Genet. Res., Camb. (1983), **41**, *pp.* 1–15
With 3 text-figures
Printed in Great Britain

Theoretical study on the accumulation of selfish DNA*

By TOMOKO OHTA

National Institute of Genetics, Mishima, 411, Japan

(*Received 4 March 1981 and in revised form 6 July 1982*)

SUMMARY

The accumulation of selfish DNA in eukaryotic genomes was studied from the standpoint of population genetics. Selfish DNA is assumed to replicate itself within a haploid set. For the selectively neutral case, the fate of a single self-replicating DNA segment (unit) within a population was investigated by the method of the probability generating function, and by Monte Carlo simulation, with special reference to the probability of survival and average number of units per haploid set. For the selectively deleterious case at the organismal level, the equilibrium between new occurrence and selective elimination was studied, and the average and variance of the number of units per haploid set in the population was examined by Monte Carlo simulation. It is shown that the process of self-replication (duplication–deletion) plays an essential role for the maintenance and elimination of selfish DNA.

1. INTRODUCTION

Through the remarkable progress of molecular biology, the functional organization of the genetic material of higher organisms is being elucidated. Highly and moderately repetitive sequences may be considered to be 'selfish', in that they spread by forming additional copies of themselves within the genome, even if they are not useful to the organism (Doolittle & Sapienza, 1980; Orgel & Crick, 1980). At least two types of such DNA exist in the genomes of higher organisms; tandemly and dispersed repetitive sequences. Mechanisms of forming additional copies of themselves (replication) within the genome should be different between the tandemly and dispersed repeating sequences. The population genetics of repeating DNA sequences has been developed for the cases where such DNA has already been established in the population, by analysing how the number of repeating units per genome changes (Ohta & Kimura, 1981; Ohta, 1981). As pointed out by Orgel & Crick (1980), however, the process of spread into a population needs to be investigated.

In our previous reports (Ohta & Kimura, 1981; Ohta, 1981), selfish DNA is defined as those sequences which change in numbers by duplication–deletion process but with no directed increase. Dover & Doolittle (1980) define this in a more

* Contribution no. 1427 from the National Institute of Genetics, Mishima, Shizouka-ken, 411, Japan.

specific way; selfish DNA replicates itself faster than other DNA by a sequence specific mechanism, and its amount is expected to increase on the average. They further define 'ignorant' DNA sequences which have equal chance of duplication and deletion, through a process like unequal crossing-over. Dover (1982) has defined a process of fixation called 'molecular drive' encompassing both the selfish and ignorant mechanisms of change.

In the present report, accumulation of selfish DNA is investigated from the following two standpoints: (1) accumulation is completely selectively neutral at the organismal level, and (2) it is slightly deleterious.

For the selectively neutral case, the behaviour of a self-replicating DNA unit newly introduced into a population is investigated. Such units, i.e. self-replicating DNA segments, may be transposons, that are widely found in both procaryotes and in eukaryotes, or may be tandemly repeated DNA in eukaryotes. A unit may have more chance to replicate than that of deletion (truly selfish), or it may have equal chance of duplication and deletion (ignorant). Once such DNA spreads in sufficient amount in the haploid sets of a species, the previous formulation (Ohta & Kimura, 1981; Ohta, 1981) may be applicable for understanding the change of its amount. Therefore, in this study, the process until DNA segments spread in sufficient amount, will be investigated.

For the slightly deleterious case, it is supposed that the accumulation of selfish DNA is a burden for the organism (Orgel & Crick, 1980). However, the intensity of natural selection against the accumulation is expected to be quite small. In the present analyses, it is assumed that a selfish DNA segment appears in the population at a constant rate, and its amount is kept in balance between occurrence and selective elimination. The situation would correspond to the accumulation of pseudogenes discovered in members of many multigene families (Fedoroff, 1971; Nishioka & Leder, 1980; Vanin *et al.* 1980; Van Arsdell *et al.* 1981; Hollis *et al.* 1982), but the analysis does not deal directly with truly selfish DNA (Dover & Doolittle, 1980) which has more chance to replicate than that of deletion. Future studies are needed for such cases.

2. SELECTIVELY NEUTRAL CASE

In this section, the dispersed type of DNA is mainly considered; however, the result may also be applicable to the clustered type so long as the number of units per haploid set is small. A theoretical approach is possible only for a single genomic line, but a very rough analysis that extends to the level of population will be presented. Simulation studies will also be included. In higher organisms, the exchange of DNA between haploid sets is an important process, however, it is most difficult to analyse, and simulation studies are performed to study the effect. For the clustered type, the rate of exchange is small whereas the dispersed type would be freely recombined, therefore various levels of exchange rate are examined by an approximation procedure.

First, a single haploid line, initially with one unit, is considered, and population

dynamics will be considered later. Assume that, in one generation, a unit has the probability γ_0 of being deleted, and γ_2 of forming another copy of itself somewhere in the genome independent of the other. The model is intended for treating a duplicative transposition event and may not be appropriate for other mechanisms of generation of pseudogenes (Van Arsdell *et al.* 1981; Hollis *et al.* 1982). The method of probability generating functions (see Feller, 1957) will be used. The generating function for the number of descendant copies in a single haploid line may be expressed by

$$A(s) = \gamma_0 + (1-\gamma_0-\gamma_2)\, s + \gamma_2\, s^2, \tag{1}$$

where the coefficient of s^n ($n = 0, 1, 2, \ldots$) represents the probability of a unit leaving n descendant copies. Let q_t be the probability that the unit becomes extinct by the tth generation. Then the recurrence equation of q_t may be expressed as follows (Feller, 1957, page 275):

$$q_t = q_{t-1} + (\gamma_0 - \gamma_2\, q_{t-1})\,(1-q_{t-1}). \tag{2}$$

First, consider the simple case of $\gamma_0 = \gamma_2 = \gamma$, i.e. the unit has equal probabilities of being deleted or duplicated in the haploid set. Under the assumption that $\gamma \ll 1$, the recurrence relation may be approximated by the following differential equation:

$$\frac{dq_t}{dt} = \gamma(1-q_t)^2. \tag{3}$$

By solving this formula with the condition that $q_0 = 0$, one gets the following equation:

$$q_t = \frac{\gamma t}{1+\gamma t}. \tag{4}$$

In other words, the unit will be eventually lost from the haploid set, since $q_t \to 1$ when $t \to \infty$.

Next, consider a more interesting case where the unit has more chance of duplication than that of deletion, i.e. $\gamma_2 > \gamma_0$. Then the differential equation corresponding to the formula (3) becomes,

$$\frac{dq_t}{dt} = (\gamma_0 - \gamma_2\, q_t)\,(1-q_t). \tag{5}$$

Solving this equation with the same initial condition as before, the extinction probability becomes,

$$q_t = \frac{e^{(\gamma_2-\gamma_0)\,t} - 1}{(\gamma_2/\gamma_0)\, e^{(\gamma_2-\gamma_0)\,t} - 1}. \tag{6}$$

The above formula demonstrates that the extinction probability eventually increases and approaches γ_0/γ_2 as t gets large. The ultimate survival probability is $1-\gamma_0/\gamma_2$. A less interesting situation would be the case $\gamma_0 > \gamma_2$, i.e. the unit has more chance of extinction than that of duplication. The formula (6) is applicable even in such cases, however, the unit is quickly lost from the haploid set, and the ultimate survival probability is zero.

The above theory applies to a single haploid line, and we shall next consider the

population dynamics. Let us assume a randomly mating population consisting of N breeding individuals, and ask the following questions: (1) what is the probability of spreading a unit newly introduced into one haploid set of the population? (2) when the unit spreads into the population and the majority of the haploid sets contain the units, how many gene copies are contained in one genome on the average? At this moment, exact answers to the above questions cannot be obtained, however approximate analyses provide rough estimates, which may be useful for understanding the evolution of repetitive DNA. For simplifying the treatment, I shall, for the moment, assume that neither jumping to a different haploid set nor exchange of units between haploid sets occur. The effect of such exchange between the haploid sets will be considered later.

When the haploid sets containing units are selectively neutral, the original genomic line with the initial copy would have the probability, $1/2N$, of spreading into the population, and it would take $4N$ generations until fixation (Kimura & Ohta, 1969). Then, roughly speaking, the probability of spreading of units into the population is expected to be as follows at $t = 4N$:

$$u_{4N} = (1-q_{4N})/(2N), \tag{7}$$

where q_{4N} is given by formula (6) by putting $t = 4N$. When $\gamma_2 > \gamma_0$, the ultimate survival probability is interesting. It becomes, with no recombination between the genomes,

$$u_\infty = \frac{1-q_\infty}{2N} = \frac{\gamma_2-\gamma_0}{2N\gamma_2}. \tag{7a}$$

When the units are freely recombined and N is large such that $2N(\gamma_2-\gamma_0) \gg 1$, it would approach,

$$u'_\infty \approx 2(\gamma_2-\gamma_0). \tag{7b}$$

This is because each unit may be treated as a selectively advantageous mutant with coefficient, $\gamma_2-\gamma_0$. Thus, recombination may greatly increase the survival probability when $\gamma_2 > \gamma_0$.

The average number of units per haploid set when units spread in the population is obtained from the *unconditional* mean number of units (i.e. including the case of loss), which becomes (Feller, 1957),

$$m_t = e^{(\gamma_2-\gamma_0)t}. \tag{8}$$

Then the average number when the unit spreads in the population is,

$$\mu_{4N} = \frac{m_{4N}}{2N} \times \frac{1}{u_{4N}} = \frac{e^{4N(\gamma_2-\gamma_0)}}{(1-q_{4N})}. \tag{9}$$

In order to test the reliability and limitations of the above analyses, Monte Carlo simulations were performed. Also examined are the effects of unit exchange between the haploid sets. Unit exchange may occur by two mechanisms; transposition of units to the other haploid set in diploid cells, and unit segregation at meiosis. To simplify the treatment, however, I assume that the transposition always occurs to the same genome and therefore all exchanges are through

segregation at meiosis. The Monte Carlo experiments consist of a series of inputs of one unit followed by its subsequent increase or decrease. The fate of each unit is followed until it is lost from the population or until all genomes contain at least one unit. One generation consists of the following processes: duplication or deletion within the haploid set, exchange of units between the haploid sets and random sampling drift. By means of random numbers, duplication or deletion of units occurs following the assigned probability, and the number of units in each genome

Table 1. *Results of Monte Carlo simulations on the relative fixation probability (r.p.) and the mean number of gene units when spread ($\bar{n}$) for the cases of $\gamma_0 = \gamma_2 = \gamma$*

(Their expected values (equations 7 and 9) are also given for comparison. r.p. is taken relative to that of a neutral mutant in the population.)

	$N = 25$		$N = 50$		$N = 100$		Expected	
$N\gamma$	$\bar{n}$	r.p.	$\bar{n}$	r.p.	$\bar{n}$	r.p.	$E(n) = \mu_{4N}$	$E(\text{r.p.}) = 2Nu_{4N}$
2·0	5·13	0·04	8·24	0·08	11·76	0·11	9·0	0·11
1·0	5·46	0·16	6·03	0·14	8·03	0·14	5·0	0·20
0·5	3·49	0·30	4·74	0·23	3·67	0·21	3·0	0·33
0·25	2·24	0·40	2·47	0·41	2·29	0·43	2·0	0·50
0·125	1·58	0·60	1·59	0·62	1·48	0·49	1·5	0·67
0·0625	1·31	0·78	1·28	0·72	1·25	0·82	1·25	0·80

is scored. As to the unit exchange process, a simple method is applied, i.e. each of the units of two randomly chosen haploid sets is distributed randomly to the two daughter sets. By this method, unit exchange is maximized, since if the unit in one haploid set happens to have the partner unit at the homologous position of the other haploid set, the two units would be distributed one–one into the daughter sets. Random distribution is expected when the unit is hemizygous in the two sets. For an exact assessment, one needs more detailed Monte Carlo studies in which not only the number of units but also chromosomal location are recorded. At this moment, however, these approximations give useful information on the effect of gene exchange.

Each Monte Carlo experiment is continued until the total number of losses of the introduced units becomes 25000. The number of cases where units spread into the population, the number of generations required until spreading, and the mean number of units per haploid set when spread, are recorded. The results are shown in Tables 1–3. The proportion of cases where the units spread is calculated and divided by $1/(2N)$ to obtain relative probability of fixation to an ordinary selectively neutral mutant, and the relative probability is denoted by r.p. in the table. Its expectation is $2Nu_{4N}$. The number of generations until spreading is denoted by T, and the mean number of units per haploid set when spread, by $\bar{n}$. The first two tables give the comparison of the results of simulation and theoretical predictions (equations 7 and 9), whereas the third table show the effect of gene exchange between the haploid sets. The parameters are $N = 25$, 50 and 100 with $N\gamma_0 = N\gamma_2 = N\gamma$ in the range of 2·0–0·0625 (Table 1), and $N = 50$ and 100 with

https://doi.org/10.1017/S0016672300021029 Published online by Cambridge University Press

$\gamma_0 = 0{\cdot}001$–$0{\cdot}01$ and $\gamma_2 = 0{\cdot}01$–$0{\cdot}02$ (Table 2). As can be seen from the tables, the agreements between the expected and observed values are fairly good. Note that the expected number of generations is $4N$, which almost agrees with the observed numbers given in Table 2.

Table 2. *Results of simulations on the relative fixation probability (r.p.) and the mean number of gene units when spread ($\bar{n}$) for the cases of $\gamma_0 < \gamma_2$*

(Their expected values are given for comparison. Observed average number of generations (T) until every genome of the population contains at least one unit is also given. Note that its expectation is $4N$.)

		Observed			Expected	
γ_0	γ_2	$\bar{n}$	r.p.	T*	$E(n) = \mu_{4N}$	$E(\text{r.p.}) = 2Nu_{4N}$
			$N = 50$			
0·01	0·02	16·58	0·50	235	13·78	0·54
0·002	0·01	9·66	0·91	207	5·94	0·83
0·001	0·01	9·18	0·90	195	6·61	0·92
			$N = 100$			
0·005	0·015	38·44	0·58	399	81·40	0·67
0·005	0·01	18·62	0·59	409	13·78	0·54
0·002	0·01	34·65	0·72	377	30·42	0·81
0·001	0·01	67·61	0·83	409	40·55	0·90

* $E(T) = 4N$.

The cases with unit exchange are more complicated, and the present results provide only a rough estimation of the effect. Four cases, as given at the top line of Table 3, were examined. The exchange rate is defined as the probability that each unit is randomly distributed to the two haploid sets in one generation and is denoted by β. In practice, the units contained in two randomly chosen haploid sets are distributed randomly to the two daughter sets and the process is repeated $N\beta$ times each generation. Eight values of β were tried. As can be seen from Table 3, the effect of gene exchange is to increase fixation probability and to decrease the number of units per haploid set at the time of spreading, although their product seems to be little influenced by the exchange process. The effect becomes larger when the mean number of units increases, as expected. Beyond a certain value of β, however, the fixation probability does not become larger by increasing the exchange rate. Furthermore, note that the number of generations until spreading is only slightly influenced by unit exchange.

Although the detailed quantitative assessment awaits future studies, let us briefly consider the underlying theory to explain the above observations. In the previous analysis, a haploid line which happens to spread into the population and survival of units within this line were considered. When the units are exchanged between a particular haploid line and one of the other haploid sets in the population, the units, which are transferred to the outside of the line, are likely to be lost from the population. Hence the average number of units per haploid set

Table 3. *Results of Monte Carlo simulations performed to study the effect of gene exchange between the haploid sets on the probability of spreading and the mean number of units per genome*

(N is 50 or 100, and four sets of parameter values, each with eight levels of the exchange rate (β) were examined.)

	$\gamma_0 = \gamma_2 = 0{\cdot}01$, $N = 50$			$\gamma_0 = \gamma_2 = 0{\cdot}04$, $N = 50$			$\gamma_0 = 0{\cdot}01$, $\gamma_2 = 0{\cdot}02$, $N = 50$			$\gamma_0 = 0{\cdot}005$, $\gamma_2 = 0{\cdot}015$, $N = 100$		
β	$\bar{n}$	r.p.	T	$\bar{n}$	r.p.	T	$\bar{n}$	r.p.	T	$\bar{n}$	r.p.	T
0·0	4·74	0·23	246	8·24	0·08	258	16·58	0·50	235	38·44	0·58	399
0·04	3·55	0·25	287	6·27	0·15	254	5·98	1·01	222	6·90	2·08	343
0·08	3·39	0·32	304	5·35	0·16	291	4·78	1·24	215	5·57	2·73	330
0·12	3·32	0·29	330	4·79	0·18	261	4·31	1·47	221	5·01	3·21	331
0·16	3·37	0·36	311	4·49	0·23	285	4·24	1·44	225	4·85	3·12	323
0·2	3·58	0·27	345	4·30	0·22	297	4·07	1·64	222	4·80	3·06	334
0·4	3·50	0·29	352	4·01	0·24	299	4·01	1·54	225	4·63	3·76	325
0·5	3·75	0·27	350	3·96	0·30	344	4·00	1·84	244	4·58	3·55	340

is reduced by the exchange process. Next, consider those haploid lines which happen to be going extinct from the population. When the units are transferred from such lines to one of the other haploid sets in the population, the transferred units may have a non-zero chance to spread into the population. Thus, the probability of spreading increases by the exchange process. Note that the relative fixation probability may be more than one when $\beta > 0$ and $\gamma_2 > \gamma_0$ as in the last column of the table. In other words, when a unit has more chance of duplication than of deletion and a positive probability of being transferred into the other genome, it may be easily established in the population. Such a tendency is expected to be more pronounced in large populations as expected from equations (7*a*) and (7*b*).

The above result would give a basis for understanding how often a new type of selfish DNA may spread in the course of evolution. Once it spreads in the total population, the change of its amount may be treated by the previous theory (Ohta & Kimura, 1981; Ohta, 1981). Theoretically speaking, it will be eventually lost from the population or its amount increases more and more with smaller probability. Such a situation is unrealistic, and in the next section, the model is investigated where the accumulation of DNA has a deleterious effect on the organism.

3. SLIGHTLY DELETERIOUS CASE

In this section, I shall investigate the situation where continued occurrence of selfish DNA is balanced by selective elimination of individuals with excess amount of such DNA. As before, selfish DNA consists of replicating units, and let us assume that a haploid set with n_i units of selfish DNA has selective disadvantage sn_i, i.e. the fitness of the haploid set is $1-sn_i$. The analyses are mainly concerned with higher organisms, and with the equilibrium properties of the amount of selfish DNA between the occurrence of new units in the population and selective elimination of such DNA. However, the occurrence of new units may have a similar effect on the amount as the difference between the rate of duplication and that of deletion, i.e. the rate of occurrence per genome equals $(\gamma_2-\gamma_0)\bar{n}$, where $\bar{n}$ is the average number of units per haploid set, provided that there is an equilibrium. The problem of the existence of equilibrium is not simple and needs further study as the previous works on meiotic drive (Hartl, 1974; Prout & Bundgaard, 1976; Crow, 1979) indicate. Various mechanisms may be responsible for the occurrence of new selfish DNA; insertion of a unit into the genome from another source (Van Arsdell *et al.* 1981; Lueders *et al.* 1982; Hollis *et al.* 1982), deterioration of members of a multigene family (Fedoroff, 1979), and transposition of gene members from a clustered gene family, i.e. 'orphons' (Childs *et al.* 1981).

The procedure of the simulation experiment is as follows. In each generation, selfish DNA is assumed to undergo the duplication–deletion process and exchange of DNA takes place between the haploid sets. As in the previous studies (Ohta & Kimura, 1981; Ohta, 1981), the clustered and dispersed types are treated separately. Although an analytical approach such as that of Takahata (1981) may

be possible if selection is strong, the present analyses resort to Monte Carlo simulations because selection is very weak and also because finite population size may have some effect.

Three sets of Monte Carlo experiments were performed to study (i) the effect of duplication–deletion (self-replication) for the clustered families, (ii) the effect of the rate of occurrence of selfish DNA for the clustered and the dispersed classes,

Table 4. *Parameters used in simulation studies, all measured per generation*

Occurrence of new unit	p per genome
Selective disadvantage	s per unit
Duplication–deletion of dispersed class	α_1 per unit
Unequal crossing-over of clustered class	α_2 per unit
Inter-chromosomal recombination of clustered class	β per cluster With $\beta = \beta_I + \beta'_I$*

* See Fig. 3.

and (iii) the effect of natural selection for both classes. In all cases, the mean and variance of the number of repeating units per haploid set are recorded and are given in the following. All experiments start from a population with no selfish DNA, and in each generation, selfish DNA is newly introduced at a constant rate. Eventually, the amount of selfish DNA reaches an equilibrium value due to new occurrence and selective elimination. Table 4 gives the parameters used in the experiments, and the details of the experimental procedures are given in the Appendix.

The results of Monte Carlo experiments are explained below. Fig. 1 shows the effect of the duplication–deletion process on the clustered type. The abscissa is the rate of unequal crossing-over per one unit per generation (α_2), and the ordinate is the average number of units per haploid set ($\bar{n}$). Other parameters are, $N = 250$, $s = 0{\cdot}0002$, $\beta = \beta_I + \beta'_I = 0{\cdot}01$ and $p = 0{\cdot}04$ ($2Np = 20$) or $0{\cdot}004$ ($2Np = 2$). All experiments start from a population with no selfish DNA, and the observed values of $\bar{n}$ is the average over the period of 251st $\sim$ 5250th generations. Since each experiment starts from a population free of selfish DNA, it takes some time to accumulate units. In the initial 250 generations, both the average and the variance of the number of units are smaller than those in the following period used for recording the data. This period may not represent complete equilibrium, however, it is considered to be close to equilibrium.

From the figure, it is very clear that the duplication–deletion process is quite effective in making the amount of selfish DNA smaller. This is because the variance of the number of units per haploid set within the population increases through this process, and selection becomes effective as the variance increases. Thus, it seems that the duplication–deletion process is essential in eliminating selfish DNA by natural selection. In this regard, let us examine the variance of n_i relative to the mean, i.e. the ratio $\sigma_n^2/\bar{n}$ in the above experiments. When $2Np = 20$, the ratio is $3{\cdot}86$ for $\alpha_2 = 0$, but increases rapidly and becomes $30{\cdot}81$ for $\alpha_2 = 0{\cdot}004$. When $2Np = 2$, it increases from $1{\cdot}55$ for $\alpha_2 = 0$ to $11{\cdot}23$ for $\alpha_2 = 0{\cdot}004$. Therefore, it is clear that the variance of n_i greatly increases by duplication–deletion process. It

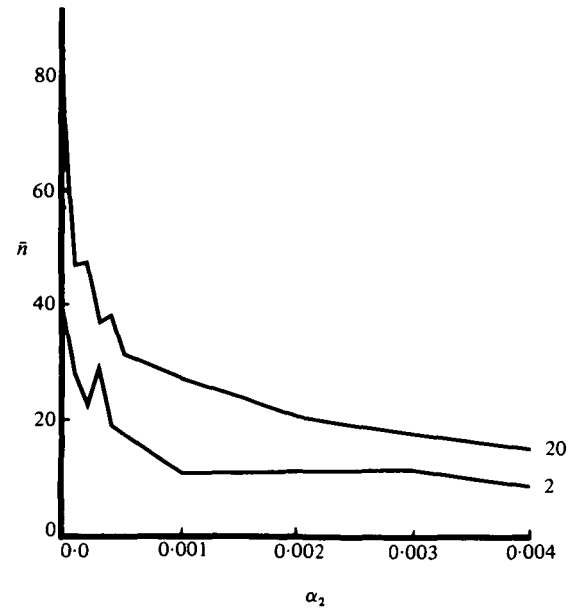

Fig. 1. Results of simulations on the effect of duplication–deletion process (α_2) on the amount ($\bar{n}$) of selfish DNA (clustered type) at equilibrium between new occurrence and natural selection. Parameters are, $N = 250$, $s = 0{\cdot}0002$, $\beta = 0{\cdot}01$ and $2Np = 20$ or 2 as given at the left of the figure.

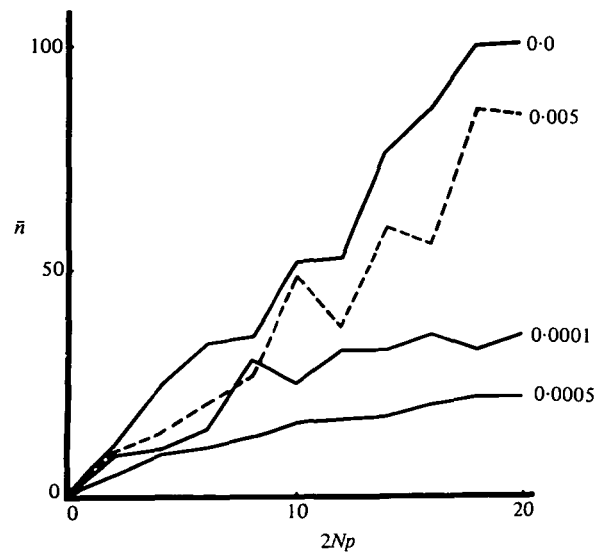

Fig. 2. Results of simulations on the effect of the rate of occurrence of new unit (p) on the amount of selfish DNA ($\bar{n}$). Solid line is for the clustered type, and the broken line, for the dispersed type. Parameters are, $N = 250$, $s = 0{\cdot}0002$ and $\beta = 0$ for the clustered type, $\alpha_2 = 0{\cdot}0$, $0{\cdot}0001$ or $0{\cdot}0005$, and $\alpha_1 = 0{\cdot}005$ as given in the figure.

https://doi.org/10.1017/S0016672300021029 Published online by Cambridge University Press

should also be noted that, unlike the selectively neutral case (Ohta & Kimura, 1981), the relationship of $\bar{n}$ and σ_n^2 is difficult to obtain analytically when selection is involved as in the present model. In addition to the experiments given in the figure, the experiments were carried out with various rates of recombination between the haploid sets ($\beta \geqslant 0$). However, it was found that the recombination process has only a small effect on $\bar{n}$ as long as $\beta \ll 1$. The condition is usually satisfied for the clustered families.

Table 5. *Results of simulations on the mean and standard error of the number of units per haploid set in the population, when new occurrence of selfish DNA is balanced by selective elimination*

(Parameters are, $N = 250$, $p = 0{\cdot}04$, and $\beta = 0$ for the clustered type, and free recombination for the dispersed type.)

		Dispersed		Expectation under
s	Clustered ($\alpha_2 = 0{\cdot}001$)	$\alpha_1 = 0{\cdot}005$	$\alpha_1 = 0{\cdot}02$	random assortment
0·0001	14·54 ± 7·22	91·08 ± 9·57	92·81 ± 9·65	400·00
0·0002	16·35 ± 7·96	77·53 ± 8·80	77·26 ± 8·81	200·00
0·0003	16·07 ± 7·86	88·09 ± 9·41	75·04 ± 8·67	133·33
0·0004	13·45 ± 6·49	48·33 ± 6·96	62·12 ± 7·91	100·00
0·0005	13·03 ± 5·89	52·22 ± 7·23	56·75 ± 7·55	80·00
0·001	11·88 ± 6·34	33·77 ± 5·82	30·79 ± 5·59	40·00
0·002	9·20 ± 4·37	17·34 ± 4·05	19·01 ± 4·44	20·00
0·003	8·76 ± 3·75	12·32 ± 3·52	12·14 ± 3·55	13·33
0·004	7·24 ± 3·08	10·03 ± 3·17	9·23 ± 3·09	10·00
0·005	6·43 ± 2·82	7·82 ± 2·80	7·30 ± 2·75	8·00

Fig. 2 shows the effect of rate of occurrence (p) on $\bar{n}$. The abscissa is $2Np$ and the ordinate is $\bar{n}$. Both clustered (solid line) and dispersed types (broken line) were studied. Three levels of α_2 (0·0, 0·0001 and 0·0005) and one level of α_1 (0·005) were examined. Other parameters are, $N = 250$, $s = 0{\cdot}0002$, and $\beta = 0$ for the clustered type and free recombination for the dispersed type. The observed values are again the averages for the period between 251st ~ 5250th generations starting from a population free of selfish DNA.

The figure shows that $\bar{n}$ seems to increase linearly with $2Np$ for the dispersed class whereas it does so only when $\alpha_2 = 0$ for the clustered class. When $\alpha_2 > 0$, $\bar{n}$ does not increase linearly with $2Np$ for the clustered type. Again this is because, the variance of n_i increases for larger α_2, and selective elimination becomes more efficient. Note that selection does not operate on an individual unit but on the total *number* of units in each haploid set. Let us again examine the ratio, $\sigma_n^2/\bar{n}$. It is about unity for the dispersed class. This suggests Poisson distribution of n_i. For the clustered class, the ratio is smaller than unity when $\alpha_2 = 0$, but larger than one for other cases studied. When $\alpha_2 = 0$, the ratio is 0·11 ~ 0·37 with no tendency of increase or decrease with change of $2Np$. When $\alpha_2 = 0{\cdot}0001$, the ratio is 1·52 for $2Np = 2$, gradually increases, and becomes 5·99 for $2Np = 20$. When $\alpha_2 = 0{\cdot}0005$, it increases from 1·71 for $2Np = 2$ to 9·78 for $2Np = 20$.

Simulation experiments were also performed to examine the effect of intensity of natural selection. The clustered class ($\beta = 0, \alpha_2 = 0{\cdot}001$) and the dispersed class ($\alpha_1 = 0{\cdot}005$ or $0{\cdot}02$) were studied with $N = 250$ and $p = 0{\cdot}04$. The selection coefficient (s) are varied from $0{\cdot}0001$ to $0{\cdot}005$. Table 5 gives the observed values of $\bar{n}$ again for the period of 251st ~ 5250th generations, starting from a population free of units. The last column of the table gives the expected $\bar{n}$ at equilibrium under random assortment of units among the haploid sets. That is,

$$E(\bar{n}) = p/s. \tag{10}$$

From the table, it can be seen that, when selection is strong enough, the observed and the expected values roughly agree. As the selection becomes weaker, the disagreement becomes larger. Also the disagreement is larger for the clustered type than for the dispersed type. Again this is caused by larger variance of n_i relative to $\bar{n}$ of the clustered type than that of the dispersed type. In addition, it is interesting to note that the magnitude of α_1 has little effect on $\bar{n}$ for the dispersed type. The reason would be that, due to free recombination of this class, the random assortment of units is attained irrespective of the duplication–deletion process. In fact, Table 5 shows that the variance of n_i is almost equal to $\bar{n}$ for the dispersed class, suggesting Poisson distribution of the number of units per haploid set in the population.

4. DISCUSSION

The models studied here may be a limited sample for describing the dynamic nature of genome evolution of higher organisms (Dover & Flavell, 1982). However, the analyses show that the duplication–deletion process within a genome plays an important role in the evolution of selfish DNA. This process has also been considered to be a major mechanism for creating new genes (Ohno, 1970) and for the evolution of multigene families (Ohta, 1980), and it now seems that the progressive evolution of higher organisms depends much on this process. At any rate, it produces raw material for evolution. However, it would be reasonable to suppose that only a very small minority of such material would be used for organismal adaptation, and thereby selfish DNA accumulates.

The problem of the presence or absence of polarity of self-replication ($\gamma_2 > \gamma_0$ or not) seems to be most important in recent discussions of the evolution of repeated sequences (Dover, 1982). Namely, unequal crossing-over is considered to be not directional, whereas polarity is sometimes found for gene conversion and transposition, and may have a large effect on the evolution of dispersed repeating sequences. In order to have correct understanding, one would need not only theoretical analyses as in this report, but also experimental studies planned in the framework of population genetics.

I thank Dr A. Robertson and other anonymous referees for their many valuable comments to improve the presentation. Supported by Grant-in-Aids 57120009 from the Ministry of Education, Science and Culture of Japan.

APPENDIX: METHOD OF SIMULATION EXPERIMENT

One generation of the experiment consists of the following processes; new occurrence of selfish DNA, sampling and selection of gametes, duplication–deletion of existing units of sampled and survived haploid sets, and recombination among the haploid sets.

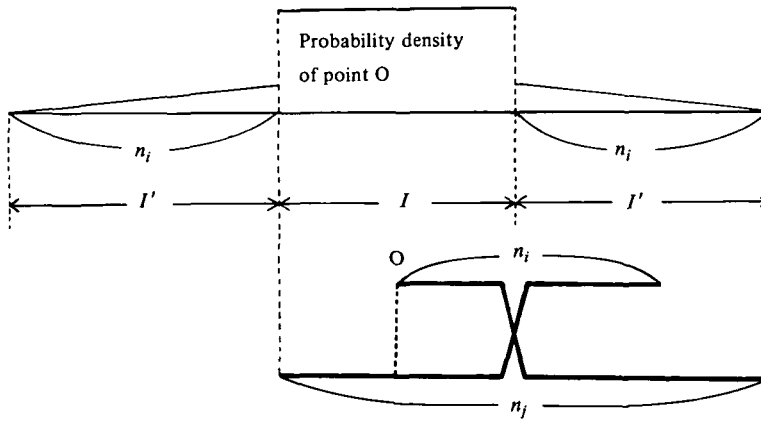

Fig. 3. Diagram illustrating the model of exchange process of the clustered selfish DNA. Upper figure shows the probability density of point O, and the lower one, the crossing-over between the DNA segments with n_i and n_j units.

Introduction of new units was carried out by choosing random haploid sets from the population $2Np$ times, and by increasing the number of units by one in each of the chosen haploid sets, where $2N$ is the population size and p is the rate of occurrence of a unit of selfish DNA per haploid set. Sampling and selection were done simultaneously by randomly choosing a haploid set and by determining its survival, where the probability of survival is $1-sn_i$ if n_i is the number of units. This process was repeated until the size of the population became $2N$.

The duplication–deletion process is different for the clustered and dispersed classes (Ohta & Kimura, 1981). For the dispersed class, duplication–deletion occurs independently for each unit, i.e. an individual unit of selfish DNA has a constant probability α_1 of either duplicating or being deleted. For the clustered class, unequal crossing-over is considered to be the main mechanism, whereby a certain number of units are simultaneously duplicated or deleted in one cluster. This number (x) obeys a distribution $2(1-x/l)/l$, where l is the maximum number being duplicated or deleted and is taken to be $0{\cdot}9n_i$ if a haploid set has n_i units. The distribution implies that a smaller shift at unequal crossing-over is more likely to occur than a larger shift. The occurrence of unequal crossing-over in a haploid set is determined by the probability $\alpha_2 n_i$, where α_2 is the rate of unequal crossing-over per one unit.

As mentioned earlier, the occurrence of new units in the population is almost equivalent to the replicational advantage, $\gamma_2-\gamma_0$, as far as the amount of selfish

https://doi.org/10.1017/S0016672300021029 Published online by Cambridge University Press

DNA at equilibrium is concerned. Here the rate of occurrence of a unit (p) equals $(\gamma_2-\gamma_0)\bar{n}$ if $\bar{n}$ is the average number of units per haploid set, provided that an equilibrium exists.

The recombination process among the haploid sets is also different between the clustered and the dispersed classes. For the dispersed class, free recombination is assumed. The free recombination is carried out by choosing N mating pairs randomly and, at a mating, distributing randomly the existing units into the two haploid sets. It maximizes recombination, since random distribution is expected for hemizygous units, but not for homozygous units. For the clustered class, the model of Ohta and Kimura (1981) is applied. The number of units being duplicated or deleted was determined so that it obeys the probability density as given in Fig. 3. Let β_I and β'_I be the rates of balanced and skewed recombination per cluster. Recombination was simulated by randomly choosing two haploid sets and repeating the process $N(\beta_I+\beta'_I)$ times. In the experiments $\beta_I = 10\beta'_I$, i.e. the balanced type is ten times more likely to occur than the skewed one.

REFERENCES

CHILDS, G., MAXSON, R., COHN, R. H. & KEDES, L. (1981). Orphons: dispersed genetic elements derived from tandem repetitive genes of eucaryotes. *Cell* **23**, 651–663.

CROW, J. F. (1979). Genes that violate Mendel's rules. *Scientific American* **240**(2), 134–146.

DOOLITTLE, W. F. & SAPIENZA, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603.

DOVER, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* **299**, 111–117.

DOVER, G. & DOOLITTLE, W. F. (1980). Modes of genome evolution. *Nature* **288**, 646–647.

DOVER, G. & FLAVELL, R. B. (eds.) (1982). *Genome Evolution*. London: Academic Press.

FEDOROFF, N. V. (1979). On Spacers. *Cell* **16**, 697–710.

FELLER, W. (1957). *An Introduction to Probability Theory and its Applications*, vol. 1. New York: John Wiley.

HARTL, D. L. (1975). Modifier theory and meiotic drive. *Theoretical Population Biology* **7**, 168–174.

HOLLIS, G. F., HIETER, P. A., MCBRIDE, O. W., SWAN, D. & LEDER, P. (1982). Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. *Nature* **25**, 321–325.

KIMURA, M. & OHTA, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771.

LUEDERS, K., LEDER, A., LEDER, P. & KUFF, E. (1982). Association between a transposed α-globin pseudogene and retrovirus-like elements in the BALB/c mouse genome. *Nature* **295**, 426–428.

NISHIOKA, Y., LEDER, A. & LEDER, P. (1980). An unusual alpha globin-like gene that has cleanly lost both globin intervening sequences. *Proceedings of the National Academy of Sciences of the U.S.A.* **77**, 2806–2809.

OHNO, S. (1970). *Evolution by Gene Duplication*. Berlin, New York: Springer-Verlag.

OHTA, T. (1980). *Evolution and Variation of Multigene Families*. Lecture Notes in Biomathematics, vol. 37. Berlin, New York: Springer-Verlag.

OHTA, T. (1981). Population genetics of selfish DNA. *Nature* **292**, 648–649.

OHTA, T. & KIMURA, M. (1981). Some calculations on the amount of selfish DNA. *Proceedings of the National Academy of Sciences of the U.S.A.* **78**, 1129–1132.

ORGEL, L. E. & CRICK, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.

PROUT, T. & BUNDGAARD, J. (1977). The population genetics of sperm displacement. *Genetics* **85**, 95–124.

TAKAHATA, N. (1981). A mathematical study on the distribution of the number of repeated genes per chromosome. *Genetical Research* **38**, 97–102.

VAN ARSDELL, S. W., DENISON, R. A., BERNSTEIN, L. B., WEINER, A. M., MANSER, T. & GESTELAND, R. F. (1981). Direct repeats flank three small nuclear RNA pseudogenes in the human genome. *Cell* **26**, 11–17.

VANIN, E. F., GOLDBERG, G. I., TUCKER, P. W. & SMITHIES, O. (1980). A mouse α-globin-related pseudogene lacking intervening sequences. *Nature* **286**, 222–226.