J. Mol. Evol. 15, 181–196 (1980)

Journal of Molecular Evolution

A Computer Simulation of Evolutionary Forces Controlling the Size of a Multigene Family

J.M. Hood[1], H.V. Huang[2], and L. Hood[2]

[1] Mathematics Department, Cal Poly State University, San Luis Obispo, CA 93407, USA
[2] Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Summary. A Monte Carlo-type simulation of the evolution of a multigene family was performed. The model was designed to study the selective forces which may control the size of a multigene family. As expected, we find that direct selection on the size of the multigene family can control its size. More important, we find that selection acting upon the family as a single functional unit, in conjunction with homologous but unequal crossing over, can also control the size of a multigene family.

Key words: Multigene Family, size – Selective forces – Homologous but unequal crossing over – Monte Carlo simulation

Introduction

A multigene family is defined as a group of nucleotide sequences or genes that demonstrate four properties - multiplicity, close linkage, sequence homology, and related or overlapping phenotypic functions (Fig. 1)(Hood, 1972; Hood et al., 1975). Such families have arisen in eukaryotic evolution by employing genetic mechanisms which include mutation, selection, homologous but unequal crossing over and duplication of the gene family in toto or in part. As a chromosomal unit, the multigene family encompasses a broad spectrum of gene families, some of which have simple phenotypic traits, while others are very complex in character (Hood et al., 1975; Table 1).

Hitherto, investigation has focused on the forces that control the gene composition and the equilibrium properties in regard to gene composition of multigene families. We want to open up another inquiry, namely to ask what controls the size of a multigene family, be it one of very homogeneous (e.g., rRNA genes) or very heterogeneous (e.g., antibodies) nature.

0022-2844/80/0015/0181/$ 03.20

Table 1. Classification and properties of multigene families[a]

Category	Gene products	Multiplicity	Gene or protein homology	Information content	Examples of coincidental evolution	Examples of change in family size
Simple-sequence satellites	None known	10^3-10^6	80-100%		Different satellites of *Drosophila*	Mouse satellite
Multiplicational	RNA, protein					
18S-28S RNA		100-600	97-100%	one Unit	Spacer regions of *X. laevis* and *X. mulleri*	
5S RNA		100-1200	97-100%	One unit	Spacer regions of *X. laevis* and *X. mulleri*	Gene number in *X. mulleri* and *X. laevis*
tRNA		6-400		One unit		
histones		10-1200	87-99%	Few units	Histone mRNAs of two species of sea urchin	Gene number in different sea urchin species
Informational	Proteins					
antibodies		10^2-10^3	30-100%	Many units	Rabbit and mouse κ chains	λ_γ and V_κ in mammals
hemoglobins		~10	< 75-100%	Few units	Human and cow δ chains	Human and rabbit β-like genes

[a] In all instances characteristics refer to a single closely linked set of genes. (From Birnstiel et al. 1974)

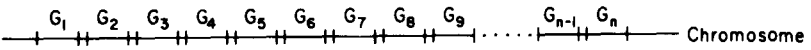

Fig. 1. Representation of a multigene family on a chromosome. Genes (G) may be separated by spacers

What is the Origin of Multigene Families?

The most likely genetic mechanism to explain the origin of multigene families is unequal crossing over (Fig. 2). Unequal crossing over occurs when chromosomes mispair and cross over to yield one chromosome with a duplication and a second with a deletion. With time, unequal crossing over generates a family of closely linked repeats on the chromosome. Thus an existing gene may be duplicated tandemly to form a multigene family. Subsequently, this original family may be duplicated in part or in toto to generate new multigene families. Unequal crossing over may operate upon random DNA sequences to generate de novo repeats (Smith, 1973, 1976; Ohta, 1976, 1978a). The repeats, once generated, may be maintained for long periods of time even in the absence of selective forces. This may be the origin of the satellite DNA families, for which no function is known.

Forces which Control the Properties of Multigene Families

Multigene families may be distinguished from one another by two general properties: the gene composition and the size of the family.

Gene Composition. Certain multigene families such as the ribosomal RNAs and histones contain repeats that are extremely similar to each other (see Table 1; Birnstiel et al., 1974; Ford and Southern, 1973; Tartof and Dawid, 1976). Other families contain repeats that are extremely diverse such as the immunoglobulin gene families.

The evolution and equilibrium properties of the gene composition in multigene families have been well investigated, expecially by Ohta (1977, 1978a,b,c) and Kimura and Ohta (1979). Their studies deal with the influence of unequal crossing over (both

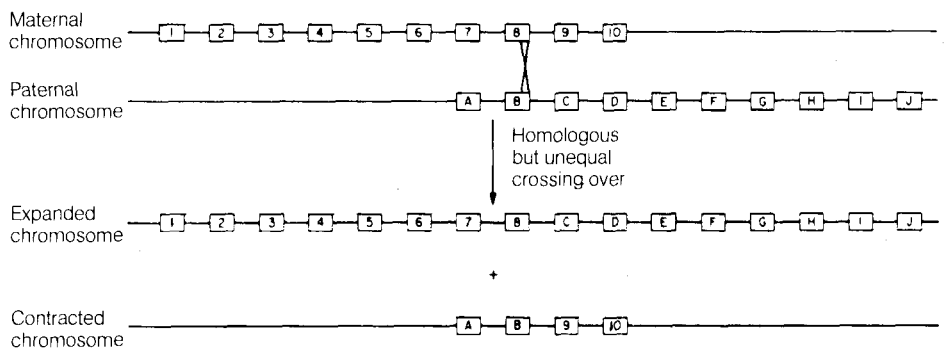

Fig. 2. A model for homologous but unequal crossing over. (From Birnstiel et al., 1974)

intra- and interchromosomal), mutations and selection upon the similarity of genes within a family and the similarity of genes from related but distinct multigene families. Black and Gibson (1974), and Perelson and Bell (1977) have shown that unequal crossing over in the absence of selection may in a multigene family consisting of individually distinguishable genes lead to the loss of some genes and the expanded representation of others. After an adequate interval of time most of the genes in the multigene family form discrete groups of closely related genes. In the extreme case, all genes in a contemporary family are descended from a single ancestral gene. Because of this shared ancestry, multiple genes may appear to evolve coincidentally as a single gene. This phenomenon is called coincidental evolution, and is observed, for example, as species-specific residues in antibody gene families.

Family Size. The size of a multigene family is generally a property independent of the similarity of genes in the family. Thus, informational and multiplicational multigene families may be large or small. A given multiplicational family (e.g., 5S rRNA) may be of different size from individual to individual, and from one sibling species to another (Brown and Sugimoto, 1973). For immunoglobulins, the kappa light chain family is much larger than the lambda light chain family in the mouse, but smaller in the horse (Hood, 1973). Thus multigene families can expand or contract their sizes rapidly in terms of revolutionary time.

What then controls the size of multigene families? Once again, an important mechanism for explaining the gene expansion and contraction is unequal crossing over. When it occurs one chromosome will have an increased number of genes and the other will have a decreased number of genes (Fig. 2). If, in addition, there were selective pressures for an expanded (or contracted) number of genes, those chromosomes with increased (decreased) numbers of genes would be favorably selected in a population. In time the size of the multigene family would change.

Selection may operate directly upon the size of the multigene family. However, our operating hypothesis is that selection cannot act with precision in directly controlling the size of a multigene family. Instead, we show that selection, working upon the multigene family as a single functional unit, is sufficient to control the size of the multigene family.

Structure of the Model

There is no a priori indication as to which evolutionary forces will be necessary and sufficient to control the size of a multigene family. We therefore constructed a relatively detailed model, with the intention of testing as many variables as possible.

The model follows the evolution of a multigene family in diploid populations. These idealized populations consist of up to 10 individuals, with this number being defined as the effective population size (Crow and Kimura, 1970; Nei, 1975). Each individual has two haplotypes or haploid sets of genes which make up the multigene family. The multigene family consists of two classes of genes, functional and nonfunctional. The functional genes are further divided into two subclasses, designated G_1 and G_2. This division of the functional genes into two subclasses serves as an internal check of our mutational model (see below). The class of nonfunctional genes is denoted by N. Nonfunctional genes are considered deleterious (see section D). The three types of genes

are assumed to differ by single event mutations. Thus, G_1, G_2 and N represent sets of genes distinguishable by characteristic base substitutions. The members in each set are not necessarily identical. They may differ from one another by still other genetic differences but these are disregarded. In most of the simulations, G_1 and G_2 are considered selectively equal.

The total number of genes of all types in a given haplotype is defined to be the size of that haplotype. Initially, all haplotypes are assigned 500 functional genes with $G_1:G_2 = 9:1$, and no N genes. Thus each animal starts with a total of 1000 functional genes in its two haplotypes and all animals have identical genotypes. Due to limited computation time, we assume that the N, G_1 and G_2 gene types are uniformly distributed throughout the haplotype, and no attempt is made to follow their linear order during the simulations. Thus, we cannot study coincidental evolution.

The simulations are organized into cycles (Fig. 3) which represent a generation cycle of an animal. A cycle consists of four consecutive segments: mutation, crossing over, breeding, and selection. Each cycle represents one or more generations, depending upon the rates chosen for mutations and crossovers. Each simulation run or 'evolutionary period' consists of 2000 to 10,000 cycles. Typically we perform two replications for each set of starting conditions. The standard parameters used in a normal run are listed in Table 2 and are explained below. The details of each segment in a cycle is described below.

A. Mutations. In most of the simulations, we use the mutation rate of 1.65×10^{-4} functional gene^{-1} cycle^{-1}. Nonfunctional genes, N, enter the system only through mutation. Both types of functional genes, G_1 and G_2, are allowed to mutate to N genes as well as to the other type of functional gene. For the sake of simplicity, half the mutations are assumed to convert functional to nonfunctional genes, the other half result in interconversion of the two functional gene types. Back mutations from

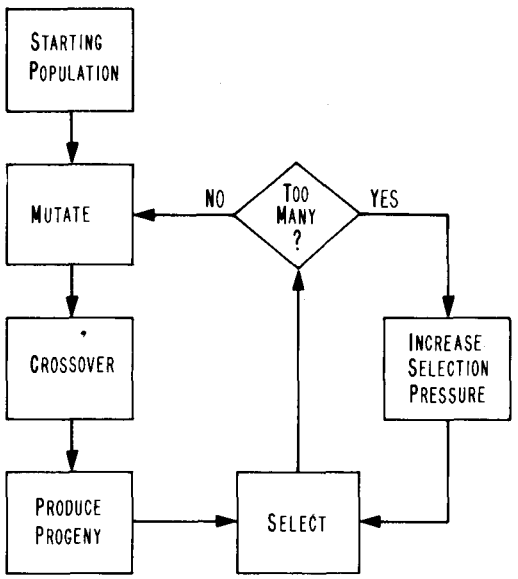

Fig. 3. Organization of the model. See text for description of each step in the cycle

Table 2.

Standard Parameters (1 cycle = 1-100 generations = 1-100 years)	
Mutation rate	$\mu = 1.65 \times 10^{-4}$ gene^{-1}cycle^{-1}
Crossover rate	$\chi = 1 \times 10^{-4}$ x sum of sizes of shorter haplotypes
Minimum required number of functional genes	500
Maximum allowed number of nonfunctional genes	1000
Maximum allowed haplotype size	1500
Effective population size (P_e)	10
Randomly-picked dominant male breeding	

nonfunctional genes to functional genes are ignored. Hence, nonfunctional genes leave the system through selection and random sampling during breeding.

The number of mutations for one cycle for the whole population and for each type of gene is calculated. These mutations are then assigned to randomly picked haplotypes. The resulting changes only affect the gene composition of the progeny of individuals chosen to undergo mutation. Parental individuals are discarded after each cycle.

B. Crossovers. After undergoing mutation, the haplotypes are altered by unequal crossing over. The number of crossovers in the population per cycle is calculated by multiplying the crossover rate (Table 2) by the sum of the size of the shorter haplotype of each individual. (The two haplotypes of each individual need not be of identical size.) Since in each cycle the sum of the size of the shorter haplotype in each individual is 7×10^3 to 1×10^4, the simulations actually use a rate of about one unequal crossing over per cycle. If each gene is 300 base pairs (bp) long, and one cycle is one generation, then the rate of crossing over is 3.3×10^{-7} to 4.8×10^{-7} bp^{-1} generation^{-1}. The rRNA multigene family on the X chromosome of *Drosophila melanogaster* may undergo unequal crossing over to result in gene reduction at a rate of 0.003 generation^{-1}; the reversion of mutants by gene magnification from a reduced number to the wild-type level occurs at a rate of 0.24 (Tartof, 1974). The former rate corresponds to 8×10^{-10} bp^{-1} generation^{-1}, since the rRNA gene repeat is 16 Kbp long, and there are about 230 such repeats in the family (Tartof, 1974). Thus the simulations use a crossing over rate which is 400-600 times higher than the rate of gene reduction found in the rRNA multigene family. The use of a high rate of crossing overs is analogous to the use of high mutation rates when dealing with small populations in simulations (see e.g., Ohta 1978a).

The crossovers are performed on randomly selected individuals. The simulation of crossover consists of randomly aligning an individual's two haplotypes and then randomly choosing a recombination site internally, within the region where the haplotypes overlap. All crossovers are assumed to be intergenic. This method of interchromosomal crossing over is similar to that of Ohta (1978a). In reference to Fig. 2, this means that the choice of apposition of genes, 7 with A, 8 with B, etc., was made randomly, and the choice of a crossover event between genes 8 and 9 also was random. Thus in most cases the two resulting haplotypes are of unequal sizes. The lengthened (shortened) haplotype (containing both functional and nonfunctional genes) is then the element on which selection acts.

C. Breeding. Following crossover, the population is bred. The model employs two basic breeding methods. The first employs random pairing of individuals. In the second method, one animal is randomly picked as the dominant male (Klein, 1975), and it is paired with all the others. In other simulations a dominant male is chosen with dominance being defined as some selected attribute. The method of selection of the dominant male is varied. Different choices for the dominant male include, for example, the individual with the maximum number of a particular type of gene, or the one with the longest haplotype. In either case, all potential progeny genotypes are generated. Parental animals are discarded after breeding, so only the progeny undergo selection to determine the individuals that will comprise the next generation.

D. Selection. During breeding, the size of the potential progeny population has been amplified fourfold over the old population. This number is then reduced to the effective population size ($\leqslant 10$ individuals). Two methods are used for this reduction: a random elimination of some fixed percent (usually 50%) of the progeny, and a selection process based upon the genotype of each individual.

The random elimination of some of the potential progeny reflects the random sampling nature of breeding, where not all possible progeny genotypes are produced. As a check, several runs were made without using this random elimination. These simulations gave essentially the same results as those runs using the random elimination method.

The selective elimination of possible progeny involves several criteria. First, each individual is checked for the size of its two haplotypes. We arbitrarily set a maximum allowable haplotype size and eliminate any individual having a haplotype whose size exceeded the maximum (see next section). Second, we assume an additive effect of gene fitness, that is, progeny are selected according to the number of functional and nonfunctional genes in their haplotypes. The number of functional genes and the number of nonfunctional genes (the gene composition) of each offspring are compared to some prescribed standards (Table 2). Individuals are eliminated if they have either too few functional genes or too many nonfunctional genes. If the remaining number of individuals is less than or equal to 10, the next cycle begins; otherwise, a stepwise process of increasing the selective pressures is instituted. This is done by increasing the minimum number of functional genes required and decreasing the maximum number of nonfunctional genes allowed by some fixed amount (usually five genes). This process continues until the population size falls into an acceptable range.

In many simulations, this competition step is followed by another random elimination step to reduce the number of individuals to the effective population. If a population is very homogeneous, as for example, at the beginning of a run, an increase in the selective pressure could totally annihilate the population. A check is incorporated in the model to see if this event will occur. If it does, the current round of selection is aborted and instead the population is randomly reduced to 10. This random culling technique is used only after the initial selective pressure has been increased. This procedure guarantees that any population which meets the minimal requirements will not die out during sibling competition. If all the progeny of the previous population have a genetic makeup that did not meet the originally prescribed standards (before any increase in selective pressure), then the population is allowed to die. This situation actually occurred in several simulations, as, for example, when the randomly chosen

dominant male happened to be the individual with the fewest number of functional genes.

In other simulations, competition is stopped once the remaining individuals are essentially alike, e.g., less than 5% difference in the number of G or N genes that each of the individuals have. If the population is greater than 10, then 10 are sampled randomly to be the effective population for the next cycle.

E. Parameter Values. The values of certain parameters used in the simulations either cannot be confidently estimated or are not known. These parameters include the maximum allowable size of a haplotype, the number of functional genes needed for survival, and the number of nonfunctional genes that makes a haplotype lethal. Clearly, these parameters must have a natural limit - e.g., the size of a haplotype is bounded by the length of the chromosome that carries it. In both *Xenopus laevis* and *X. mulleri* the 5S rRNA genes are found on the telomeres of most of the chromosomes. *X. mulleri* has about 9000 5S rRNA genes while *X. laevis* has about 24,000 such genes (Brown and Sugimoto, 1973). It is not clear why *X. laevis* should have more than 2.5 times as many 5S rRNA genes as *X. mulleri.* Furthermore, we cannot safely conclude that the maximum allowed haplotype size for 5S rRNA genes is also 24,000 in *X. mulleri*, since there might be some selective pressure operating against a large haplotype size in *X. mulleri* such that the size of its 5S rRNA family stays relatively small. Thus we have had to assign arbitrary values to some of the parameters. In all such cases they were systematically varied to ascertain the effect of different choices. For some variables, no noticeable effect was detected; other cases had a significant impact. All of these results are described below.

Results

Behavior of the Model

The contribution of the various parameters representing evolutionary processes are described by noting their effects on the two major properties of the multigene family - gene composition and haplotype size. The simulations showed that these properties were influenced by distinct sets of evolutionary processes. Simulation runs of duration 2×10^3 to 10^4 cycles were performed using the values of the parameters in Table 2. The results are shown in Fig. 4 for the gene composition and Fig. 5 for the mean haplotype size.

A. Progressive Changes in Gene Composition. As the simulation progresses, the number of G_1 functional genes in an average haplotype steadily decreases while the number of nonfunctional genes increases linearly. The number of G_2 functional genes reaches a plateau level by cycle 3000 and remains stable for the duration of the run. Since G_1 and G_2 genes are interconvertible, it can be shown that in the limiting case, for an infinite number of cycles, their numbers will become equal. Nonfunctional genes are derived from both G_1 and G_2 genes and therefore will increase until the number of N genes approaches the boundary value specified by the maximum number of nonfunctional genes allowed. Neither of these limiting cases were reached in any of the simulations.

B. Progressive Changes in Haplotype Size. Starting with an initial haplotype size of 500 genes, unequal crossovers generate size heterogeneity on which selection can act. There is an initial transient peak in haplotype size of 1200-1400 genes centered about 100 cycles. Thereafter the size stabilizes to a level of 700-900 genes for the duration of the simulations (Fig. 5). This transient peak is explained below (section G).

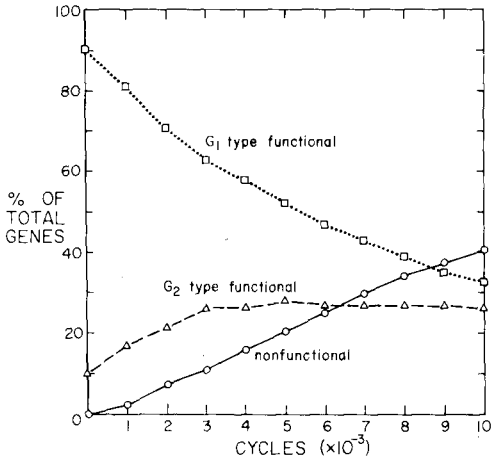

Fig. 4. Evolution of gene composition up to cycle 10,000 under standard conditions

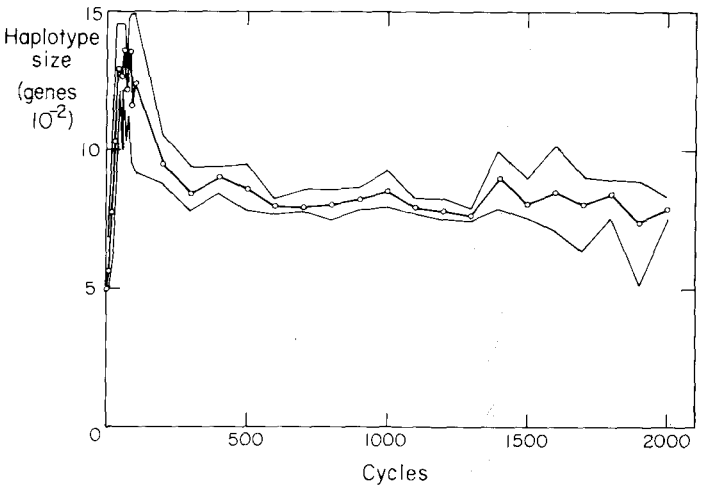

Fig. 5. Evolution of haplotype size during the first 2000 cycles under standard conditions. No further change in the mean haplotype size was observed after ca. 500 cycles. Note the initial rapid rise and decline in haplotype size. Thin lines indicate the largest and smallest sizes encountered in two replications

The size of the multigene family fluctuates under constant conditions. Differences occur between individuals in the population but are usually smaller than the fluctuations from generation to generation. These short-term fluctuations are grouped around a mean value which is stable for up to 10^4 cycles (10^4-10^6 generations). This kind of size heterogeneity is observed in nature, for example, in the amphibian rRNA families. Family size heterogeneity is present between individuals of a given species while two sibling species, *Xenopus laevis* and *X. mulleri*, have very different average 5S rRNA family sizes (Brown and Sugimoto, 1973).

C. Breeding Behavior. We compared random pairing of the population with that where a random individual is chosen as the dominant male and paired with all other individuals. Both models followed the same evolutionary path (Figs. 4 and 5), but the random pairing model was kinetically slower. We therefore used the randomly-chosen dominant-male model in the majority of our simulations to conserve computation time. As would be expected, the gene composition is influenced by the choice of the dominant male if the dominant male is not chosen randomly. For example, picking the individual with the most G_2 functional genes as the dominant male results in a large increase in the proportion of G_2 genes in the population while G_1 genes and nonfunctional genes show proportional decreases.

When the dominant male is chosen as having the largest haplotype size, the average size of the multigene family in the population increased to the maximum allowable size with no change in gene composition. This increase shows that direct selection for haplotype size does indeed work. However, we consider this a trivial case since we show below that selection for functional genes and against nonfunctional genes can also control the size of the family. This is achieved without having to assume any direct size selection beyond those used as boundary parameters (e.g., maximum allowed haplotype size).

D. Mutation Rates. The simulations tested mutation rates of $0 < \mu \leqslant 0.005$ gene^{-1} cycle^{-1}. If we assume that 1 cycle = 1 generation = 1 year, and that there are 300 bp per gene, then $0 < \mu \leqslant 1.7 \times 10^{-5}$ bp^{-1}y^{-1}. With $P_e = 10$, then $0 < P_e\mu \leqslant 0.05$ gene^{-1} cycle^{-1}. In comparison, Ohta (1978a) used $0.025 < P_e\mu \leqslant 0.1$. When the mutation rate (μ) is increased (Fig. 6), there is an increase in the rate of evolution. For example, with $\mu = 1 \times 10^{-3}$ gene^{-1} cycle^{-1} (6 times the standard rate) the gene composition and haplotype size at 1000 cycles are similar to those at 6000-7000 cycles using the standard rates. Identical evolutionary pathways (Figs. 4,5) are observed, but in a much shorter time period. Similarly when the mutation rate is reduced, the rate of evolution is reduced. Both the gene composition and the haplotype sizes under all these conditions remain the same as those using the standard conditions with appropriate time adjustments. Within the range of mutation rates we tested, the effect of a change in mutation rates is to change the scale of the X-axis in Figures 4 and 5. The evolutionary path is unaffected. This is an important observation since the absolute mutation rate for multigene families is not known with certainty. (The estimated μ for proteins is about 5×10^{-9} bp^{-1} y^{-1}; Cavali-Sforza and Bodmer, 1971; Dayhoff, 1972). The results here indicate that using a different set of mutation will affect the kinetics but not the path of evolution of the multigene family.

We compared the behavior of the model in 2000-cycle runs using standard parameters where the last 1000 cycle either had $\mu = 1.65 \times 10^{-4}$ gene^{-1} cycle^{-1} or $\mu = 0$. As expected, the progressive change in gene composition (Fig. 4) stopped when mutations ceased. There is no significant difference in the haplotype sizes, which fluctuated around an equilibrium level (Fig. 5, and data not shown). Thus the maintenance of an equilibrium haplotype size is independent of ongoing mutations, provided that functional and nonfunctional genes are present, on which selection may act.

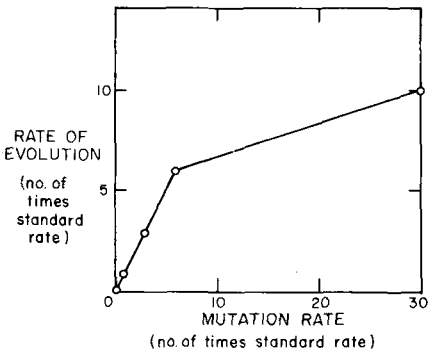

Fig. 6. Effect of changing mutation rates on the rate of evolution of the multigene family gene composition

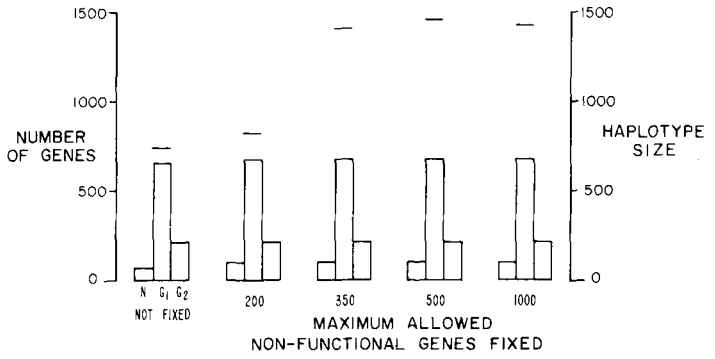

Fig. 7. G-type competition. Situation after 1000 cycles. The maximum allowed number of nonfunctional genes was fixed at 200, 350, 500, and 1000. This is compared with the standard condition where the maximum was allowed to decrease with increasing selective pressure (left histogram). From left to right each histogram consists of columns corresponding to the relative proportion of N, G_1, and G_2 genes respectively. The number of each is expressed as the number of genes of each type per thousand total genes. Note the increase in haplotype size (bar above each G_1 column) in the former cases

E. Selection. Changes in the selection processes used to reduce the progeny at each cycle to the effective population size have little, if any, effect on the gene composition of the individuals (Figs. 7-9). Partly, this is a consequence of our model in that we consider only three broad types of genes, rather than individual variants.

However, selection does control the absolute number of nonfunctional genes carried in a multigene family; i.e., natural selection can detect and operate against individuals carrying an excessive number of nonfunctional genes. An important corollary resulting from this observation is that most multigene families will contain nonfunctional genes (Nei, 1975; Ohta, 1978b). This constraint especially may be true for large multigene families whose constituent genes have overlapping functions.

F. Constraints on Haplotype Size. An arbitrary constraint placed on the haplotypes was a maximum size limit of 1500 genes. Any individual with one or both haplotypes greater than this size is eliminated. In simulations with the standard parameters, this boundary limit was rarely a factor. In fact, when the maximum size limit was fixed instead at 1000 or 2000 genes, the results are identical to those with a maximum limit of 1500 genes. The exception is that with a limit of 1000 genes, the initial transient peak was much lower. The stabilized haplotype sizes (700-900 genes) are identical using any of these boundary limits. The function of this maximum size limit is to model a putative natural limit, and is useful to stop any runaway increase in haplotype size occurring under special conditions (see next section).

G. G-Type Competition. Individuals in the population are required to have a minimum number of G genes. This constraint sets a lower limit to the family size. During the selection process, individuals may compete with each other through a comparison of the number of functional genes (G-type competition). An individual with more functional genes is assumed to have a selective advantage. Thus, there is a selective drive towards longer haplotype lengths.

The effects of G-type competition are accentuated by allowing individuals to carry relatively large numbers of nonfunctional genes, e.g., by fixing the maximum allowable number of nonfunctional genes at relatively high levels (Fig. 7). Under these conditions, the mean haplotype size approaches the maximum allowable size. In the absence of a maximum size limit this increase could continue indefinitely. Note that the gene composition is unaffected when compared to standard conditions. Individuals also are forced to undergo G-type competition when they all have very few or no nonfunctional genes, as at the start of the simulations. The consequent G-type competition accounts for the initial rapid increase in the mean haplotype size. With time, the number of

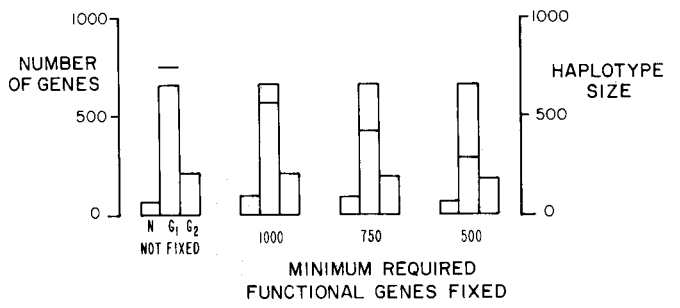

Fig. 8. N-type competition. Situation after 1000 cycles. The minimum required number of functional genes was fixed at 500, 750, and 1000. This is compared with the standard condition where the minimum was allowed to increase with increasing selective pressure (left histogram). Note the decrease in haplotype size in the former cases

nonfunctional genes increases and the increasing extent of N-type competition (section H) forces the mean haplotype size to decrease. This is the reason for the initial transient peak in the mean haplotype size.

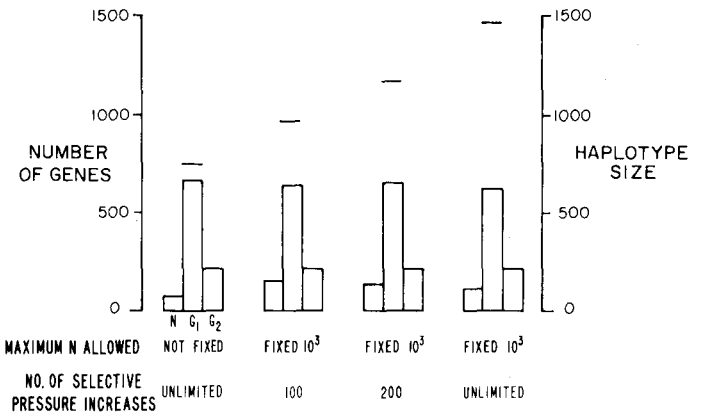

Fig. 9. Effect of competition intensity on haplotype sizes. Situation after 1000 cycles. The histogram on the left shows results using standard parameters and is included for reference. The histogram on the right shows typical results of G-type competition under unrestricted intensity of competition. The middle histograms demonstrate the effect of limiting the intensity of competition under G-type competition. The haplotype size varies directly with the intensity of competition

H. N-Type Competition. Individuals are allowed to carry some nonfunctional (N-type) genes as long as the number they have is below a prescribed limit. During the selection process, individuals may compete with each other by comparing the number of nonfunctional genes they have (N-type competition). Individuals having fewer nonfunctional genes are assumed to be selectively favored. N-type competition tends to restrict the expansion of the haplotype size since a larger size is likely to contain a greater number of nonfunctional genes. Larger haplotypes are also more likely to suffer mutations of functional genes to nonfunctional gels.

N-type competition can be accentuated by fixing the required number of functional genes such that all individuals which have this minimum number are selectively equal in respect to functional genes. The individuals are thus forced to compete by comparing the number of nonfunctional genes they have (Fig. 8). In these cases, the mean haplotype sizes show dramatic decreases to levels where just enough functional genes are carried in each individual's two haplotypes to satisfy selective requirements. Any individual with more than one or two nonfunctional genes is eliminated. Again, the gene composition is unaffected.

I. Intensity of Competition. At equilibrium (beyond ca. cycle 500) the N- and G-competition balance each other and a stable average haplotype size results (Fig. 5). Whenever one or the other type of competition is emphasized, the balance is shifted, and the haplotype size changes to a new stable position. The adjustment is towards a larger family size if individuals are selected for having more functional genes (G-type

competition). Conversely selection against carrying nonfunctional genes (N-type competition) tend to decrease the size of the family. Under equilibrium conditions random sampling in a small population may cause the average haplotype size to increase or decrease, however the prevailing selective conditions will emphasize N-type or G-type competition respectively and therefore to return the haplotype size to equilibrium. The G-type and N-type competitions provide the *direction* for change under selection. The *magnitude* of change is provided by the intensity of competition or selection.

The model is structured to increase the intensity of competition among individuals by imposing progressively stringent selective criteria when reducing the number of progeny to the effective population size. Limiting the intensity of competition limits the magnitude of the effects of the competitive strategies that are present (Crow and Kimura, 1970). For example, if the individuals are engaged in G-type competition, then limiting the intensity of competition prevents the haplotype size from increasing to its usual level (Fig. 9), although it is still clearly larger than under standard conditions.

Discussion

We have shown that the size of a multigene family may be controlled by two types of selective forces. First, selection may assay directly for the size of the family, as exemplified by picking as the dominant male the one with the largest family size. As expected this type of selection resulted in a much expanded family size throughout the population. Second, and more important, selection may act upon the family as a single functional unit. This type of selection is sufficient to maintain the size of the multigene family in a dynamic equilibrium in our model. The dynamic equilibrium is achieved by the balance of two types of drives. One is G-type competition whereby animals compete with one another by comparing the number of functional genes they carry. Since we assume an additive effect of genes (i.e., 2x functional genes are selectively advantageous compared to x functional genes), this kind of competition provides a drive towards carrying more functional genes and thus a drive to increase the size of the multigene family. The opposite drive originates from N-type competition whereby animals compete with one another by comparing the number of nonfunctional genes they carry. Since we assume an additive effect of genes (that 2y nonfunctional genes are selectively disadvantageous compared to y nonfunctional genes), N-type competition provides a drive towards a smaller multigene family, since a large multigene family potentially may carry more nonfunctional genes.

The actual equilibrium size of the multigene family is determined by the balance between the intensitites of the two opposing drives. An adjustment of the selective intensities will cause a shift of the equilibrium size. Our standard model results in an equilibrium family size of about 800 genes. We were able to shift progressively the equilibrium size to as low as 300 genes, and to as high as 1500 genes, by varying the appropriate selective parameters. These extreme values are in fact imposed by the boundary conditions, that an animal must carry at least 500 functional genes (an average of 250 functional genes per family), and that the family size must be maximally 1500 genes. Thus, the model can maintain an equilibrium family size at any point between the boundary conditions.

Mutation and homologous but unequal crossing over generate the variability upon which selection may act. Mutation generates the primary variability in gene composition. Homologous but unequal crossing overs generate family size variability. It also reassorts genes from the paternal and the maternal chromosomes, and thus also generates variability in gene composition. This is important in our model since we chose not to emphasize direct selection for family size, but primarily to use selection which is sensitive only to the gene compositions of individuals. As discussed above, the equilibrium size of the multigene family can be controlled by this relatively indirect process. The interesting prediction is that given a population containing both functional and nonfunctional genes, and at equilibrium for family size, ongoing mutations may cease, without affecting the model's ability to maintain an equilibrium family size, since variability in gene composition is continually generated by homologous but unequal crossing overs. This was observed in our simulations. Thus, homologous but unequal crossing over by itself can generate the variability in gene composition, as well as in family size, upon which selection may act.

Our model of selection involves (a) random birth of genotypes; (b) truncation selection, to make all individuals conform to minimal criteria; (c) sib competition where an additive effect of genes is assumed; and, if necessary, (d) random sampling of breeding individuals for the next generation. Of these steps, sib competition is the one which provides the opposing forces (N-type and G-type competition) whose balance controls the size of the multigene family. It would be interesting to test other models of selection to determine if similar opposing forces can be found to control the size of a multigene family. We expect a theoretical description of the model will further our understanding of the forces controlling the size of a multigene family.

Acknowledgements. This work was supported by NSF grant PCM 76-81542 and NIH grants AI 1781 and GM 06965. H.V.H. is a Danforth Fellow.

References

Birnstiel, M., Telford, J., Weinberg, E., Stafford, D. (1974). Proc. Nat. Acad. Sci. USA **71**, 2900–2904

Black, J.A., Gibson, D. (1974). Nature **250**, 327–328

Brown, D.D., Sugimoto, K. (1973). Cold Spring Harbor Symp. Quant. Biol. **38**, 501–506

Cavali-Sforza, L.L., Bodmer, W.F. (1971). The genetics of human populations. San Francisco: W.H. Freeman

Crow, J.F. and Kimura, M. (1970). An introduction to population genetics theory. New York: Harper and Row

Dayhoff, M.O. (1972). Atlas of protein sequence and structure 1972. Washington, D.C.: The National Biomedical Research Foundation

Ford, P.J., Southern, E. (1973). Nature New Biol. **241**, 7–12

Hood, L. (1972). Fed. Proc. (FASEB) **31**, 177–187

Hood, L. (1973). Stadler Symp. **5**, 73–142
Hood, L., Campbell, J.H., Elgin, S.C.R. (1975). Ann. Rev. Genet. **9**, 305–353
Kimura, M., Ohta, T. (1979). Proc. Nat. Acad. Sci. USA, (in press)
Klein, J. (1975). Biology of the mouse histocompatibility-2 complex. Berlin-Heidelberg-New York: Springer
Nei, M. (1975). Molecular population genetics and evolution. Amsterdam: North-Holland
Ohta, T. (1976). Nature **263**, 74–76
Ohta, T. (1977). Nature **267**, 515–517
Ohta, T. (1978a). Genetics **88**, 845–861
Ohta, T. (1978b). Genet. Res. Camb. **31**, 13–28
Ohta, T. (1978c). Proc. Nat. Acad. Sci. USA **75**, 5018–5112
Perelson, A.S., Bell, G.I. (1977). Nature **265**, 304–310
Smith, G.P. (1973). Cold Spring Harbor Symp. Quant. Biol. **38**, 507–513
Smith, G.P. (1976). Science **191**, 528–535
Tartof, K.D. (1974). Proc. Nat. Acad. Sci. USA **71**, 1272–1276
Tartof, K.D. (1975). Ann. Rev. Genet. **9**, 355–385
Tartof, K.D., Dawid, I.B. (1976). Nature **263**, 27–30

Received February 5, 1979; revised January 2, 1980