J. Mol. Evol. 14, 273–285(1979)

Journal of Molecular Evolution

Selection and Recombination in Populations Containing Tandem Multiplet Genes*

Arthur L. Koch[1]

Division of Biological and Medical Research, Argonne National Laboratory, Argonne, Il 60439 U.S.A.

Summary. Computer simulation for selective conditions that may apply in nature yielded three generalizations for prokaryotic organisms with recombinant mechanisms. (1) Selective forces can suffice to maintain a tandem gene family with the nearly optimum number of genes with little variance within the population. (2) Tandem genes will occur within the population unless the population is frequently cloned or unless the function due to a single copy is capable of over-providing the needs of the organism. (3) Even when there is no selective advantage or disadvantage due to extra gene copies, the population distribution becomes more skewed with time; and organisms with only single copies of the gene comprise a progressively larger fraction of the total. This may be the case with genes that function under strong cellular regulation.

Evolutionary implications of these calculations are that the occurrence of unequal recombination of tandem genes would greatly slow evolution via duplication of genetic material. This difficulty and its possible resolutions are discussed.

Introduction

Tandem genes, in the presence of recombination systems, offer the possibility of rapid expansion and loss in numbers of gene copies as selective conditions alter. In experimental studies of prokaryote evolution, duplicate genes arise, usually in tandem (see review in Anderson and Roth, 1977). Even when legitimate recombination is prevented by the absence of the *rec* system, duplications can occur (Beeftinck, et al., 1974). Such duplications were probably important early in the evolution of life because, no matter how rarely they were produced, a potential two-fold growth advantage insures that organisms with doublet genes can take over the population.

* Supported by the U.S. AEC under contract No. W-31-109-ENG-38.

[1] Current and permanent address: Department of Microbiology, Indiana University, Bloomington, IN 47405 U.S.A.

0022–2844/79/0014/0273/$ 02.60

In the presence of an effective recombination system, the further expansion of a doublet to form higher multiplets is a much more probable process. I have developed a computer simulation to understand the kinectics of the development and loss of multiplet classes under different selective conditions. The program has been used to calculate the kinetics of change of population distributions and the eventual steady states to try to understand the role of unequal crossing over in early prokaryote evolution. In particular I consider the impact of unequal crossing over between sister chromosomes of growing prokaryotes on the role of duplications in the evolution of new functions and on the mechanism proposed earlier (Koch, 1972, 1974) for the role of duplications in the evolution of more efficient enzymes.

A Semi-Realistic Model for Selective Advantage

The growth rate of an organism bearing a variable number of gene copies of a locus depends on the cell's need for that gene. For a gene coding for a metabolic enzyme whose synthesis is not under regulatory control it is reasonable to assume that the specific growth rate depends in a hyperbolic way on the number of gene copies. This assumption of the rectangular hyperbola has the property of diminishing return from extra gene copies. There must be an unfavorable effect of extra genes when unneeded protein is formed. It is reasonable to assume that this burden is simply proportional to the dilution of useful protein synthesis. On these bases, it is assumed that

$$\lambda_n = \frac{V_{max} n}{C + n} \cdot \frac{1}{1 + Bn}, \qquad (1)$$

where the first factor on the right is the hyperbolic increase, and the second is the protein burden factor. In equation (1) λ_n is the specific growth rate of a class of organisms containing n tandem genes (n-tuplets). C is the equivalent of a Michaelis-Menten constant and is to be measured in the same units as n, but it need not be an integer. The protein burden is designated by B, and in most instances for the purposes of illustration has been set at 0.01 of the total rate of macromolecular synthesis of cell constituents in organisms with only one copy of the gene. A single gene for β-galactosidase, when fully induced, represents a burden of about 0.02 (Novick and Weiner, 1957), so that 0.01 is not an unreasonable value. V_{max} would be the maximal growth rate in the absence of a proten burden but with an infinite number of gene copies present. Equation (1) seems a logical choice, but only for cases where the genes are not under effective cellular regulation. In that case, growth will be limited by other aspects of cell psysiology, and the hyperbolic portion of equation (1) may be omitted. Alternatively equation (1) can be retained and C assigned a small or zero value.

Equation (1) has the feature that λ_n rises with increasing number of gene copies up to a maximum growth rate and then progressively, but slowly, declines to zero as the protein burden increases. The number of genes giving the maximal growth rate is $\sqrt{B/C}$.

The specific growth rate, λ_n, is defined, as usual in microbiology, as satisfying

$$\frac{dN_n}{dt} = \lambda_n N_n \quad \text{or} \quad N_n = N_n' e^{\lambda_n t}, \qquad (2)$$

where N_n is the number of organisms of the n-tuplet class at time t, and N'_n is the number at the time taken as zero. These equations do not take into account recombination. This will be taken into account under the approximation that cell recombination takes place at the end of a unit of time. The unit of time has been chosen as the time in which organisms bearing only a single copy doubles exactly once; therefore,

$$\lambda_1 = 1n\ 2 \tag{3}$$

Consequently from equation (1), equation (1) with n chosen equal to 1, and equation (3), V_{max} can be eliminated and

$$\lambda_n = 1n\ 2\ \frac{(C + 1)(1 + B)n}{(C + n)(1 + Bn)} \tag{4}$$

The value of the growth factor for each n-tuplet class during one time unit can be calculated from $e^{\lambda n}$ and equation (4).

Algorithm for Unequal Crossing-Over

Table 1 diagrammatically shows two sister chromosomes, each with four copies of the same gene 'X' arranged in tandem. Any one of the genes in one chromosome may engage in recombination with any one of the genes in the other chromosome by homologous pairing. If pairing between homologous genes is equally probable and independent of position of each in its gene string, then each of the 16 ways is just as probable as any other. Pairing can take place, however, in seven different ways as shown in Table 1.

Table 1. Possibilities for unequal crossing over with four tandem genes

Type of overlap	Probability of occurring	Products
——————X X X X—————— ————————X X X X ——	1/16	Singlet and Septuplet
——————X X X X—————— ————————X X X X———————	2/16	Doublet and Sextuplet
——————X X X X—————— ————————X X X X———————	3/16	Triplet and Quintuplet
——————X X X X—————— ——————X X X X———————	4/16	2 Quadruplets
——————X X X X—————— ————————X X X X———————	3/16	Quintuplet and Triplet
——————X X X X—————— ————X X X X———————————	2/16	Sextuplet and Doublet
——————X X X X—————— ————X X X X———————————	1/16	Septuplet and Singlet

The probabilites for the general case of sister n-tuplets that undergo pairing are given by:

$$P_n(i) = \begin{cases} i/n^2 & ; i \leqslant n \\ (2n-i)/n^2 & ; n < i < 2n \end{cases} \quad (5)$$

This distribution is referred to as the triangular distribution and is isosceles triangular in shape with the apex at $P_n(n) = 1/n$.

Population Simulation

A Fortran IV program was written (Fig. 1) that applies the triangular distribution of equation (5), the recombination probability, and the growth rate dependency on n from equation (4), to compute the multiplet distribution at the end of a time unit for a population initially containing an arbitrary distribution of tandem multiplet gene copies. The program starts a cycle by taking the number of individuals in each currently extant multiplet class and computing the total number of recombinant events occurring in that class during a unit of time. The total number of organisms engaging in recombination of a particular multiplet class of genes depends on five multiplicative factors. The first is the total number of organisms in the multiplet class at the beginning of the unit of time. The second is the growth advantage, $e^{\lambda n}$. The third is the basic recombination frequency, r, determining the probability that a particular gene on one chromosome will recombine with a particular gene on the sister chromosome. The fourth factor is n^2, since if there are n copies, then there are n^2 times as many possible recombination events as if there were only one copy. The fifth factor is the number of opportunities to recombine, which depends on how many replications have occurred in the multiplet class during the time unit. It is $e^{\lambda n} - 1$, since this is the number of new organisms in the class that have been created during the interval.

It has been arbitrarily assumed that organisms grow at a rate dependent on the number of genes at the *beginning* of a time unit, since this number determines the phenotype of the growing organism. This assumption is in error if there are multiplet classes in the population growing so fast that they double many times during a unit of time. In the latter case organisms will enter new multiplet classes by recombination while the program calculates their growth on the basis of their ancestral n-tuplet class.

The total population resulting from recombination of a n-tuplet class is then partitioned into contributions to n-tuplet classes from 1 to 2n - 1. The resultant recombinant population is combined with the population of organisms that did not undergo recombination, but did, of course, multiply according to the same set of growth factors (Fig. 1). The computer then normalizes the numbers in each class and uses this as the initial distribution for the next time cycle.

The size of the memory array needed to store the final distribution increases by 2-fold minus 1 on each unit of time for a precise calculation. But for computability, the program was arranged so that the portion of the memory actually used is only increased when the calculated number of individuals in the largest class of the array allotted for storage is greater than 10^{-6} of total. This allows the program to be run for many thousands of generations for those cases where selective forces act with sufficient strength against organisms with many copies of the gene.

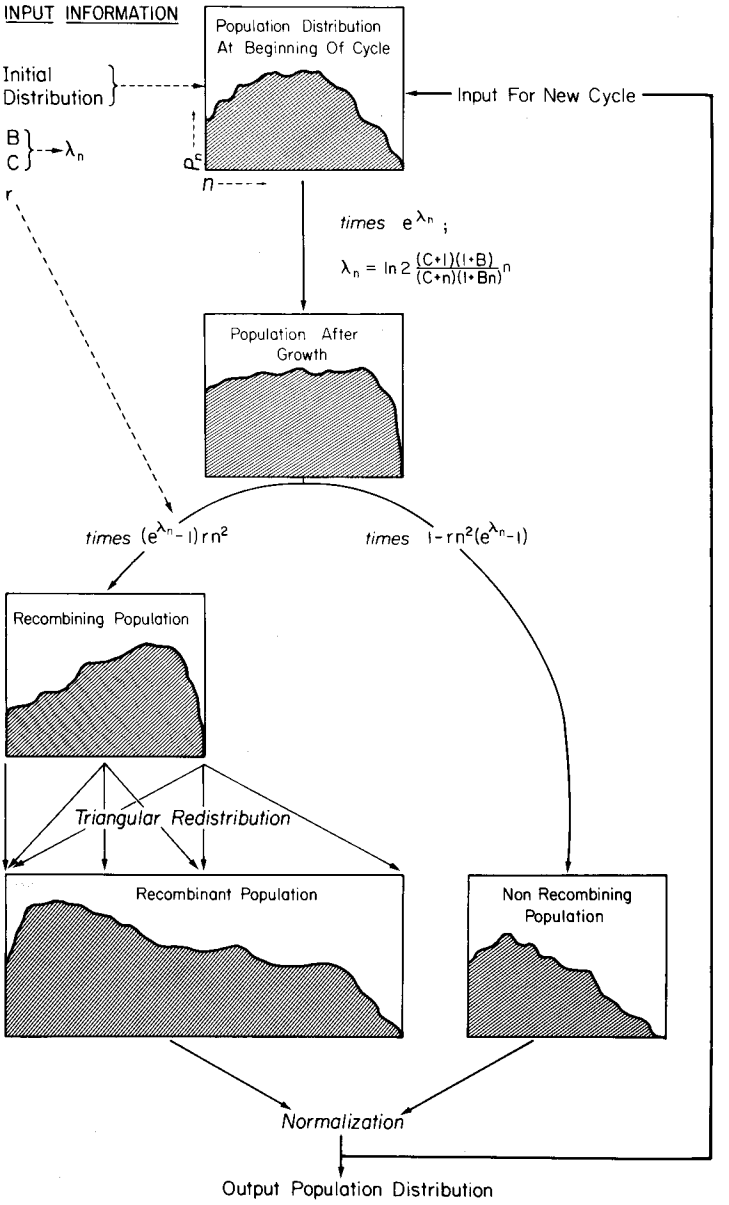

Fig. 1. Schema for calculation of population distribution. See text

Illustrative Cases

Figure 2 shows the computer simulation for a case where the basic recombination frequency was assumed to be 10^{-6}, C was set at 2 and B was set at 0.01. For these values

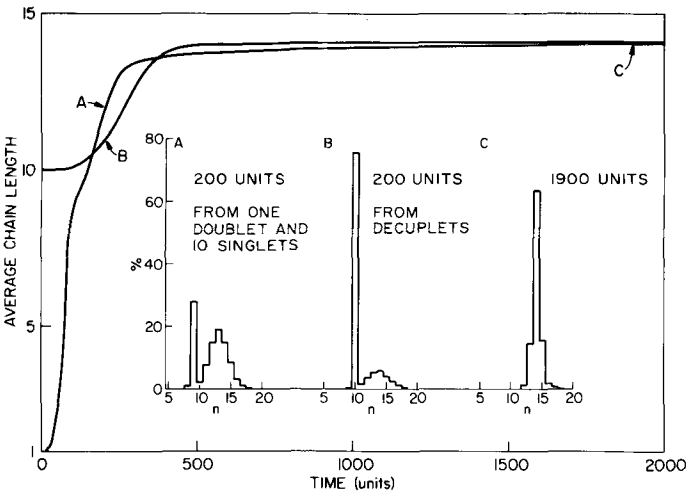

Fig. 2. Kinetic course of population. Computer calculation of the average multiplet class in two populations under the conditions that $r = 10^{-6}$, C = 2, and B = 0.01. This corresponds to a maximmum selective advantage for organisms with 14 tandem repeat genes. One population initially contained only individuals with exactly ten genes and the other consisted of one doublet organism and 10^6 singlet organisms. The population distributions at various indicated points are shown in the insets

of C and B, the growth factor during a unit of time $e^{\lambda n}$, rises from 2 at n = 1, to a maximum of 5.01 for n = 14, and thereafter declines gradually, reaching 2 again at n = 200. After 1900 generations, no matter what initial distribution was assumed, the average number of genes approaches a value a little greater than 14 and very nearly equal to $\sqrt{C/B}$, or $\sqrt{200} = 14.14$. The distribution around this mean is surprisingly sharp (see inset C). The time course of the evolution of the population mean from two initial distributions is shown. In one case, the original population was assumed to consist only of individuals with 10 genes. After 200 time units the population mean had increased slightly and the distribution (see insert B) had become bimodal, with one peak still at 10 genes and another broader peak at 14 genes. In the other case, the assumed initial distribution was two-valued, and consisted of one individual with a single duplicate together with 10^6 individuals each possessing a single copy of the gene. Nonetheless, it took only 50 generations for the descendents of the organism bearing a single duplication to become 50% of the population. At this point the population had a larger standard deviation than at any other time. Further development of the population led to the progressive narrowing of the distribution. By 200 time units the population distribution was stable (see insert A), had a mean of 12 genes, and was bimodal with peaks at 9 and 13 genes.

The most surprising conclusion from this example is how fast the population distribution reaches a time-independent steady state even when only a very small basic recombination frequency has been assumed[1]. Evidently, several hundred generations

[1] Footnote see next page

are a sufficient time to approach the limiting distribution for any reasonable initial population excluding only that case where duplicates are never produced and where no multiplets are initially present[2]. Also somewhat surprising, considering the gradualness with which growth rate depends on the number of genes near the maximum, is the sharpness of the distribution of multiplet genes. This, of course, results from selection driving the population towards the number of copies with the maximum specific growth rate. Countering this is the recombination process that tends to broaden the steady-state distribution.

When the calculations were repeated changing the value of r, it was found that the variance decreases in nearly exact proportion to the recombination frequency, and the distribution becomes increasingly more positively skewed and more positively kurtotic (broad shouldered)[3].

Most structural genes in prokaryotes occur in single copies. Evidence is accumulating that this also holds for the structural genes of eukaryotes (Davidson, 1976). To account for this, one must assume a value of C well below 0.1 to make singlets the predominate form in the steady state.

For this reason another example is shown in Fig. 3 with a low value of C. The initial distribution consisted entirely of a population of doublet organisms. At time zero, selection with C = 0.01 and B = 0.005 is imposed with several indicated recombination frequencies. For these choices of C and B, organisms with one and two gene copies have exactly the same survival value. Cells with higher numbers of gene copies have a lower growth rate, but the selective disadvantage against more than 2 gene copies is very slight. Nonetheless, this drives the population to lose its duplication since unequal crossing over creates singlets which grow exactly as fast as the doublets and triplets that grow more slowly. As is evident in Fig. 3, increasing the recombination frequency results in speeding the loss of the duplicate gene.

---

1 One important reason for setting up the detailed computer simulation instead of depending on analytical solutions of simpler cases was to study the change of frequency distributions of multiplets when recombination is very rare, but when long evolutionary times are involved. It is not selfevident how time and recombination frequency do interact. In several cases where parameters are chosen so that steady state distributions would take millions of generations to be approached, it was found that time and recombination frequency exhibit reciprocity before the steady state was approached. For example, for the case of equal growth advantage for all multiplet classes, and when the initial population consisted entirely of doublets, the same distribution of multiplets is obtained for fixed values of the product r·t. In fact, the first 4 moments for $t = 1000$; $r = 10^{-4}$ and when $t = 10{,}000$; $r = 10^{-5}$ are the same to 6 decimal places.

2 To approach equilibrium from any arbitrary distribution such as the two examples chosen for Fig. 2 takes a time that depends on all the parameters. For example, if C is reduced from 2 to 0.1 and the other parameters left the same, several hundreds of thousands of generations are needed. However, 2000 generations suffice if r is then increased from 10^{-6} to 10^{-4}.

3 Comparing the steady state distribution for a variety of parameters several rules emerge as approximations over certain ranges. Besides the rule given in the text, it is found that as C is increased the distribution tends to become more normal and the standard deviation increases approximate in proportion to C. Calculation of the steady state distributions takes a good deal of computer time. For the present studies the computation time was minimized by progressively decreasing r through the desired range of values as the calculation proceeds.

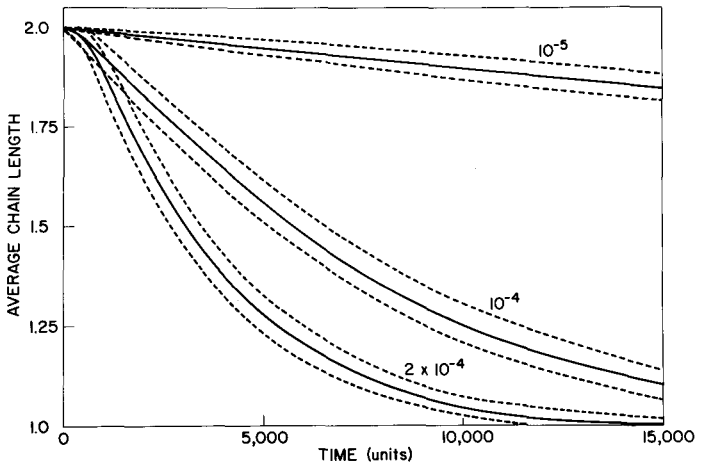

Fig. 3. Time course of population where singlets and doublets have equal specific growth rates. The initial population consists entirely of doublets. C = 0.01 and B = 0.005, which gives equal selective coefficients for singlet and doublets and lower selective values for higher multiplet classes. Recombination frequency and ± one standard deviation are indicated on the figure. This is a case where the population mean changes because of the recombination process and not because of selective growth advantage of organisms with singlet genes

Prevalence of Organism with Singlet Genes in the Absence of a Differential Growth Advantage

Figure 4 shows a case where there is no growth advantage, and exactly one recombination occurs in each genome every generation. Initially, every organism had 10 tandem genes. In the first generation a triangular distribution of organisms results with 1 to 19 genes with the maximum frequency at 10 genes according to equation (5). In successive generations each n-tuplet class generates a triangular distribution from 1 to 2n-1 and the population becomes enriched with smaller numbers of genes. However, the average number of genes does not change because of the assumption that organisms of all classes have the same selective advantage. This phenomenon is related to the gambler's ruin problem (Feller, 1968) in that the singlet state is an absorbing state: once a singlet organism is produced, all its descendents will be singlet, while multiplet organisms can continue to throw off progeny with singlet genes.

These distributions deviate very much from the normal Gaussian distribution and the median, mode, and mean are quite different from each other. Fig. 4 shows how the median and mode change with successive generations while the mean remains at its inital value of 10. Note that by only the fourth generation the most typical organism in the population has only a single copy of the gene. (At the eighteenth generation 50.5% of the population are singlets.) Consequently, as we consider growth of microorganisms in nature, colonization of new habitats would have the effect of recloning the population and would almost always result in a culture containing only pure singlets. Even for the case where initially every organism had 100 tandem genes, singlets

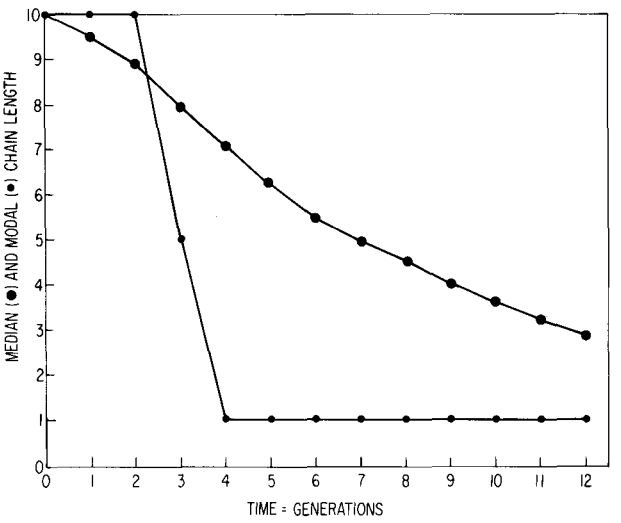

Fig. 4. Shift of median and mode brought about by continued recombination with no selective advantage. A population was assumed initially consisting of organisms each with 10 tandemly repeated genes that engage in recombination once every generation. The distribution becomes progressively more positively skewed. This is reflected in a decrease in median and mode even though the mean number of tandem genes remains at 10

quickly become the dominant type in the population; the mode was reduced to 3 in only 7 generations[4].

Discussion

The only theoretical treatment of this type of problem in the literature of which I am aware is given in Crow and Kimura (1970). They make very different assumptions and envisage a much different problem. (1) They assumed that there is an optimal number of gene copies and that the selection is quadratic and symmetric about this optimum; whereas I have assumed equation (1). (2) They assumed recombination to be independent of n; whereas I have assumed it to be proportional to n^2. (3) They assumed that selective advantages are small; whereas I have allowed high selective

[4] The shape of these distributions is sufficiently abnormal as to prevent the important aspects being depicted on any single type of graph. Consequently, I present some numerical aspects for a single case of 100 tandem genes after 12 generations of nonselective growth where recombination occurs in each generation. The resultant population contains 6.3% singlets and 3.4% doublets. The frequency gradually and monotonically decreases with increasing n. One out of two organisms have more than 30 genes, 1/10 more than 250; 1/100 more than 1040; $1/10^3$ more than 2540; $1/10^4$ more than 4575; $1/10^5$ more than 6425 and $1/10^6$ more than 7680. Thus choosing a single organism at random to clone a new culture or habitat has a 50% chance of reducing the mean number to below 30 genes, but a 10^{-6} chance of increasing the number to above 7680.

advantages by including exponential growth factors over single units of time. (4) They assumed that recombination only changes the number of gene copies by a small fraction; whereas I have allowed for all possible recombinants from 1 to 2n-1 by assuming equation (5). (5) They make no allowance for the extra recombination events due to rapid growth leading to multiple generations during a unit of time; whereas I include the factor $(e^{\lambda n} - 1) \cdot r \cdot n^2$ to take into account multiple doublings during a unit of time.

Their model is appropriate for tandem gene families in interbreeding populations which are large and cannot change much; whereas my model applies specifically to recombination between sister chromosomes in a prokaryote cell that has not yet divided to separate the chromosomes into daughter cells. The model has been developed for potentially rapidly growing organisms when some gene function severely limits growth. This must have occurred frequently during early evolution, under natural conditions today, as well as under laboratory conditions[5].

It is worth contrasting the predictions of the two models. That of Crow and Kimura leads to a normal distribution about the optimum number of genes with a variance which is proportional to the square root of the recombination frequency. My model leads to no closed analytical form for either the intermediate or final distributions, although Witten (1975) has made some attempts in this direction. However, numerical calculations showed that the variance decreases in proportion to the recombination frequency, when r is small. The steady state distributions for my model are non-normal and become increasingly more positively skewed and more positively kurtotic as r is decreased.

The main conclusions to be drawn from the computer simulation of this model are:

1. A very low incidence of duplications in the population and/or very low mutation rates producing tandem duplications together with very low recombination rates suffice, in the presence of selective pressures, to produce populations of multiplets whose mean number of genes closely approaches the optimum number. Only a few hundred to a few thousand generations are required for close approach to the steady-state distribution, even if the initial distribution is extremely different from the final one and recombination is rare.

2. The resulting distributions may be quite narrow, although not normal. With reasonable choices for parameters, one could explain maintenance of gene numbers of multiplet genes such as rRNA and γ-globulin in eukaryotes by this selection process.

3. Recombination continues to throw off organisms with only single gene copies. This statistical factor makes the distribution J-shaped and positively skewed when the growth rate is independent of the number of gene copies. For microorganisms subject to repeated cloning under natural conditions, this serves as the equivalent of a strong selective force to produce populations with single gene copies. Because most structural genes are only present in single copies in prokaryotes and in diploid eukaryotes, this process may be a (partial) explanation for the prevalence of single copy structural genes.

---

5 Cases where the number of genes copies strongly effects growth rate are surely of importance in early evolution and are not amenable to approximate analytical solutions. Failure to take into account exponential growth and the extra recombination events during a unit of time greatly alters the calculated evolution of the population distribution of multiplets for circumstances like that of Fig. 2 for the case of an initial population consisting of a single doublet in a population of 10^6 singlet containing organisms.

4. In order for a gene to be present as a single gene copy in most of the individuals of an uncloned population, the gene's function must be such that the effect of a single copy of the gene produces a functional capacity many times in excess of the cells actual needs, i.e., C must be near zero so that $\sqrt{C/B}$ is small. Of course, the gene may be regulated most of the time and/or the function of the gene product controlled, but if in times of need the gene cannot function to produce enough product so that more would be of negligible advantage, then tandem genes will contribute to the population structure no matter how slowly the original gene duplications arise.

This last is possibly the most important conclusion of the analysis and implies that at any earlier time when enzymes were less efficient, high multiplet organism might have been common. Moreover, reduction to today's situation where singlets are the rule may have been concomitant with the development of highly efficient regulatory systems permitting high production under special conditions, but preventing overproduction in most circumstances.

The existence of regulatory systems has its repercussion in ecological study of modern microorganisms. Most often when a strong selection is imposed on organisms, regulatory mutations overproducing the limiting enzyme arise (first observed by Novick and Horiuchi, 1961). Alternatively, and usually secondarily, a tandem duplicate arises. When duplication is the primary event, the duplicate may not be regulated in the original way, and instead yields fully constitutive and/or promoted levels in one step (Beeftinck et al., 1974). Either change yields much more gene product than the copy retaining the original regulation and therefore has a tremendous selctive advantage in an environment in which the function is limiting for growth. In the case of tandem duplications in which the regulation has been lost, when conditions change back, the unregulated copy would quickly be lost by unequal crossing over since the singlet retaining the original regulation has the selective advantage.

To return to the relevance of unequal crossing over to the evolution of enzymes, it is shown here [and found in practice (Jackson and Yanofsky, 1973; Anderson et al., 1976)] that genes in tandem will be lost rapidly if not maintained in the population by selection. They will be lost both as a type of selective process if $\sqrt{C/B}$ is small and as a result of the cloning phenomenon. Therefore, at least for modern organisms capable of recombination and regulation, tandem genes are not likely to serve an important role in the type of intrapeptide evolution I have proposed previously (Koch, 1972), nor indeed for the other evolutionary roles assigned to duplications.

There is a second reason to discount tandem genes in these evolutionary roles in organisms capable of facile recombination. This is because a series of unequal crossing over events also decreases the variability between the homologues even when it does not lead to a change in the number of gene copies. This role of unequal crossing over has been modeled and discussed by Smith (1974). His work has been extended by Black and Gibson (1974) and Ohta (1978). It is no doubt the explanation of 'coincidental evolution' (Hood, 1976). In Smith's process, a mutant gene would eventually become fixed or lost as the result of the random walk made possible by unequal crossing over where the total number of gene copies is maintained more or less constant by unspecified processes.

Reanney (1976) in a recent review of extrachromosomal elements as possible agents of evolution and development, argues strongly and effectively against duplication

serving as the source of genetic material for evolution in modern microorganisms. He argues that the variability is not engendered by mutation, but rather by transfer from other organisms via viruses, plasmids, etc. Clearly, modern use of antibiotics has not caused the evolution de novo of resistance genes within the pathogenic microorganisms, but it has caused the mobilization of resistance genes that preexisted elsewhere (Falkow, 1975). Even such apparently simple changes of an enzyme so that it acetylates an antibiotic instead of some other substrate is accomplished by gene mobilization and not contemporary gene evolution. Thus, gene evolution toward better or different functions has been rendered de facto ineffective by gene transfer processes from other organisms. We must look to other paradigms that must have been important in earlier times when gene evolution was taking place de novo than the study of medem organisms.

Let us consider the circumstances under which a gene duplication may serve in the improvement of the original gene function or the evolution of a new function. The first is the possibility that the majority of evolution *via* a duplication mechanism preceded the development of effective recombination mechanisms, i.e., it antedated the development of crossing over mechanisms, repair mechanisms, integration mechanisms, etc., and incidentally preceded the development of viral and plasmid mechanisms for gene transfer from organism to organism. The second is the entirely conjectural possibility that mechanisms have evolved specifically blocking recombination in certain regions of the genome, and de novo evolution is restricted in the main to these regions. Thomas (1970) has discussed this as one of several possibilities for maintaining gene number in a tandem series that does require gene function.

The remaining special circumstances involve the formation of non-tandem genes either immediately or secondarily. A DNA fragment may be illegitimately added near the end of a linear chromosome by recombination or fusion to produce non-tandem duplication. This is the common process in eukaryotes (Hansche, et al., 1978). A tandem duplication may arise, followed by a mutation so that DNA between two gene copies becomes essential, thus selecting against loss by unequal crossing over. Such a region could become essential to the organism by differentiation to some new vital function or by inactivation of its duplicate. Finally, an originally tandem gene may be transposed to a second site via a transposon or some other excision and reintegration mechanism. This last possibility has great appeal. It could allow a duplicate to be removed from the influence of its regulatory genes, allow the accumulation of individually nonadaptive changes, and yet permit a return to the original regulation. Even though the possibilities listed in the preceeding paragraphs collectively allow evolution *via* the silent duplicate, they do slow it relative to a genetic system where unequal crossing over is very improbable.

Finally, the present calculations have considered only unequal crossing over between sister chromosomes. We have thus omitted consideration of the effect of recombination between gene copies within a tandem series. Such recombination has the property of only reducing the number of gene copies unless the excised DNA is brought under control of some replicon. Internal recombination also leads to elimination of all but one of a series of tandem duplicates in the absence of selective pressures.

Acknowledgements: This work was carried out in the Division of Biological Research at the Argonne National Laboratory while the author was visiting as the Argonne University Association Distinguished Appointee for 1974. Preliminary work was carried out at Indiana University with

Matthew Witten. The author owes a debt of thanks to Carol Fox, Martin Kraimer, Jan Ruston and Matthew Witten for aid in learning how to force a big computer to do one's will. I especially wish to thank Herbert Kubitschek in whose group I spent an enjoyable year at Argonne and whose budget paid for the time on the IBM 195. A big debt is due Barry Dancis who found a flaw in the logical development of equation (5) in time so that all the figures presented here could be recalculated and corrected. Finally I would like to thank Nicolas Glandsdorff, Stine Levy, Rollin Richmond, Marcia Rosenthal, and Matthew Witten for constructive comments on the manuscript. Work at Indiana University is supported by NSF under grant PCM 72 01852-A03.

References

Anderson, R.P., Roth, J.R. (1977). Ann. Rev. Microbiol. **31**, 473–505
Anderson, R.P., Miller, G.C., Roth, J.R. (1976). J. Mol. Biol. **105**, 201–218
Beeftinck, F., Cunin, R., Glandsdorff, N. (1974). Mol. Gen. Genet. **132**, 241–253
Black, J.A., Gibson, D. (1974). Nature **250**, 327–328
Crow, J.F., Kimura, M. (1970). An introduction to population genetics theory. New York: Harper
Davison, E.H. (1976). Gene activity in early development. New York: Academic
Falkow, S. (1975). Infectious multiple drug resistance. London: Pion
Feller, W. (1968). An introduction to probability theory and its application. Vol. 1, 3rd ed. New York: John Wiley
Hansche, P.E., Beres, V., Lange, P. (1978). Genetics **88**, 673–687
Hood, L. (1976). Federation Proc. **35**, 2158–2161
Jackson, E.N., Yanofsky, C. (1973). Ann. Rev. Microbiol. **31**, 473–505
Koch, A.L. (1972). Genetics **72**, 297–316
Koch, A.L. (1974). Genetics **77**, 127–142
Novick, A., Horiuchi, T. (1961). Cold Spring Harbor Symp. Quant. Biol. **26**, 234–245
Novick A., Weiner, M. (1957). Proc. Natl. Acad. Sci. **43**, 553–556
Ohta, T. (1978). Genetics **88**, 845–861
Reanney, D.C. (1976). Bacteriol. Rev. **40**, 552–590
Smith, G.P. (1974). Cold Spring Harbor Symp. Quant. Biol. **38**, 507–513
Thomas, C.A., Jr. (1970). In: The Neurosciences: Second Study Program, F.O. Schmitt, ed., p. 973. New York: Rockefeller University Press
Witten, M. (1975). Growth kinetics of tandem multiplicate genes: A possible model for the evolution of macromolecular complexities. Master thesis submitted to Biophysical Sciences at State University of New York at Buffalo

Received April 13, 1979