Istituto di Matematica Guido Castelnuovo, University of Rome, Italy

GENE DECAY

*I. Stochastic Model of Gene Decay**

CARLA ROSSI**

SUMMARY

On the basis of the parameters of gene stability suggested by Gedda-Brenci's model on the *Ergon/Chronon System*, a mathematical and a stochastic model of gene decay are worked out.

INTRODUCTION

According to gerontologists and geneticists such as Burch (1968) and Szilard (1959), aging and death are caused by specific mechanisms of gene decay. The whole population would be exposed to the risk of some mechanisms of decay, while additional risk would be confined to subpopulations constituted by a not well determined number of predisposed subjects. In this model, only predisposed subjects would be hit by a well-determined clinical situation connected with the *aging* process.

This model of gene decay, restricted to predisposed subpopulations, has been overcome by a model based on gene stability: the *Ergon/Chronon System.*

In 1961, at the Second International Congress of Human Genetics, Gedda suggested the hypothesis that the gene may have a lifetime (*chronon*) determined by its own stability (*ergon*). Then, in " Biology of the Gene ", Gedda and Brenci (1969) suggested that the stability of a gene be determined by the following factors:

(*a*) *Code degeneracy*, i.e., the differential stability of synonyms coding for the same information through different ratios of G-C and A-T bases, the former being more stable;

(*b*) *Redundancy*, i.e., the number of repetitions of the DNA molecule representing a given gene;

(*c*) *Repair*, i.e., the efficiency of the enzymes and cofactors conditioning the repair of a gene's specific molecular structure.

* The present work is essentially an abridged version of the thesis in mathematics discussed by the author with Professor B. de Finetti (University of Rome: 12 July 1971).

** Presently at: Istituto di Calcolo delle Probabilità, Università di Roma.

Rec. 16. IX. 1972

Acta Genet. Med. Gemellol. (1972), 21: 191-196

THE α INDEX OF STABILITY OF A DNA MOLECULE

A DNA molecule is meanly constituted by 1,500 nucleotides, i.e., 500 codons. Generally, every codon has two bases, specifically coding for a given aminoacid, and a third base that may vary. Considering that A-T and G-C pairs have not the same stability, G-C being much more stable,[1] it is clear that individuals differing in their A-T/G-C ratio are predisposed in a different way to those random events determining gene decay. Then, the index of stability of a DNA molecule, α, corresponds to the number of G-C pairs in the third position of a triplet: the stability of a gene depends on the value that α assumes.

But α can't be easily observed; therefore, we must consider its probability distribution.

This may be determined by considering that the third nucleotide in a triplet is either A-T or G-C with a probability $\frac{1}{2}$, and that the triplets are incorrelated one to the other. We thus have a binomial distribution, that can be approximated to a normal one, the expected value of which is $m = N$ ($m = 250$), and the variance, $\sigma^2 = N/2$ ($\sigma^2 = 125$, $\sigma = 11.2$).

It may be concluded that a high proportion of individuals (consider, about 99%) have their index of stability in the range of $m \pm 3\sigma$, i.e., 250 ± 34. We thus obtain for the density function $f(\alpha)$:

$$f(\alpha) = \frac{1}{\sqrt{(\pi N)}} \, e^{\frac{-(\alpha - N)^2}{N}} \qquad [1]$$

$$f(\alpha) = \frac{1}{5\sqrt{(10\pi)}} \, e^{\frac{-(\alpha - 250)^2}{250}} . \qquad [1']$$

THE β NUMBER OF MOLECULES EFFICIENT AT THE TIME $t = 0$

Since the chemical estimates give us a number of DNA molecules greater than the number of genes that can be estimated by biological methods, we suppose that every biological unit (gene) is composed by some identical molecules.

For instance, the number of DNA molecules in a human individual is estimated at about 1,200,000, while the number of genes can be evaluated at about 60,000: then, we can say that a human gene is meanly composed by 20 DNA molecules. We shall denote this number by β_0.

It may be that some of the molecules constituting a gene be not efficient for coding; i.e., there may be some errors that one may inherit from his parents, i.e., at the time $t = 0$.

[1] The base A is bonded to the base T through only two hydrogen-bonds, while G is bonded to C through three hydrogen-bonds.

The redundancy β is equal to the number of efficient molecules in a gene at the time $t = 0$.

We can't easily observe the parameter β; then, we must consider its probability distribution.

Let us call p the probability that a molecule be efficient at the time $t = 0$, and suppose that the errors in the molecules are incorrelated one to the other; then, we obtain a binomial distribution between 0 and $\beta_0 = 20$:

$$p_h = \binom{\beta_0}{h} p^h (1 - p)^{\beta_0 - h} \qquad [2]$$

$$p_h = \binom{20}{h} p^h (1 - p)^{20-h} . \qquad [2']$$

Now, we should consider that, if β is less than, or equal to, a well defined limit value L, the gene is not active; it is useful to pool together the cases where $h \leq L$, for in such cases we can't study the gene decay.

The probability p' that a gene is not active at the time $t = 0$, without any consideration about the number of efficient molecules, is then:

$$p' = \sum_{i=0}^{L} p_i , \qquad [3]$$

and the probability $1 - p'$ of a gene being active, is the sum of the other terms given by the binomial distribution.

PROBABILITY DISTRIBUTION OF $M(t)$

The Number of Molecules Efficient at the Time t *and Expectation of the Time of Decay*

With time, random events produce errors in the DNA molecules. The accumulation of such random errors, that can be produced either in an efficient molecule, that they make not efficient (mutant), or in a not efficient one, making no change, is *gene decay.*

The velocity μ of the process that produces errors in the molecules is a constant.[2] Then the probability $p(t)$ of a molecule to be still efficient at the time t is $p(t) = pe^{-\mu t}$.

To study the process of gene decay, let $M(t)$ be the number of molecules still efficient at the time t. In particular, when $t = 0$, $M(0)$ is the random number β we have considered before, and $M^* = 20$ is the maximum value of it.

[2] The velocity of a stochastic process with independent increments, as the one we are considering now, is the expectation of the number of events which occur in a time unit interval. Then, μt is the expectation of the number of events in an interval of length t, and $\mu dt + o(dt)$ is the probability that an event occurs in the interval $(t, t + dt)$.

https://doi.org/10.1017/S1120962300010891 Published online by Cambridge University Press

We assume the errors to be incorrelated one to the other; the probability distribution of $M(t)$ can be obtained in the same way as the one of $M(0) = \beta$; only, we must replace the probability p by the probability $p(t) = pe^{-\mu t}$:

$$p_h(t) = \binom{M^*}{h} [p(t)]^h [1 - p(t)]^{M^*-h} \qquad [4]$$

$$p_h(t) = \binom{20}{h} pe^{-\mu ht}(1 - pe^{-\mu t})^{20-h} \quad . \qquad [4']$$

By studying the probability distribution of $M(t)$, we can easily obtain the density function of the time of gene decay. Such an event occurs when a random error reduces $M(t)$ from $L + 1$ to L. Then, the probability that such decay occurs in any time interval $(t, t + dt)$ is:

$$f(t)\, dt = \frac{L+1}{20}\, p_{L+1}(t)\mu dt, \qquad [5]$$

from which we obtain:

$$f(t) = \frac{L+1}{20} \binom{20}{L+1} p^{L+1}\, e^{-(L+1)\mu t}\, (1 - pe^{-\mu t})^{20-(L+1)} \quad . \qquad [6]$$

We can now calculate the expectation of the time of gene decay:

$$E(t) = \int_0^{+\infty} tf(t)dt = \mu\, \frac{L+1}{20} \binom{20}{L+1} p^{L+1} \int_0^{+\infty} te^{-(L+1)\mu t}\, (1 - pe^{-\mu t})^{20-(L+1)}\, dt. \qquad [7]$$

We must consider the integral:

$$\int_0^{+\infty} t \left(\sum_{i=0}^{20-(L+1)} A_i\, e^{-B_i t} \right) dt = \sum_{i=0}^{20-(L+1)} \int_0^{+\infty} tA_i\, e^{B_i t}\, dt.$$

Then we can integrate by parts:

$$\int_0^{+\infty} tA_i\, e^{-B_i t}\, dt = -\, [t\, (A_i/B_i)\, e^{-B_i t}]_0^{+\infty} + (A_i/B_i) \int_0^{+\infty} e^{-B_i t}\, dt.$$

The first term in the sum in the second member is equal to zero, and the second is:

$$(A_i/B_i) \int_0^{+\infty} e^{-B_i t}\, dt = -\, [(A_i/B_i^2)\, e^{-B_i t}]_0^{+\infty} = A_i/B_i^2$$

and, by writing $B_i = \underline{B}_i \mu$, we obtain:

$$E(t) = \frac{1}{\mu}\, \frac{L+1}{20} \binom{20}{L+1} p^{L+1} \sum_{i=0}^{20-(L+1)} (A_i/\underline{B}_i^2) = \frac{K}{\mu}. \qquad [7']$$

STOCHASTIC MODEL OF GENE DECAY
How the Parameters α and β Influence the Process

We are now interested in representing the process of gene decay through a stochastic model, such as a drawing of balls out of urns.

Such a model will allow us to better study the process and, by simulating it, to gather some results that may be used to continue our work without necessarily going into complex experimentation.

In order to represent gene decay by a stochastic model, let us consider an urn containing 20 balls. At the time $t = 0$, we have b black balls and $20 - b = r$ red balls. Now, let the black balls represent the efficient molecules and the red ones the inefficient molecules.

Let us then choose a time unit t^*, such that $\mu t^* = 1$.

We may now start drawing balls: if we draw a red ball, we put it back into the urn; if we draw a black ball, we substitute it and put a red ball into the urn.

The process terminates when the number of black balls remained into the urn is equal to the lower limit, L.

The number of experiments made, n, multiplied by t^*, gives us t', i.e., the time of gene decay (calculated in years), according to the stochastic model.

Using the same data, we also simulated the process through the Monte Carlo method, so as to study how the parameters α and β influence the process.

The parameter α influences the process by modifying μ: we then obtain a function $\mu(\alpha)$, i.e., the velocity of the process as a function of α, or, as a function of the human individual we want to consider. It is a monotone decreasing function of α.

We can make the hypothesis that the function $\mu(\alpha)$ is a function of the form: $\mu(\alpha) = qe^{-\lambda\alpha}$. The constants q and λ must be estimated.

The parameter β influences directly the expectation of the time of gene decay, we suppose linearly.

Now we have the complete expression of the expectation of the time of gene decay, as a function of the two parameters α and β:

$$E(t) = \frac{K}{\mu(\alpha)} + \theta\,[\beta - E(\beta)] \qquad [8]$$

and $\theta = t^*$.

REFERENCES

Burch P.R.S. 1968. Growth, Disease and Aging. Oliver & Boyd, Edinburgh.

De Finetti B. 1970. Teoria della Probabilità. Einaudi, Torino.

Feller W. 1957. An Introduction to Probability Theory and Its Application. Wiley & Sons, New York.

Gedda L., Brenci G. 1969. Biology of the gene: the Ergon/Chronon System. Acta Genet. Med. Gemellol., 18: 329-379.

Marrott Sinex F. 1961. Biochemistry of aging. Science, 134: 1402-1405.

Streheler B: 1960. General theory of mortality and aging. Science, 132: 14-21.

Szilard L. 1959. On the nature of the aging process. Proc. Natl. Acad. Sci. USA., 45: 34-45.

RIASSUNTO

Sulla base dei parametri della stabilità del gene, suggeriti nel modello di Gedda-Brenci sul *Sistema Ergon/Chronon*, vengono costruiti un modello matematico ed un modello stocastico del decadimento del gene.

RÉSUMÉ

Sur la base des paramètres de la stabilité du gène, suggérés dans le modèle de Gedda-Brenci sur le *Système Ergon/Chronon*, un modèle mathématique et un modèle stochastique de l'épuisement du gène sont élaborés.

ZUSAMMENFASSUNG

Auf die von Gedda-Brenci im Modell des *Ergon/Chronon-Systems* vorgeschlagenen Parameter gestützt, wird sowohl ein mathematisches als ein stochastisches Modell für den Genverfall ausgearbeitet.

Dr.ssa Carla Rossi, Istituto di Calcolo delle Probabilità, Facoltà di Scienze Statistiche dell'Università, Rome, Italy.

https://doi.org/10.1017/S1120962300010891 Published online by Cambridge University Press