Genet. Res., Camb. (1983), **41**, *pp.* 47–55
With 3 text-figures
Printed in Great Britain

Time until fixation of a mutant belonging to a multigene family

By TOMOKO OHTA

National Institute of Genetics, Mishima 411, Japan

(*Received 5 April 1982*)

SUMMARY

Time until fixation of a mutant that occurs in one copy of a multigene family was investigated from the standpoint of population genetics. Because of the complexity of the double process of random drift (on the chromosome and in the population), an approximate method based on the rate of steady decay of genetic variability is applied. The simple model of gene conversion with constant gene family size is also used. The expectation based on the approximate method is shown to be valid by extensive Monte Carlo simulations, and the results are useful for understanding the mechanisms of turnover of multigene families, when comparison is available between closely related species.

1. INTRODUCTION

It is now known that there exist many repeated gene families in the genomes of higher organisms (see Dover 1982 for review). Gene families are characterized by coincidental or concerted evolution, and population genetical theory has been worked out under the assumption of constant occurrence of homologous but unequal crossing-over (see Ohta, 1980 for review). The theory is mainly concerned with the probability of gene identity (identity coefficient) of multigene families. However, for experimental geneticists, time until fixation of a mutant gene appears to be more meaningful than identity coefficient in applying the theory to real data (e.g. Coen, Thoday & Dover, 1982). The process of mutant spread is complicated, since it occurs simultaneously on the chromosome and in the population, and the explicit solution for the time until fixation comparable to Kimura & Ohta (1969*a*) for a single mutant in finite populations (see also Ewens, 1979), cannot be obtained. A simpler approach is to estimate approximately the time using the rate of decay of variability. In this report, the steady rate of decay of variability is obtained, and the estimate, based on this rate, for the time until fixation of a mutant gene is examined by extensive Monte Carlo experiments. The results suggest that this method is promising. Recent observations on species comparisons of gene families will be discussed in the light of these analyses.

* Contribution no. 1442 from the National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan.

2. MODEL AND ANALYSES

In my original analysis of multigene families (Ohta, 1978, 1980), the process of mutant gene spreading was treated as a double diffusion; diffusion of genes on the chromosome through unequal crossing-over, and diffusion of chromosomal types in the population by random genetic drift. The analysis is not exact, and the shift of position of loci on the chromosome by unequal crossing-over is not precisely treated in calculating the transition equations of the identity coefficients. Even in more exact analyses (Kimura & Ohta, 1979; Ohta, 1982*b*), the treatment is not completely precise. The new model of gene conversion (Ohta, 1982*a*) is straightforward and precise, since no shift of position of loci on the chromosome occurs – merely the transfer of a gene segment. Nevertheless, the model describes the same process of concerted evolution of multigene families without change in gene family size, and it is used here for the analysis of the time until fixation.

By the analysis of a single chromosomal line, Nagylaki & Petes (1982) have shown that small conversional advantage or disadvantage has a large effect on the time for spreading. Dover *et al.* (1982) also suggested the significance of polarity in certain gene families. However, for simplicity, the completely random process is treated here; neither polarity of conversion nor natural selection at the organismal level is considered.

Let N be the effective population size, and n the number of tandemly arranged genes on a chromosome. Thus there are $2nN$ genes in the population. Identity coefficients increase through intrachromosomal gene conversion and random genetic drift. As in the previous study, let λ be the rate in one generation at which a gene is converted by any one of the remaining $(n-1)$ genes with equal likelihood. Here I assume that conversion is 'asymmetric' (see Nagylaki & Petes, 1982). The real process of conversion is likely to involve one piece of a split gene (Miyata *et al.* 1980; Slightom, Blechl & Smithies, 1980; Yamawaki-Kataoka *et al.* 1982). Therefore λ is the average rate at which any small unit such as an amino acid site or a nucleotide site is converted by the homologous unit of another locus of the multigene family. In addition to random drift and gene conversion, recombination occurs at meiosis, and is assumed always to be 'equal'. Let β be the rate at which recombination occurs between the two adjacent loci of a multigene family, thus the rate per one family is $(n-1)\beta$. Note that this definition of β is different from, and more general than the previous one (Ohta, 1982*a*), as shown later. No mutation is assumed and the rate of steady decay of genetic variability is obtained as follows.

As in the previous report (Ohta, 1982*a*), let f be the average probability of allelic identity, C_1 the average identity probability of genes at different loci of the multigene family on the same chromosome, and C_2 that of two genes taken from different loci of two homologous chromosomes of the population. Note that the unit to compare identity needs to be smaller than the ordinary gene and may be a nucleotide site, an amino acid site or a piece (exon) of a gene, since the unit to compare identity must be smaller than or equal to a region converted at a time.

Then the transition equation of the identity coefficients from one generation to the next takes the following form, by letting $\mathbf{c} = (f, C_1, C_2)$ (Ohta, 1982*a*),

$$\mathbf{c}_t = \mathbf{A}\,\mathbf{c}_{t-1} + \mathbf{b}, \tag{1}$$

where

$$\mathbf{A} = \begin{bmatrix} 1-1/(2N)-(n-1)\alpha & 0 & (n-1)\alpha \\ 0 & 1-\alpha-\dfrac{(n+1)}{3}\beta & \dfrac{(n+1)}{3}\beta \\ \alpha & 1/(2N) & 1-\alpha-1/(2N) \end{bmatrix} \tag{2}$$

and

$$\mathbf{b} = 1/(2N,\ \alpha,\ 0), \tag{3}$$

with $\alpha = 2\lambda/(n-1)$, and the subscript t denotes the tth generation. The coefficient of $2\lambda C_2 = (n-1)\alpha C_2$ of equation (3) of Ohta (1982*a*) is erroneous and should be '+', and agrees with the present formula. Note also that the coefficient of β is different from that of the previous formulation (equation (5) of Ohta, 1982*a*). This is because β is now defined as the rate between two adjacent loci, whereas the previous one is that per total family. Also, the coefficient $(n+1)/3$ is the average recombination value between two randomly chosen loci from the family, and reduces to 1 when $n = 2$.

The rate of decay of variability or the rate of increase of identity coefficients may be obtained from the eigenvalues of the **A** matrix, which satisfy the following cubic equation.

$$x^3-(3-a_1)x^2+(3-2a_1+a_2)x-\left\{1-a_1+a_2-\frac{a}{2N}\left(a_1-\frac{1}{2N}-\alpha\right)\right\}=0, \tag{4}$$

where

$$a_1 = (n+1)\alpha+\frac{1}{N}+\frac{(n+1)}{3}\beta$$

and

$$a_2 = \frac{(n+2)}{2N}\alpha+\frac{n(n+1)}{3}\alpha\beta+n\alpha^2+\frac{(n+1)}{6N}\beta+\left(\frac{1}{2N}\right)^2.$$

By using the subroutine DLOWP of FACOM M140 at the National Institute of Genetics, the three roots of the above equation were numerically obtained. Of particular importance is the largest root, $\lambda_{\max}$, and the rate of steady decay becomes

$$k = 1-\lambda_{\max}. \tag{5}$$

Time until fixation of an ordinary mutant gene is $4N$, as Kimura & Ohta (1969*a*) have shown (see also Ewens, 1979). This value is twice the reciprocal of the rate of steady decay. Thus it is conjectured that the time until fixation roughly becomes

$$T \approx 2/k \tag{6}$$

for the case of multigene families. In the next section, the expectation based on the conjecture is examined by Monte Carlo simulation.

When there is no recombination, the equation (4) can be explicitly solved, and the three roots become

$$\lambda_1 = 1-1/(2N), \quad \lambda_2 = 1-\alpha \quad \text{and} \quad \lambda_3 = 1-n\alpha-1/(2N). \tag{7}$$

Therefore, the time until fixation is determined either by λ_1 or λ_2, depending on their magnitudes, and we have

$$T = 4N, \quad \text{for} \quad 1/(2N) < \alpha$$

and

$$T = \frac{2}{\alpha} = \frac{n-1}{\lambda}, \quad \text{for} \quad 1/(2N) > \alpha. \tag{8}$$

The former is the same as the time until fixation of an ordinary mutant, and the latter is that in a single chromosomal line (see Nagylaki & Petes, 1982, for exact value). The result implies that, in the former, the mutant spreads on the chromosome before a chromosomal line spreads in the population, and the time is therefore determined by the rate of random drift in the population, whereas in the latter the rate of spreading of the mutant on the chromosome is the critical factor for the fixation time.

3. MONTE CARLO EXPERIMENTS

The simulated population has $2N$ chromosomes, each with n tandem genes (units), and each experiment starts with the introduction of a mutant gene in one chromosome of the population. The fate of the mutant gene is traced until it is lost from the population, or it is fixed in all loci of the family of all chromosomes in the population, i.e. the total number becomes $2Nn$. The number of generations until loss, and that until fixation are recorded, and the experiments are repeated until the number of cases of fixation reaches to a specified value.

One generation consists of gene conversion, inter-chromosomal recombination and random sampling of gametes for reproduction. Gene conversion was carried out, for each chromosome with probability $n\lambda$, by randomly choosing two units, and one of them, again randomly chosen, is converted by the other one. Recombination was performed, between two randomly chosen chromosomes, at the randomly determined site. The recombination was repeated $N(n-1)\beta$ times. Sampling was done as in the ordinary Monte Carlo studies of population genetics (*e.g.* Kimura & Ohta, 1969a).

Tables 1 and 2 show the results of simulations to compare with expectation. (equation (6)). Table 1 gives the comparison for various values of N and β. Note that, from the theory in the previous section, the important factors are $N\lambda$, $N\beta$ and n. Therefore the values of N are too small and those of λ and β too large; however, their products are thought to be realistic. From the tables it is clear that the agreements between the expected and the observed values are satisfactory. Standard deviation of the time is also given in the tables, and it is noted that this is about half of the mean as in the case of a single locus (Kimura & Ohta 1969*b*).

Table 1. *Comparison of the expected and the observed times until fixation for various values of* N and β

The expected value is obtained from the largest root of equation (4), i.e. by equation (6). N is the effective population size, n is the number of tandem genes on a chromosome, λ is the rate in one generation, at which any unit is converted by the homologous unit, and β is the inter-chromosomal recombination rate between the adjacent loci in the gene family.

N	Exp.	Obs.	Exp.	Obs.
	$\beta = 0$		$\beta = 0{\cdot}025$	
10	80·0	85·9 ± 43·2	121·4	122·0 ± 61·4
20	80·0	125·8 ± 53·0	187·7	202·4 ± 104·7
30	120·0	159·1 ± 62·5	258·5	356·6 ± 170·0
40	160·0	189·5 ± 61·2	330·8	368·3 ± 184·3
50	200·0	223·3 ± 103·0	403·6	493·5 ± 274·3
	$\beta = 0{\cdot}05$		$\beta = 0{\cdot}075$	
10	144·7	148·4 ± 71·9	161·1	152·0 ± 66·0
20	235·3	208·5 ± 113·7	268·1	269·5 ± 129·6
30	329·9	360·9 ± 205·3	377·7	386·6 ± 235·6
40	423·7	429·0 ± 198·6	488·0	461·6 ± 273·2
50	518·8	501·1 ± 191·2	598·6	775·5 ± 486·2

Parameters $n = 5$, $\lambda = 0{\cdot}05$ and 20 replications.

Table 2. *Comparison of the expected and the observed times until fixation for various values of* λ *and* β

	$\beta = 0{\cdot}00625$		$\beta = 0{\cdot}01875$		$\beta = 0{\cdot}0375$	
λ	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.
0·02	393·0	372·2 ± 189·9	551·0	485·0 ± 329·4	669·0	563·8 ± 234·2
0·04	268·5	290·5 ± 145·9	384·1	357·4 ± 186·4	489·0	499·2 ± 278·9
0·06	228·0	269·7 ± 130·8	316·9	315·6 ± 157·9	406·8	406·4 ± 210·6
0·08	208·8	263·9 ± 157·8	280·3	269·5 ± 156·1	357·9	377·4 ± 177·4
0·10	197·9	194·2 ± 106·2	257·3	226·2 ± 96·0	325·2	303·1 ± 160·0

Parameters; $N = 40$, $n = 5$ and 40 replications.

4. NUMERICAL EXAMPLES

In order to find out the relationship between the time until fixation and various parameter values in more detail, numerical calculations are carried out for some interesting cases. Figs. 1–3 show the results. Fig. 1 represents the relationship between $N\lambda$ and the time until fixation (T). From the figure it is clear that T rapidly decreases as Nλ increases. The minimum value of T is $4N$ – that is, the time until fixation of an ordinary mutant gene (Kimura & Ohta, 1969*a*) – and is expected when $N\beta = 0$ and $N\lambda \gg 1$.

Fig. 2 shows the relationship of T and $N\beta$. As $N\beta$ becomes larger, T increases. However when $N\lambda$ is small, the upper limit of T seems to be rapidly attained with

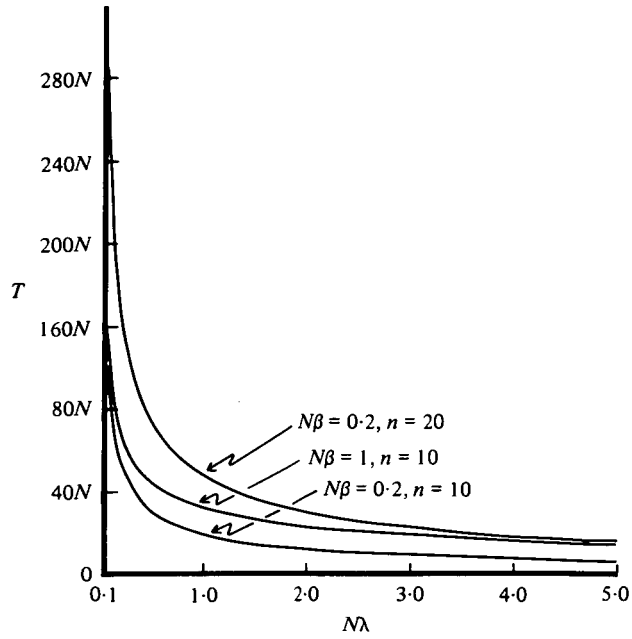

Fig. 1. Relationship between $N\lambda$ and the time until fixation (T).

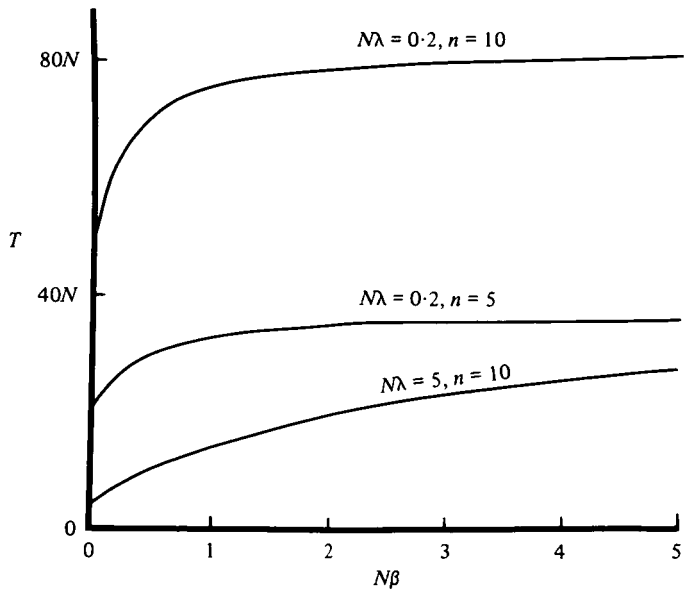

Fig. 2. Relationship between $N\beta$ and the time until fixation (T).

https://doi.org/10.1017/S0016672300021054 Published online by Cambridge University Press

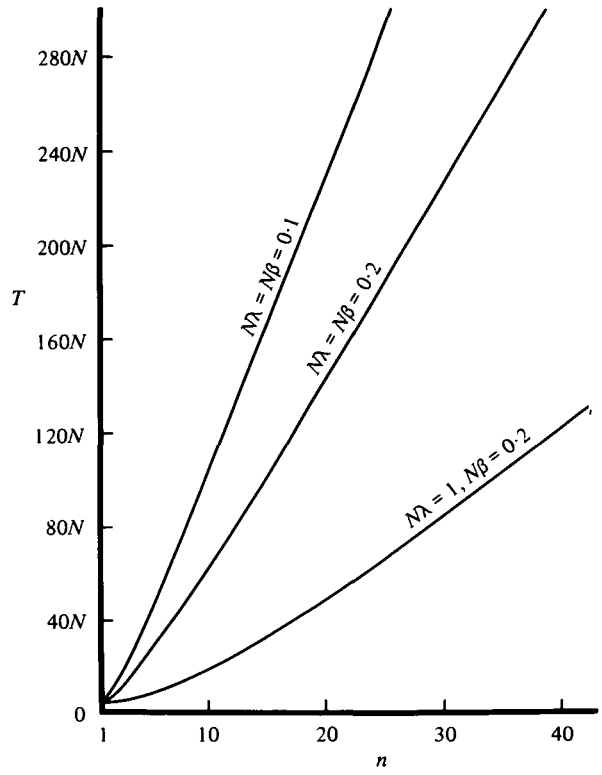

Fig. 3. Relationship between the time until fixation (T) and the number of genes on the chromosome (n).

increasing $N\beta$. When $N\lambda$ is large, T increases slowly, but the maximum is not yet reached in the figure. The situation would be comparable to the model of unequal crossing-over (Ohta, 1980), since the previous equations for equilibrium identity coefficients suggest that genetic variation increases more and more by increasing the rate of inter-chromosomal recombination. In other words, when $N\lambda$ is large, the allelic identity (f) approaches to the non-allelic identity (C_2), and the present model of gene conversion becomes closer to the previous one of unequal crossing-over. Under such a condition, inter-chromosomal recombination is very effective in retarding the decay of genetic variability. The present analysis also suggests that the previous result on identity coefficients for the model of unequal crossing-over is biased when the rate of unequal crossing-over is low and the shift of positions of loci on the chromosome is infrequent compared to the rate of random drift.

Fig. 3 shows the relationships between T and the number of genes on the chromosome (n). The figure indicates that, beyond certain value of n, T increases almost linearly with n, and that the slope is determined by $N\lambda$ and $N\beta$. Note that the point at $n = 1$ is $T = 4N$, therefore the figure gives a picture how increase in copy number is effective in maintaining genetic variation.

5. DISCUSSION

As shown in the previous sections, the simple method of using the rate of steady decay of genetic variation is very useful for estimating the time until fixation of a mutant in multigene families. In studying the effectiveness of over dominance on maintenance of genetic variability, Robertson (1962) called the reciprocal of the rate of steady decay 'retardation factor'. When selection is involved, the present method of estimating the time may not be appropriate, and the concept of retardation factor rather than fixation time should be used in the analyses. For selectively neutral case with no polarity in coversion, the rate of steady decay is used for estimating the time until fixation, since the time seems to be most useful to experimental geneticists (Coen, Thoday & Dover, 1982: Dover 1982).

When there is no interchromosomal recombination ($N\beta = 0$), the present estimate may be biased in some cases as explained below. If $4N \gg 2/\alpha$ or $4N \ll 2/\alpha$, the estimate should give a reasonable value. However, when the two parameters are similar, our result is likely to give an underestimate. This is because, in the process of spreading into the population of a chromosomal line that contains the original mutant, differentiation occurs among the descendant chromosomes, and the complete fixation is attained only after all descendant chromosomes contain n copies of the mutant. This is the case where the two eigenvalues, λ_1 and λ_2 (equation (7)), take similar values. When $N\beta > 0$, the three eigenvalues are all different and our estimate seems to be unbiased.

In discussions on species comparisons of gene families, the possible polarity of the conversion rates seems to be a most important problem (Dover *et al.* 1982; Coen *et al.* 1982). The present result is useful for estimating how rapidly turnover of gene members occurs under no polarity. Care has to be taken here since the process is stochastic and accompanies a large variance. As stated before, the standard deviation of the time until fixation is roughly half of the mean, therefore even if the time of species separation appears to be too short to explain the divergence of the gene families, the data may be an outcome of chance.

Finally, the correspondence between the present model of gene conversion and the previous one of unequal crossing-over needs to be mentioned. The parameter, $\alpha = 2\lambda/(n-1)$, roughly corresponds to $\alpha + \alpha'$ of the model of unequal crossing-over, in which

$$\alpha = m\gamma/n^2$$

and

$$\alpha' = m'\gamma'/n^2,$$

where γ and γ' are rates, and m and m' are the mean numbers of genes shifted, at intra-and inter-chromosomal unequal crossing-overs respectively (see Ohta, 1980). The correspondence is not exact, because of the shift of positions of the loci on the chromosome in the model of unequal crossing-over. As an example, let us consider the following case of a family of 200 tandem genes ($n = 200$) in a population with $N = 5 \times 10^4$. If $\gamma = 10^{-3}$ and $m = 20$ for intra-chromosomal unequal crossing-over, and $\gamma' = 10^{-4}$, $m' = 20$ and $(n-1)\beta \approx n\beta = 0{\cdot}7 \times 10^{-4}$ (see

Ohta, 1980, page 32 for the relationship of γ' and $n\beta$) for inter-chromosomal unequal crossing-over, we have $2N\lambda \approx 5{\cdot}5$ and $2N\beta \approx 0{\cdot}035$. Then the mean time until fixation of a mutant is roughly $205N \approx 10^7$ generations. This would correspond to the turnover of ribosomal gene family of some species of *Drosophila* (Coen *et al.* 1982). The problem of this gene family, however, is that genes are on the sex chromosomes, X and Y, and may not be treated as a unison. In general, the turnover of dispersed repeated sequences is difficult to assess, and it is possible that directional gene conversion or transposition may be responsible for their evolution (Dover 1982).

I thank Dr A. Robertson, Dr T. Nagylaki and Dr. G. Dover for their many valuable suggestions on the manuscript. Supported by Grant-in-Aids 57/20009 from the Ministry of Education, Science and Culture of Japan.

REFERENCES

Coen, E. S., Thoday, J. M. & Dover, G. (1982). Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster*. *Nature* 295, 564–568.

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

Dover, G., Brown, S., Coen, E., Dallas, J., Strachan, T. & Trick, M. (1982). The dynamics of genome evolution and species differentiation. In *Genome Evolution* (ed. G. A. Dover and R. B. Flavell, pp. 343–374. London: Academic Press.

Ewens, W. (1979). *Mathematical Population Genetics*. Berlin: Springer-Verlag.

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

Kimura, M. & Ohta, T. (1969*b*). The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* **63**, 701–709.

Kimura, M. & Ohta, T. (1979). Population Genetics of multigene family with special reference to decrease of genetic correlation with distance between gene members on a chromosome. *Proceedings of the National Academy of Sciences of the U.S.A.* **76**, 4001–4005.

Miyata, T. Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M. & Honjo, T. (1980). Nucleotide sequence divergence of mouse immunoglobulin γ_1 and γ_{2b} chain genes and the hypothesis of intervening sequence-mediated domain transfer. *Proceedings of the National Academy of Sciences of the U.S.A.* **77**, 2143–2147.

Nagylaki, T. & Petes, T. D. (1982). Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* **100**, 315–337.

Ohta, T. (1978). Theoretical study on genetic variation in multigene families. *Genetical Research* **31**, 13–28.

Ohta, T. (1980). *Evolution and Variation of Multigene Families*, Lecture Notes in Biomathematics, no. 37. Berlin, New York: Springer-Verlag.

Ohta, T. (1982*a*). Allelic and non-allelic homology of a supergene family. *Proceedings of the National Academy of Sciences of the U.S.A.* **79**, 3251–3254.

Ohta, T. (1982*b*). Further study on genetic correlation between gene members of a multigene family. *Genetics* **99**, 555–571.

Robertson, A. (1962). Selection for heterozygotes in small populations. *Genetics* **47**, 1291–1300.

Slightom, J. L., Blechl, A. E. & Smithies, O. (1980). Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequence suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**, 627–638.

Yamawaki-Kataoka, Y., Nakai, S., Miyata, T. & Honjo, T. (1982). Nucleotide sequences of gene segments encoding membrane domains of immunoglobulin γ chains. *Proceedings of the National Academy of Sciences of the U.S.A.* **79**, 2623–2627.