delta$ as an effective mutator in Saccharomyces cerevisiae, Genes Genet. Syst. 81 (2006) 391–397.
[6] H. Horuma, M. Hirai, S. Nakamura, H. Hikimi, S. Kawanu, N. Palacpac, H. Hinoeda, H. Matsuda, S. Kawai, H. Endo, T. Yanoaga, J. Ohishi, T. Mita, T. Hori, M. Furusawa, K. Tanabe, Generation of rodent malaria parasites with a high mutation rate by destructing proofreading activity of DNA polymerase $\delta$, DNA Res. 21 (2014) 439–446.
[7] A. Uchimura, M. Higuchi, Y. Minakushi, M. Ohno, A. Toyoda, A. Fujiyama, I. Miura, S. Wakana, J. Nishino, T. Yagi, Germline mutation rates and the long-term pheno-typic effects of mutation accumulation in wild-type laboratory mice and mutator mice, Genome Res. 25 (2015) 1–10.
[8] J. Holland, Adaptation in Natural and Artificial Systems, University Michigan Press, Ann Arbor, MI, 1975.
[9] M. Eigen, J. McCaskill, P. Schuster, The molecular quasispecies, Adv. Chem. Phys. 75 (1989) 149–263.
[10] M.A. Nowak, Evolutionary Dynamics: Exploring the Equations of Life, President and Fellows of Harvard College, Cambridge, 2006.
[11] K. Wada, H. Doi, S. Tanaka, Y. Wada, M. Furusawa, A neo-Darwinian algorithm: asymmetrical mutations due to semiconservative DNA-type replication promote evolu-tion, Proc. Natl. Acad. Sci. USA 90 (1993) 11934–11938.
[12] J. Snedeker, M. Wooten, M. Xin, X. Chen, The inherent asymmetry of DNA replica-tion, Annu. Rev. Cell Dev. Biol. 33 (1) (2017) 291–318.
[13] M. Furusawa, Implications of fidelity difference between the leading and the lagging strand of DNA for the acceleration of evolution, Front. Oncol. 2 (2012) 144.
[14] B. Sankar, M. Wooten, X. Chen, The nature of mutations induced by replication–transcription collisions, Nature 535 (2016) 178–181.
[15] I. Fujihara, M. Furusawa, Disparity mutagenesis model possesses the ability to real-ize both stable and rapid evolution in response to changing environments without altering mutation rates, Heliyon 2 (8) (2016) e0014.
[16] P. Hohenlohe, A. Storfer, Rapid evolutionary response to a transmissible cancer in Tasmanian devils, Nat. Commun. 7 (2016) 12684.

Declarations

Author contribution statement

Ichiro Fujihara: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper. Mitsuru Furusawa: Analyzed and interpreted the data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.