J Mol Evol (1984) 20:274–280

Journal of
Molecular Evolution

Population Genetics Theory of Concerted Evolution and Its Application to the Immunoglobulin V Gene Tree[1]

Tomoko Ohta

National Institute of Genetics, Mishima, 411, Japan

Summary. The previous simple model for treating concerted evolution of multigene families has been revised to be compatible with various new observations on the immunoglobulin variable region family and other families. In the previous model, gene conversion and unequal crossing-over were considered, and it was assumed that genes are randomly arranged on the chromosome; neither subdivision nor correlation of gene identity and chromosomal distance were considered. Although this model satisfactorily explains the observed amino acid diversity within and between species, it fails to predict the very ancient branching of the mouse immunoglobulin heavy chain V-gene family. By incorporating subdivided structure and genetic correlation with chromosomal distance into the simple model, the date of divergence may be satisfactorily explained, as well as the rate of nucleotide substitution and the amino acid diversity. The rate at which a V-gene is duplicated or deleted by conversion or by unequal crossing-over is estimated by the new model to be on the order of 10^{-6} per year. The model may be applicable to other multigene families, such as those coding for silkmoth chorion or mammalian kallikrein.

Key words: Multigene family — Subfamily — Gene family branching — Amino acid divergence — Nucleotide divergence — Time for spreading of a gene copy

Introduction

It has been established that multigene families evolve in concert as sets, and that the mechanisms responsible for this phenomenon are unequal (but homologous) crossing-over, gene conversion, and duplicative transposition (for reviews, see Kedes 1979; Long and Dawid 1980; Dover 1982; Jones and Kafatos 1982; Arnheim 1983; Ohta 1983a). Among various multigene families, the variable region gene families of immunoglobulins represent some of the most interesting and best studied examples (Hood et al. 1975; Honjo 1983), and a population genetics model for treating them has been developed (Ohta 1980a, 1983a).

The model considers gene conversion and unequal crossing-over as mechanisms of concerted evolution. The transient and equilibrium properties of identity coefficients were formulated by assuming constant gene family size and finite population size (see Ohta 1983a for a review). Application of these results to the within- and between-species amino acid diversity of reported sequences of immunoglobulin (Ig) variable regions compiled by Kabat et al. (1976) led to the conclusion that the observed amino acid diversity may be well explained under a set of reasonable parameter values for both the framework and the hypervariable regions, and that the rate of amino acid substitution in the latter is about three times higher than that in the former region (Ohta 1978, 1980a,b). In particular, the rate of amino acid substitution in the hypervariable region (now called the complementarity-determining region) is roughly the same as that in fibrinopeptides, which are known to have the highest evolutionary rate so far determined among proteins. This

[1] Contribution no. 1560 from the National Institute of Genetics, Mishima, 411 Japan

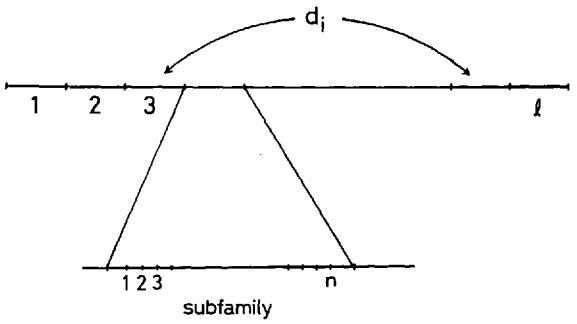

Fig. 1. Diagram showing the model of subdivided structure of a multigene family. A family consists of ℓ subfamilies, and a subfamily, of n gene units. d_i, identity coefficient between genes belonging to subfamilies i steps apart on the same chromosome

finding implies that there is little selective constraint on this region in the light of the neutral theory of molecular evolution (see Kimura 1983 for a review).

The population genetics theory of concerted evolution has now been extended to estimate the time until fixation of a mutant gene copy, i.e., the time necessary for spreading of a single gene into all copies of a multigene family of a species (Ohta 1983b; Ohta and Dover 1983). It is expected that the time is extremely long for large multigene families with variable gene members, such as IgV families (V is a part of the variable region excluding the D and J segments; see Honjo 1983 for a review). On the other hand, the rapid accumulation of DNA sequence data has enabled the construction of gene family trees, and Gojobori and Nei (1984) successfully worked out a phylogenetic tree of the IgV genes of heavy (H) chains of mouse and human. An important finding of their study is that the earliest divergence of the IgV family of mouse H can be estimated to be approximately 3×10^8 years ago. This implies that the time for the spreading of a gene copy into the IgV family is of about this length. My previous model and estimated parameter values cannot explain such a long time for the spreading of a single gene copy.

In this article, I present some modifications of the previous simple model of concerted evolution that enable it to explain the long period of time for expansion of a gene copy, as well as the observed nucleotide substitution rate and amino acid diversity of Ig sequences of mouse and human.

Model and Analyses

In the previous simple models of unequal crossing-over (Ohta 1980a) and gene conversion (Ohta 1982, 1984; Nagylaki 1984), it was assumed that there is neither correlation between gene identity and chromosomal distance nor subdivision of the gene members of a multigene family into several subgroups. On the other hand, it is very clear that there do exist several subfamilies within such multigene families as human IgVH, mouse IgVκ (Kabat et al. 1976), and silkmoth chorion (Kafatos 1983). It is also expected that those genes that are tightly linked on the chromosome are similar to each other. Also, predictions based on the simple model disagree with observed facts, as reflected in the present problem of the excessively ancient (compared with the predictions of the old model) branching of the IgV gene tree. Thus, it is necessary to improve the simple model to take subdivision and genetic correlation with chromosomal distance into account. This is what we have done below.

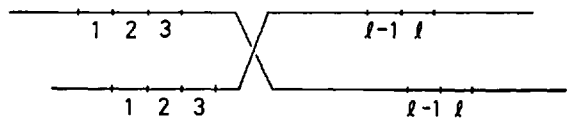

Fig. 2. Model of unequal crossing-over. The lines represent sister chromatids

Let us consider a multigene family comprising ℓ subfamilies tandemly arranged on the chromosome, with each subfamily consisting of n tandemly arranged genes. Figure 1 shows such a multigene family. It is assumed that there is no genetic correlation within a subfamily, i.e., that the previous simple model applies within a subfamily. I use this conversion model within each subfamily, letting λ be the rate at which a gene is converted by one of the remaining $(n - 1)$ genes belonging to the same subfamily. It is also assumed that conversion is asymmetric and not biased. (For a more general model of conversion, see Nagylaki and Petes 1982; Nagylaki 1984; Ohta 1984.)

In addition to gene conversion within each subfamily, the multigene family is assumed to be evolving by unequal crossing-over; this is responsible for the concerted evolution of genes belonging to different subfamilies. For our purpose, I modify the model of Kimura and Ohta (1979), in which the shift at unequal crossing-over is one gene unit and therefore gene identity is inversely correlated with distance on the chromosome. In the present model, instead of one gene unit, the shift at unequal crossing-over is assumed always to be *one subfamily*, and all unequal crossing-overs are assumed to be intrachromosomal. Figure 2 shows such an unequal crossing-over. As in the model of Kimura and Ohta (1979), the probability of duplication and that of deletion are equal, i.e., 1/2. In this model the total number of subfamilies may change, but it is possible to assume constant family size, since the identity coefficients simply depend on the chromosomal distance between the two subfamilies, just as in the

case of Kimura and Ohta (1979). Let γ be the rate of unequal crossing-over per multigene family in one generation.

Within-subfamily gene conversion and between-subfamily unequal crossing-over are the two important processes for concerted evolution in this model. They are assumed to take place *within a chromosome*. Now we consider concerted evolution at the level of the population (species). Let the effective population size be N; i.e., there are 2N homologous multigene families, and therefore a total of $2N\ell n$ genes. The population is evolving under mutation, random genetic drift, and interchromosomal recombination, in addition to gene conversion and unequal crossing-over. Let v be the mutation rate per generation per unit whose identity is to be compared. The infinite allele model (Kimura and Crow 1964) is assumed; i.e., every mutation is to a new, not a preexisting, state. The unit whose identity is to be compared may be a nucleotide site, an amino acid site, or an exon region, and the mutation rate must be appropriately chosen according to the unit length.

All interchromosomal recombinations at meiosis are assumed to be equal. Let β be the rate of recombination between adjacent subfamilies. Recombination also occurs within subfamilies; let b be the rate of recombination between adjacent loci within a subfamily. In the present analysis, interchromosomal unequal crossing-over is not considered. When it occurs, it is effective in duplication/deletion of genes like intrachromosomal unequal crossing-over. It is also effective in recombining genes between chromosomes as is interchromosomal equal crossing-over (see Ohta 1981 for analyses of interchromosomal unequal crossing-over).

To formulate concerted evolution of the IgVH gene family, the following set of identity coefficients is needed. The within-subfamily coefficients are given the same labels as before (Ohta 1982, 1983a, 1984; Nagylaki 1984): f is the allelic identity, C_1 is the coefficient of identity between genes of the same subfamily on the same chromosome, and C_2 is that between nonallelic genes on different chromosomes but belonging to the same subfamily. In addition to these three, the coefficients between subfamilies need to be defined. Let d_i be the coefficient between subfamilies i steps apart on the same chromosome (see Fig. 1) and g_i that between subfamilies on different chromosomes. It is possible to formulate the equations of transition of gene identity from one generation to the next in terms of the above set of identity coefficients. In these formulations, it is assumed that all evolutionary forces are weak; i.e., v, λ, γ, b, β, $1/N \ll 1$.

The changes of identity coefficients by conversion may be obtained in a manner similar to that for the simple conversion model (Ohta 1982, 1983a, 1984; Nagylaki 1984). Through within-subfamily conversion with the rate λ, the first three identity coefficients change according to the following equations, where we assume that $\lambda \ll 1$:

$$\left.\begin{aligned} \Delta_{conv}(f) &= 2\lambda(C_2 - f) \\ \Delta_{conv}(C_1) &= \frac{2\lambda}{n-1}(1 - C_1) \\ \Delta_{conv}(C_2) &= \frac{2\lambda}{n-1}(f - C_2), \end{aligned}\right\} \quad (1)$$

where $\Delta_{conv}(x)$ is the expected change resulting from within-subfamily conversion.

The expected coefficient changes resulting from random genetic drift are given by

$$\left.\begin{aligned} \Delta_{drift}(f) &= \frac{1}{2N}(1 - f) \\ \Delta_{drift}(C_2) &= \frac{1}{2N}(C_1 - C_2) \\ \Delta_{drift}(g_i) &= \frac{1}{2N}(d_i - g_i) \end{aligned}\right\} \quad (2)$$

The changes caused by interchromosomal recombination are as follows. The within-subfamily coefficients are assumed to be independent of chromosomal distance, and the average change of C_1 has been shown to be

$$\Delta_{rec\cdot w}(C_1) = \frac{(n+1)}{3}b(C_2 - C_1) \quad (3)$$

where $\Delta_{rec\cdot w}(x)$ is the expected change by within-subfamily recombination (Ohta 1983a). Applying the method of Kimura and Ohta (1979) to the between-subfamily coefficients, one gets

$$\Delta_{rec\cdot b}(d_i) = i\beta(g_i - d_i) \quad (4)$$

where $\Delta_{rec\cdot b}(x)$ is the expected change resulting from between-subfamily recombination.

The expected identity coefficient changes resulting from mutation are given simply by

$$\left.\begin{aligned} \Delta_{mut}(f) &= -2vf \\ \Delta_{mut}(C_1) &= -2vC_1 \\ \Delta_{mut}(C_2) &= -2vC_2 \\ \Delta_{mut}(d_i) &= -2vd_i \\ \Delta_{mut}(g_i) &= -2vg_i \end{aligned}\right\} \quad (5)$$

Finally, the change caused by unequal crossing-over is calculated by the method of Kimura and Ohta (1979) and of Ohta (1980a, 1981). Noting that our subfamily corresponds to their gene unit at unequal crossing-over (see Fig. 2), we see that the shift of positions results in the following equations [see Eq. (5) of Kimura and Ohta 1979 and Eq. (3') of Ohta 1981]:

$$\left.\begin{aligned}\Delta_{uneq}(d_i) &= \frac{i\gamma}{\ell}\left(\frac{d_{i-1} + d_{i+1}}{2} - d_i\right)\\ \Delta_{uneq}(g_i) &= \gamma\left(\frac{g_{i-1} + g_{i+1}}{2} - g_i\right)\end{aligned}\right\} \quad (6)$$

for $1 \leq i \leq \ell - 1$, where $\Delta_{uneq}(x)$ is the expected change resulting from unequal crossing-over. However, special caution is needed in dealing with the terminal classes ($i = 1$ and $i = \ell - 1$). It can be seen that d_0 and g_0 may be expressed by the formulas

$$\left.\begin{aligned}d_0 &= \frac{1}{n}[1 + (n - 1)C_1]\\ g_0 &= \frac{1}{n}[f + (n - 1)C_2]\end{aligned}\right\} \quad (7)$$

which may be used in Eq. (6) when $i = 1$. When $i = \ell - 1$, I follow Kimura and Ohta (1979) and let $d_{\ell-1} = d_\ell$ and $g_{\ell-1} = g_\ell$.

In addition to the changes given above by Eqs. (6) and (7), the within-subfamily coefficients need to be transformed; f and C_2 change by unequal crossing-over according to the equations

$$\left.\begin{aligned}\Delta_{uneq}(f) &= \gamma(g_1 - f)\\ \Delta_{uneq}(C_2) &= \gamma(g_1 - C_2)\end{aligned}\right\} \quad (8)$$

Equations (1)–(8) give the changes in the set of identity coefficients in one generation. The solution obtained by setting the changes to zero gives the identity coefficients at equilibrium, i.e., when the various evolutionary forces balance each other. The numerical results are given in the next section. The average values may be compared with the observed values of amino acid identity of Ig sequences. By using Eqs. (1)–(8), it is also possible to estimate the time until fixation of a single gene copy within a multigene family. The method uses the rate of steady decay of genetic variability (see Ohta 1983b; Ohta and Dover 1983; Nagylaki 1984). The results of this numerical study are also given in the next section, where they are examined to see whether the model satisfactorily explains the extremely long time for spreading of a gene copy discussed earlier.

Numerical Examples That Fit the Observed Facts About the IgV Family

Equilibrium identity coefficients and the time for spreading of a gene copy were numerically obtained for several sets of plausible values of parameters. The rate of amino acid substitution (v under selective neutrality) in the framework region of IgV is estimated to be about 2×10^{-9} per amino acid site per year (Ohta 1980a,b), and that of nucleotide substitution, 1.01×10^{-9} for the first position, 0.65×10^{-9} for the second position, and 2.63×10^{-9} for the third position of a codon (Gojobori and Nei 1984). Since the sum of the substitution rates of the first and the second positions should be roughly equal to the rate of amino acid substitution, these two estimates approximately agree with each other, and this parameter should be taken to be in the range of these two values, i.e., $1.6–2 \times 10^{-9}$ per year. The effective population size (N) in the course of vertebrate evolution is usually assumed to be 10^4–10^5, based on data on enzyme polymorphisms (Kimura and Ohta 1971; Nei 1975).

The number of subfamilies (ℓ) in an IgV family is conjectured to be a few tens, with subfamily size (n) on the order of 10 (Honjo 1983). The unknown parameters are the rates of conversion, unequal crossing-over, and interchromosomal recombination. Of these three, some information is available on the rate of occurrence of interchromosomal recombination. Weigert and Riblet (1977) obtained 14 recombinants from 3414 hybrids between the constant and the variable gene markers. Thus, it is reasonable to suppose that the recombination rate per one gene family ($\ell - 1)\beta$ is on the order of 10^{-3} per generation. The rates of conversion and unequal crossing-over are important in determining the speed of concerted evolution, whereas that of interchromosomal recombination has a minor effect. Here it should be noted that, by assuming almost fixed values of substitution rate and gene family size, and a not very flexible population size, the permissible values of the rates of conversion and unequal crossing-over become very limited if the predicted value of the amino acid diversity of the IgV framework region of human or mouse is to approximate that observed. Our problem is to examine the possibility of choosing a set of parameter values that simultaneously explains the observed amino acid identity (0.75–0.8) and the time for spreading of a single gene copy (3×10^8 years).

The average value of predicted amino acid identity is obtained by the following formulas (Kimura and Ohta 1979):

$$\left.\begin{aligned}\bar{d} &= \frac{2}{\ell(\ell+1)}\left(\ell\frac{1 + (n-1)C_1}{n} + (\ell - 1)d_1 + \ldots + d_{\ell-1}\right)\\ \bar{g} &= \frac{2}{\ell(\ell+1)}\left(\ell\frac{f + (n-1)C_2}{n} + (\ell - 1)g_1 + \ldots + g_{\ell-1}\right)\end{aligned}\right\} \quad (9)$$

where f, C_1, C_2, d_i and g_i are equilibrium values

Table 1. Numerical examples that plausibly explain the time of branching (t_l), the rate of nucleotide substitution (= mutation rate, v), and the degree of amino acid identity of reported sequences of immunoglobulins

Parameter	Case						
	1	2	3	4	5	6	7
ℓ	10		10	20	20		20
Nv	10^{-5}		4×10^{-5}	4×10^{-5}	4×10^{-5}		4×10^{-5}
Nλ	0.005		0.05	0.075	0.1		0.05
Nγ	0.005		0.016	0.075	0.06		0.06
Nb	0.006		0.03	0.06	0.06		0.06
Nβ	0.12		0.6	0.6	0.6		0.6
$\bar{g}$	0.74		0.76	0.80	0.78		0.76
t_l (in N gns.)	42,794N		11,722N	9,690N	11,560N		12,285N
N	6.25×10^3	5×10^3	2.5×10^4	2.5×10^4	2.5×10^4	2.0×10^4	2.5×10^4
v	1.6×10^{-9}	2×10^{-9}	1.6×10^{-9}	1.6×10^{-9}	1.6×10^{-9}	2×10^{-9}	1.6×10^{-9}
λ	8×10^{-7}	1×10^{-6}	2×10^{-6}	3×10^{-6}	4×10^{-6}	5×10^{-6}	2×10^{-6}
t_l (in gns.)	2.67×10^8	2.14×10^8	2.93×10^8	2.42×10^8	2.89×10^8	2.31×10^8	3.07×10^8

See text for explanation of parameters. The total gene number of one multigene family (ℓn) is assumed to be 200. $\bar{g}$, average predicted value of amino acid identity; gns., generations

numerically calculated using the equations in the previous section. It turns out that in the case of multigene families with extremely variable members, e.g., the IgV family, $\bar{d}$ and $\bar{g}$ are almost the same, and thus we will tabulate only $\bar{g}$. The situation is different for gene families with uniform members, such as the rRNA family.

The time until fixation or spreading of a gene copy (t_l) is obtained by the following formula:

$$t_l = \frac{2}{1 - \xi_{max}} \qquad (10)$$

where ξ_{max} is the maximum eigenvalue of the transition equations of the set of identity coefficients given in the previous section (see Ohta 1983b; Ohta and Dover 1983; Nagylaki 1984).

Table 1 gives the results of numerical studies for several sets of parameter values. It should be noted that it is the products such as Nv, Nλ, and Nβ that are important, rather than the individual parameter values. Also, the time until fixation is obtained in units of N generations. Therefore, I assume that one generation is equal to a year in the course of the spreading of a gene copy in mouse or human IgV families. This assumption does not hold for contemporary species, but for a very long period, on the order of 10^8 years, it must have been valid. The number of subfamilies is assumed to be either 10 or 20, with the total gene number held constant at ℓn = 200. These values are plausible estimates for IgV families (Honjo 1983).

The values Nλ, Nγ, and Nv are crucial in determining the within-species gene identity ($\bar{g}$) and the time for the spreading of a single gene copy (t_l), whereas Nβ and Nb have a relatively minor effect on t_l and $\bar{g}$. In other words, when the total gene number (ℓn) is fixed, mainly Nλ and Nγ determine t_l, and the rate of accumulation of mutations in those gene lineages that are spreading into the multigene family depends on Nv. Cases 1–4 of Table 1 show that, for small values of N, all three products Nv, Nλ, and Nγ must be correspondingly small to predict realistic values of t_l and $\bar{g}$. Also, it can be seen, from cases 1 and 2 or cases 5 and 6 of Table 1 that when N is varied but Nv, Nλ, Nγ, Nβ, and Nb are held constant, different predictions of t_l are obtained that are positively correlated with N. Note that N can be varied only in a small range, since the value of v is in the range $1.6–2.0 \times 10^{-9}$, as discussed earlier. In general, the observed values of t_l and $\bar{g}$ can be explained by assuming that the within-subfamily conversion rate per gene (λ) and the between-subfamily unequal crossing-over rate per family (γ) are both on the order of 10^{-6} per year (or generation, in our case).

Table 2. Predicted equilibrium identity coefficients (amino acid identity) within a subfamily

Coefficient	Case				
	1, 2	3	4	5, 6	7
f	0.997	0.983	0.981	0.983	0.980
C_1	0.909	0.920	0.952	0.963	0.935
C_2	0.909	0.918	0.948	0.959	0.932

Cases are the same as in Table 1. f, allelic identity; C_1 and C_2, nonallelic identities

Let us examine in more detail the expected identity coefficients (amino acid identity) at equilibrium. Table 2 gives within-subfamily identity coefficients (f, C_1, and C_2) of the same examples as in Table 1. From Table 2, it can be seen that the identity within a subfamily is fairly high (more than 90% for C_1 or

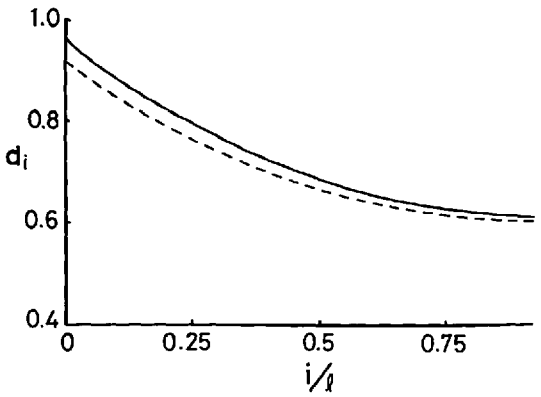

Fig. 3. Identity coefficients d_i as functions of chromosomal distance (i/ℓ). ___, case 3 of Table 1; ——, case 5 or 6

C_2). A more notable fact is that the allelic amino acid identity is expected to be 98% or more. In other words, if the amino acid sequences of the same locus of a population are compared, the amino acid difference is expected to be 2% or less; therefore, polymorphisms should not be very extensive. This prediction is quite different from that of the previous model (Ohta 1982), which assumed no correlation between genetic identity and chromosomal distance, and which predicts extraordinary polymorphisms at the loci of histocompatibility antigens. But polymorphisms are predicted to be more extensive in our multigene case than in the ordinary single-locus model.

Figure 3 shows the amino acid identity between the subfamilies (d_i) as a function of chromosomal distance (i/ℓ). With the present sets of parameter values, g_i values are almost the same as those of d_i, and hence are not shown. Two cases (case 3 and case 5 or 6 of Table 1) are shown. As can be seen from the figure, identity decreases with chromosomal distance between the two subfamilies. This situation is analogous to the results of Kimura and Ohta's (1979) model, in which the shift at unequal crossing-over was one gene unit, and contrasts with the prediction of the simple model (Ohta 1982, 1983a; Nagylaki 1984), which assumed no correlation between the gene identity and the chromosomal distance.

Discussion

The present analyses have clearly shown that, by incorporating subdivision and gene correlation into the previous simple model, various observed facts such as amino acid diversity, nucleotide substitution rate, and the ancient branching of the IgV gene tree may be satisfactorily explained. Nevertheless, the model analyzed here is too unrealistic in many respects. Most importantly, natural selection is not considered. It is evident that accumulation of pseudogenes in the family is prevented by natural selection. Selection is considered to operate in eliminating those individuals who have too many pseudogenes (negative selection). In fact, about one-third of all genes appear to be pseudogenes (Gojobori and Nei 1984). On the other hand, positive natural selection may be working to increase gene diversity in IgV families. I have presented elsewhere some models showing that natural selection may efficiently increase or decrease the gene diversity of a multigene family, but have not solved them quantitatively (see Ohta 1980a, pp 111–116). The effect of natural selection on the amino acid diversity of IgV families is left to future investigation. The present study is an attempt to examine to what extent the observed diversity of Ig sequences may be explained without assuming positive natural selection.

In our analyses, gene family size (ℓn) is assumed to be constant, and natural selection is considered to be responsible for the constancy. Therefore, the rate of unequal crossing-over (γ) is the effective rate; i.e., it does not include those unequal crossing-overs that result in a too small or too large gene family size and hence are eliminated by natural selection (see Ohta 1981 for a discussion on this type of natural selection).

Not only are problems posed for the model by natural selection, but our assumptions that gene conversion is restricted to genes belonging to the same subfamily and that the shift at unequal crossing-over is always one subfamily in length may be too idealistic. In practice, however, the model is applicable to cases in which the rates of such gene interactions as conversion and unequal crossing-over are much higher within a subfamily than between subfamilies and the rate of between-subfamily interaction decreases with chromosomal distance. Even if the details of the theory need modification depending on the characteristics of the model used, the predictions of gene identity and of the time for spreading of a gene copy should be valid approximations. It should also be noted that the prediction is slightly different from that of the simple model for a more explicit and general model of conversion, even if neither subdivision nor genetic correlation with chromosomal distance is taken into account (see Nagylaki and Petes 1982; Nagylaki 1984; Ohta 1984). Furthermore, in our present model, no bias in conversion is assumed.

Another limitation of our model is that it assumes no recombination *within* the V-region; i.e., the occurrence of mosaic genes is assumed negligible. Otherwise, the length of the oldest branch of the phylogenetic tree would not correspond to the time for the spreading of a single gene copy. This assumption appears to be valid for genes belonging to different

subfamilies, but not for those of the same subfamily (Ohta 1980c; Gojobori and Nei 1984). As long as the creation of mosaic genes by recombination between different subfamilies is negligible, the correspondence between the branch length and the time for spreading should hold. Still another problem is the question of whether the observed amino acid identity is influenced by somatic mutation (Tonegawa 1983). In terms of identity coefficients, however, the effect of somatic mutation seems to be quite small, even if it is effective in increasing the diversity of immunoreaction. In fact, Gojobori and Nei (manuscript in preparation) found, by comparing myeloma and germ-line nucleotide sequences reported by various investigators, that the proportion of nucleotide diversity due to somatic mutation is at most a few percent in the framework regions.

The present model may be applicable not only to IgV gene families but also to other multigene families with variable gene members, such as the chorion genes of silkmoth (Jones and Kafatos 1982; Kafatos 1983) and mammalian kallikrein genes (Mason et al. 1983). In these examples, subdivision of a multigene family into several subfamilies and functional or developmental differentiation among the subfamilies are observed.

Acknowledgments. I thank Drs. T. Gojobori and Y. Tateno for their many valuable comments on the first draft of this paper. Thanks are also due to Dr. K. Aoki for going over the manuscript and for making many useful suggestions to improve the presentation. This work was supported in part by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan.

References

Arnheim N (1983) Concerted evolution of multigene families. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland, Massachusetts, pp 38–61

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

Gojobori T, Nei M (1984) Concerted evolution of the immunoglobulin V_H gene family. Mol Biol Evol 1:195–212

Honjo T (1983) Immunoglobulin genes. Annu Rev Immunol 1:499–528

Hood L, Campbell JH, Elgin SCR (1975) The organization, expression, and evolution of antibody genes and other multigene families. Annu Rev Genet 9:305–353

Jones WC, Kafatos FC (1982) Accepted mutations in a gene family: evolutionary diversification of duplicated DNA. J Mol Evol 19:87–103

Kabat EA, Wu TT, Bilofsky H (1976) Variable regions of immunoglobulin chains. Medical Computer Systems, Bolt, Beranek and Newman, Cambridge, Massachusetts

Kafatos FC (1983) Structure, evolution, and developmental expression of the chorion multigene families in silkmoths and *Drosophila*. In: Subtelny S, Kafatos FC (eds) Gene structure and regulation in development. Alan R Liss, New York, pp 33–61

Kedes LH (1979) Histone genes and histone messengers. Annu Rev Biochem 48:837–870

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, London New York

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Kimura M, Ohta T (1971) Theoretical aspects of population genetics. Princeton University Press, Princeton

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. Proc Natl Acad Sci USA 76:4001–4005

Long EH, Dawid IB (1980) Repeated genes in eukaryotes. Annu Rev Biochem 49:727–764

Mason AJ, Evans BA, Cox DR, Shine J, Richards RI (1983) Structure of mouse kallikrein gene family suggests a role in specific processing of biologically active peptides. Nature 303: 300–307

Nagylaki T (1984) The evolution of multigene families under intrachromosomal gene conversion. Genetics 106:529–548

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

Nei M (1975) Molecular population genetics and evolution. North Holland, Amsterdam New York

Ohta T (1978) Sequence variability of immunoglobulins considered from the standpoint of population genetics. Proc Natl Acad Sci USA 75:5108–5112

Ohta T (1980a) Evolution and variation of multigene families. Springer-Verlag, Berlin New York (Lecture notes in biomathematics, vol 37)

Ohta T (1980b) Amino acid diversity of immunoglobulins as a product of molecular evolution. J Mol Evol 15:29–35

Ohta T (1980c) Linkage disequilibrium between amino acid sites in immunoglobulin genes and other multigene families. Genet Res 36:181–197

Ohta T (1981) Further study on genetic correlation between gene members of a multigene family. Genetics 99:555–571

Ohta T (1982) Allelic and non-allelic homology of a supergene family. Proc Natl Acad Sci USA 79:3251–3254

Ohta T (1983a) On the evolution of multigene families. Theor Popul Biol 23:216–240

Ohta T (1983b) Time until fixation of a mutant belonging to a multigene family. Genet Res 41:47–55

Ohta T (1984) Some models of gene conversion for treating the evolution of multigene families. Genetics 106:517–528

Ohta T, Dover GA (1983) Population genetics of multigene families that are dispersed into two or more chromosomes. Proc Natl Acad Sci USA 80:4079–4083

Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302:575–581

Weigert M, Riblet R (1977) Genetic control of antibody variable regions. Proc Cold Spring Harbor Symp Quant Biol 41: 837–846