Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 3509–3512, May 1988
Genetics

Time for acquiring a new gene by duplication

(evolution/population genetics/stochastic process)

TOMOKO OHTA

National Institute of Genetics, Mishima 411, Japan

Communicated by Motoo Kimura, January 4, 1988 (received for review November 4, 1987)

ABSTRACT **In view of the widespread occurrence of gene families in eukaryotic genomes that suggests the importance of gene duplication in evolution, a population genetic model incorporating unequal crossing-over was formulated. By using this model, the time needed for acquiring a new gene is investigated by an approximate analytical method and by computer simulations. The model assumes that natural selection favors those chromosomes with more beneficial genes than other chromosomes in the population, as well as random genetic drift, mutation, and unequal crossing-over. Starting from a single gene copy, it is found that the time for acquiring another gene with a new function is dependent on the rates of occurrence of unequal crossing-over and mutation. Within a realistic range of parameter values, the required time was at least several times $4N$ generations, where N is the effective population size. Interchromosomal unequal crossing-over at meiosis is more effective than intrachromosomal (between sister chromatids) unequal crossing-over for obtaining a new gene, provided that other parameters are the same. However, the genetic load for acquiring a gene is larger under the model of interchromosomal crossing-over. The relevance of this finding to the advantage of sexual reproduction is discussed.**

It has become increasingly evident that gene duplication and subsequent differentiation played an important role in the evolution of genetic systems. There are numerous examples that suggest that new genes were created by duplication (for reviews, see refs. 1–3). From the standpoint of population genetics, I have tried to formulate a model of evolution for a new genetic system (4–6). In these studies, unlike the previous models that assume fixed loci, unequal crossing-over is assumed to occur continuously, leading to changes in gene arrangement and number of loci. The diversity of organization observed in genes for hemoglobin (7), immunoglobulin (8), and interferon (9) calls for a model that provides a comprehensive understanding of their origin.

I have examined how a genetic system evolves (5, 6) under unequal crossing-over, natural selection, and random genetic drift, starting from a single gene copy. In this paper, the time required for spreading a beneficial mutant allele into the population at one locus of a duplicated loci will be examined. Interchromosomal unequal crossing-over through sexual reproduction accelerates evolution by duplication (6). Thus, both haploid and diploid models are investigated, and the advantage of sexual reproduction is discussed.

Theoretical Approach

As I have reported (5, 6), I assume that initially a single gene copy exists in each genome and that all these genes are identical in the population. Unequal crossing-over is assumed to occur at γ, a constant rate per gene per generation. I assume that this gene is indispensable, so that its complete loss from the genome is lethal. Except for the first production of two tandem genes, I assume that unequal but homologous crossing-over leads to duplication or deletion by one gene unit. Crossing-over is assumed to occur between sister chromatids (i.e., intrachromosomally), and a simple haploid model will be studied in this section. Interchromosomal unequal crossing-over through sexual recombination (diploid model) will be treated in the next section.

Two types of mutations will be assumed. One type is beneficial and the other is deteriorating. They occur with v_+ and v_-, the rates per gene per generation, respectively. It is assumed that if a beneficial mutation occurs in a gene, it changes to a new form according to the infinite allele model (10), and if a deteriorating mutation occurs, the gene becomes nonfunctional, i.e., a pseudogene. The term "allele" may not be quite appropriate when two or more duplicated loci exist on a chromosome; however, for convenience, I shall use it to represent mutational states of genes belonging to the gene family.

A finite population with the effective size N is assumed. Selection takes place at the haploid stage according to the following scheme, where w stands for the fitness (adaptive value).

$$w_i = 1 \quad \text{for } k_i \geq \bar{k}$$

and

$$w_i = e^{-s(\bar{k}-k_i)} \quad \text{for } k_i < \bar{k}. \qquad [1]$$

In these expressions, the subscript i refers to the ith genome, so that k_i is the number of different beneficial alleles in the ith genome, $\bar{k}$ is the population average, and s is the selection coefficient. In other words, if the number of different alleles contained in a genome is less than the population average ($\bar{k}$), then this genome is selected against according to an exponential function. This represents a type of selection for increasing gene function, since genomes having larger number of alleles tend to be favored.

As to negative selection, the gene is assumed to be indispensable, so that its complete loss due to unequal crossing-over or damaging by deteriorating mutation is lethal. However, it is considered that accumulation of pseudogenes has no effect on fitness as long as at least one copy of the normal gene is present.

I shall now investigate the following problem: How long does it take until all the haploid genomes in the population acquire two different alleles, starting from a population consisting of genomes each with a single gene copy? The process may be partitioned into two phases: (*i*) spreading of a genome carrying a duplication and (*ii*) fixation of a beneficial mutant allele at one of the two tandemly duplicated loci. In the following analysis, I assume that the products $N\gamma$, Nv_+, and Nv_- are much less than unity. Then the two phases can be assumed to occur separately without overlap.

The first phase is analogous to the spreading of neutral mutant alleles in the population under mutation pressure. Here unequal crossing-over corresponds to mutation in the ordinary model of population genetics. Specifically, $\gamma/2$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

corresponds to mutation rate in the ordinary model; here we are concerned with duplication, and we assume that deletion becomes lethal. By using the result of Kimura (11), the average time until fixation of chromosomes carrying duplication starting from a very low frequency, may be expressed as follows

$$t_1 \approx \frac{4N}{V_1 - 1}[0.577 + \psi(V_1)], \quad [2]$$

where $V_1 = 2N\gamma$ in our notation and $\psi(\cdot)$ stands for the digamma function. This represents the first phase.

The second phase is analogous to the spreading of selected mutants at a single locus. The mutation rate of the analogous single locus model corresponds to twice the rate of beneficial mutations of the present model, since there are two loci that are duplicated. The fitness function **1** implies that beneficial mutant alleles behave under selection as if they were recessively advantageous (see ref. 5). By noting the above correspondence, Kimura's formula for the time until fixation of a beneficial mutant allele for the second phase becomes as follows.

$$t_{2+} = 4N\int_0^1 e^{-S\eta^2}\eta^{-V_2}\,d\eta\int_0^\eta \frac{e^{S\xi^2}\xi^{V_2-1}}{1-\xi}\,d\xi, \quad [3]$$

where $S = 2Ns$ and $V_2 = 8Nv_+$.

The total time until fixation of a beneficial mutant allele at one of duplicated loci is not simply the sum $t_1 + t_{2+}$, because spreading of deteriorating mutations may also occur. When a deteriorating mutation spreads before a beneficial one fixes, the evolution of a new gene must start again. Let us designate the time until fixation of a deteriorating mutant allele in one of the copies t_{2-}, which may be obtained by Eq. **2** by setting $V_1 = 8Nv_-$, since the spreading is assumed to be neutral. Let us now estimate approximately the total time required. Let b be the fraction of cases where beneficial mutant alleles spread before deteriorating ones. In the remaining cases (fraction, $1 - b$), one of the redundant genes degenerates. Then the expected total time (T) for spreading may be expressed as follows.

$$T \approx b\left[(t_1 + t_{2+})\sum_{i=0}^{\infty}(1-b)^i + (t_1 + t_{2-})\sum_{i=1}^{\infty} i(1-b)^i\right]$$

$$= (t_1 + t_{2+}) + [(1-b)/b](t_1 + t_{2-}). \quad [4]$$

I have shown (5) that the ratio (R) of the numbers of beneficial mutant alleles to the pseudogenes at a duplicated locus [$\approx b/(1-b)$ in our notation] is roughly

$$R \approx \frac{b}{1-b} \approx \frac{u_+ v_+}{u_- v_-}, \quad [5]$$

where u_+ and u_- are the fixation probabilities of beneficial (+) and deteriorating (−) mutations. In the present case, they are (5)

$$u_+ \approx \sqrt{2Ns}(1.128/2N) \quad [6]$$

and

$$u_- = 1/2N. \quad [7]$$

By using Eqs. **2–7**, it is possible to estimate the expected total time for spreading of a beneficial mutant allele into the population at one of the duplicated loci. Note that the above analysis is based on the assumption that the products $N\gamma$, Nv_+, and Nv_- are much smaller than unity and that the two phases occur separately. Actually, when the assumptions are not satisfied, the two phases overlap and the theoretical expectation, as computed above, may be biased. In the next section, such bias will be examined by computer simulations.

Monte Carlo Simulations

Both haploid and diploid populations were studied. The method of simulation used was the same as I described in refs. 5 and 6. Each generation of the simulation experiments consisted of unequal crossing-over, mutation, random sampling, and selection, and all these were carried out by generating random numbers. In the haploid model, the unequal crossing-over was assumed to occur between sister chromatids (i.e., intrachromosomally) by shifting one gene unit, and no interchromosomal recombination was permitted. In the diploid population, interchromosomal unequal crossing-over was assumed to occur in the following way. The rightmost gene of one chromosome was assumed to pair with the next-right-most gene of the other chromosome at meiosis. No intrachromosomal unequal crossing-over was considered in the diploid model.

As for selection, the fate of a sampled chromosome was determined by using the fitness function **1** in the haploid population. In the case of diploids, the fitness function was modified to apply to diploid individuals. Namely, the number of beneficial alleles were counted for sampled individuals, so that i in Eq. **1** represented the ith individual in the population.

In addition to beneficial and deteriorating mutations, selectively neutral mutant alleles were also assumed in the simulations. As described by Ohta (6), 10 sites were assumed in a gene, and each site changed following the infinite allele model (10). In practice, allelic states were stored as integers, and, with each mutation, the integer was increased by one. When the integer was a multiple of 10, the allelic state was assumed to be beneficial. Therefore, $v_0 = 9v_+$, where v_0 is the neutral mutation rate (per gene per generation).

All experiments were continued until the average number of beneficial alleles contained in a chromosome (haploid case) or in an individual (diploid case) became 2.01, and the number of generations required was recorded. Note that, in the diploid case, heterozygotes have two alleles at one locus. Simultaneously, the copy number, the pseudogene number, and the proportion of beneficial mutant alleles in total divergence (i.e., beneficial plus neutral mutant alleles accumulated) on a sampled chromosome were counted. Also the genetic load was recorded all through the experiment, so that the total load for acquiring a new gene could be estimated.

The results for haploids are shown in Table 1. It is a well-known fact in theoretical population genetics (see refs. 12 and 13) that stochastic behavior of mutant alleles in finite populations is largely determined by the products $N\gamma$, Nv_+, and Nv_0, but not by N and other parameters separately when we scale time in units of N. This enabled us to save computer time by choosing a small N and relatively large mutation and recombination rates. I chose realistic values for the products $N\gamma$, Nv_+, and Nv_0 based on knowledge of molecular evolution by letting $N = 50$. According to Kimura (14), Nv_0 is often in the range from 0.01 to 0.05. This agrees with values directly estimated for enzyme loci (15–17). As for deteriorating mutation, two cases were examined: in one case $v_- = v_0 + v_+$ and in the other case $v_- = 0$. The estimated value of $N\gamma$ was in the range from 0.01 to 0.1 for immunoglobulin and other gene families with high variability (18–22). This estimate is also consistent with values directly estimated (23).

Comparison of the observed and expected Eq. **4** values of the length of time for acquiring a new gene shows that the agreement between the two is mostly satisfactory. However, Eq. **4** overestimates the true value when a deteriorating mutation occurs, but it does not do so when $v_- = 0$. This may be explained as follows. Small values of Nv_+, Nv_-, and $N\gamma$, as assumed in deriving Eq. **4**, are not valid in this case, and various polymorphisms in gene organization and primary structure are expected to occur. Therefore, the process of acquiring a new gene is not as simple as considered above.

Table 1. Results of Monte Carlo experiments for haploids on time for acquiring a new gene, number of copies and pseudogenes, proportion of beneficial mutants, and genetic load

Rate	Parameter $2N\gamma$	Parameter $2Nv_+$	Parameter $2Ns$	T, units of $4N$ generations: Observed	T, units of $4N$ generations: Formula 4	Copies, no.*	Pseudogene(s), no.*	Proportion of beneficial mutant alleles*	Genetic load,† units of N gametes
$v_- = v_0 + v_+$	0.025	0.001	20	268.36 ± 170.69	221.87	4.91 ± 1.69	1.55 ± 1.53	0.200 ± 0.234	8.21 ± 3.77
			40	107.91 ± 70.30	169.07	3.92 ± 1.51	0.67 ± 1.35	0.454 ± 0.357	10.64 ± 7.50
	0.025	0.0025	20	127.28 ± 80.79	166.37	5.52 ± 3.66	2.64 ± 3.51	0.222 ± 0.142	11.13 ± 5.81
			40	125.83 ± 104.74	129.89	3.12 ± 0.82	0.08 ± 0.26	0.238 ± 0.157	10.02 ± 4.44
	0.1	0.01	20	31.86 ± 28.45	44.44	4.65 ± 2.26	1.95 ± 2.22	0.276 ± 0.251	8.04 ± 3.86
			40	15.33 ± 14.41	34.61	3.27 ± 1.11	0.40 ± 0.92	0.330 ± 0.285	9.60 ± 7.12
	0.25	0.01	20	16.17 ± 14.00	26.33	4.40 ± 1.72	1.14 ± 1.52	0.365 ± 0.349	8.56 ± 4.43
			40	13.04 ± 8.43	20.03	4.01 ± 1.43	0.57 ± 1.14	0.259 ± 0.238	8.69 ± 4.18
$v_- = 0$	0.025	0.001	20	103.78 ± 62.29	91.52	4.01 ± 1.37	0.0	0.400 ± 0.254	6.27 ± 1.89
			40	84.01 ± 53.30	76.89	3.35 ± 1.07	0.0	0.413 ± 0.290	6.85 ± 1.71
	0.025	0.0025	20	66.45 ± 37.64	61.81	3.14 ± 1.43	0.0	0.216 ± 0.239	6.07 ± 2.53
			40	87.43 ± 64.73	55.95	3.14 ± 0.64	0.0	0.341 ± 0.357	7.01 ± 3.64
	0.1	0.01	20	16.59 ± 14.31	16.17	2.99 ± 0.92	0.0	0.286 ± 0.254	5.62 ± 2.41
			40	16.47 ± 8.69	14.62	2.95 ± 1.04	0.0	0.270 ± 0.243	6.66 ± 2.49
	0.25	0.01	20	9.25 ± 4.37	10.10	3.83 ± 1.18	0.0	0.454 ± 0.368	5.46 ± 1.88
			40	9.84 ± 7.44	8.55	3.53 ± 1.38	0.0	0.441 ± 0.375	6.52 ± 1.45

T is time for acquiring a new gene. Results are expressed as average ± SD (n = 15).
*Final value of the experiments. Proportion of beneficial mutant alleles is that among total divergence of a sampled genome.
†Total load needed for acquiring a new gene.

When the two phases mentioned above overlap because of polymorphisms, the theoretical expectation tends to overestimate the observed value. This would account for the disagreement when $v_- > 0$. Another complication is the nature of unequal crossing-over that results in either the addition or loss of one gene unit when the chromosome contains two or more duplicated genes that are functional. Formula 4 is based on the model of unidirectional mutation and does not take into account such bidirectional crossing-over. This would make the observed time longer than the predicted value, and it accounts for the slight disagreement between the observed and the theoretically predicted values of the length of time when $v_- = 0$.

It is also interesting to find that the copy number, the pseudogene number, the proportion of beneficial mutations, and the genetic load do not change very much when the parameters are widely different. Only the time for acquiring a new gene differs greatly.

The results for diploids assuming interchromosomal unequal crossing-over are given in Table 2. The cases without deteriorating mutations were not investigated. I have not been able to obtain a mathematical formula for computing the expected time. The results of simulation experiments show clearly that interchromosomal unequal crossing-over and diploid selection accelerate the evolution of a new gene. In this model the time required was only 1/5 to 1/2 of the time required in the haploid model, if the parameters were the same; and the amount of acceleration seemed to be greater when the unequal crossing-over and mutation rates were lower. When two or more beneficial mutant alleles exist at the same locus in the population, the situation would be similar to over-dominance, and selection would increase the frequencies of such mutant alleles. Eventually the two different alleles may be combined into one chromosome by unequal crossing-over (6). The situation would be similar to permanent heterozygosity (24). The acceleration may have an important bearing on our consideration of the advantage of sex (e.g., ref. 25), and it will be discussed in the next section.

It may also be noted that the total copy number and the number of pseudogenes contained are slightly smaller in the diploid than in the haploid model, while the fraction of beneficial mutant alleles was slightly higher in the diploid than in the haploid model. The genetic load is considerably larger in the diploid than in the haploid model, the magnitude of which depends much on parameters. These differences between the two models were again caused by more effective selection in the diploid through interchromosomal unequal crossing-over than in the haploid model (6).

DISCUSSION

An intriguing question that arises is whether the rate of acquiring a new gene as considered in this paper is high enough for significantly enhancing the potential of organisms

Table 2. Results of Monte Carlo experiments for diploids on time for acquiring a new gene, number of copies and pseudogenes, proportion of beneficial mutants, and genetic load

Parameter $2N\gamma$	Parameter $2Nv_+$	Parameter $2Ns$	T observed, units of $4N$ generations	Copies, no.*	Pseudogene(s), no.*	Proportion of beneficial mutant alleles*	Genetic load,† units of N gametes
0.025	0.001	20	90.79 ± 73.35	3.44 ± 2.06	1.81 ± 2.88	0.455 ± 0.326	49.40 ± 62.28
		40	68.91 ± 55.69	3.15 ± 2.68	1.88 ± 4.23	0.496 ± 0.339	97.18 ± 126.54
0.025	0.0025	20	104.18 ± 89.38	4.33 ± 3.53	2.25 ± 3.53	0.315 ± 0.298	117.54 ± 121.48
		40	23.22 ± 13.10	2.74 ± 1.19	0.24 ± 0.61	0.470 ± 0.307	155.38 ± 144.45
0.1	0.01	20	9.21 ± 7.31	4.98 ± 7.92	0.89 ± 1.53	0.387 ± 0.273	16.60 ± 15.02
		40	5.54 ± 4.41	2.06 ± 0.57	0.10 ± 0.26	0.590 ± 0.287	46.03 ± 35.25
0.25	0.01	20	6.99 ± 4.60	3.36 ± 2.66	1.17 ± 2.53	0.523 ± 0.325	13.61 ± 10.55
		40	8.00 ± 7.10	2.40 ± 0.97	0.49 ± 1.00	0.454 ± 0.368	24.07 ± 22.16

T is time for acquiring new gene. Results are expressed as average ± SD (n = 15).
*Final value of the experiments. Proportion of beneficial mutants is that among total divergence of a sampled genome.
†Total load needed for acquiring a new gene.

for their progressive evolution. Note that it takes at least some multiple of $4N$ generations. Here we should note the essential difference between the present model and the standard model of gene substitution at fixed loci. Since we are concerned with acquiring new gene functions that did not exist before, this may not be directly related to the immediate need of the organisms in a changing environment. In other words, natural selection for such potential would not be the result of an immediate response of organisms to the change of the environment, but rather it works through the advantage of the gene system having diverse function on a long term basis. Hence the process of acquiring a new gene is concerned with the evolution of complexity.

The above consideration leads me to believe that the rate of unequal crossing-over may be adjusted in the course of evolution in such a way that it is neither too low nor too high. If it is high, the gene system may evolve rapidly. However, if it is excessively high, too many duplications and deletions may occur, and this would have deleterious effects on the organisms.

The present analysis may have a bearing on the problem of the advantage of sexual reproduction that has been discussed for many years. Fisher (26) and Muller (27) suggested that sexual reproduction is advantageous because it brings together beneficial mutant alleles in one individual. Crow and Kimura (12) and Kimura and Ohta (13) supported the Fisher–Muller theory based on mathematical analysis. However, Maynard-Smith (25) and Williams (39) criticized that it does not work under realistic conditions. Their treatments are based on the standard model of gene substitution at a fixed locus, and gene duplication is not considered. The present study suggests that sexual reproduction accelerates evolution by gene duplication. Namely, unequal interchromosomal crossing-over through sexual reproduction may bring beneficial alleles together into one chromosome, in a manner analogous to but slightly different from that previously considered.

Finally, I would like to emphasize that gene duplication is not at all rare in evolution. The organization of hemoglobin genes has changed a great deal since mammalian radiation (7). So have genes for interferons (9), immunoglobulins (8, 28), T-cell receptors (29), major histocompatibility antigens (30), growth hormones (31), γ-crystallins (32), apolipoproteins (33), silkmoth chorion (34), metallothionein (35), and major urinary protein (36), to mention only a few reported cases. Various examples of gene elongation and formation of new genes and pseudogenes are discussed by Li (37). These examples suggest that gene duplications occur continuously and that they are being tested by natural selection. Depending upon the circumstance, a multigene family or a supergene family is created. Many existing multigene families evolved a long time ago and are thought to be in a steady state with respect to the copy number and genetic diversity (for reviews, see refs. 19 and 20). In addition, some evolved into supergene families. Gene members belonging to a supergene family have more differentiated functions than those of a multigene family and are not characterized by concerted evolution. Immunoglobulin supergene family is a most exciting example (38). If diversifying selection as considered in this paper continues to work, a supergene family would be created.

I thank Dr. Motoo Kimura for going over the first draft of this paper and providing many comments that greatly improved the presentation and Dr. Kenichi Aoki for his many valuable suggestions, particularly for pointing out an error in Eq. 4. Thanks are also due to Dr. Hidenori Tachida and Dr. Wen-Hsiung Li for their many useful comments. This work is supported by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan. This is contribution 1732 from the National Institute of Genetics, Mishima, Japan.

1. Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, Berlin).
2. Dayhoff, M. O., ed. (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3.
3. Doolittle, R. F. (1981) *Science* **214,** 149–159.
4. Ohta, T. (1987) *J. Theor. Biol.* **124,** 199–211.
5. Ohta, T. (1987) *Genetics* **115,** 207–213.
6. Ohta, T. (1988) *Evolution* **42,** 375–386.
7. Jeffreys, A. (1982) in *Genome Evolution*, eds. Dover, G. A. & Flavell, R. B. (Academic, London), pp. 157–176.
8. Hinds, K. R & Litman, G. W. (1986) *Nature (London)* **320,** 546–549.
9. Allen, G. & Fantes, K. H. (1980) *Nature (London)* **287,** 408–411.
10. Kimura, M. & Crow, J. F. (1964) *Genetics* **49,** 725–738.
11. Kimura, M. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 522–526.
12. Crow, J. F. & Kimura, M. (1965) *Am. Nat.* **99,** 439–450.
13. Kimura, M. & Ohta, T. (1971) *Theoretical Aspects of Population Genetics* (Princeton Univ. Press, Princeton, NJ).
14. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, England).
15. Tobari, Y. N. & Kojima, K. (1972) *Genetics* **70,** 397–403.
16. Mukai, T. & Cockerham, C. C. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 2514–2517.
17. Neel, J. V., Satoh, C., Goriki, K., Fujita, M., Takahashi, N., Asakawa, J. & Hazama, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 389–393.
18. Ohta, T. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5108–5112.
19. Ohta, T. (1980) *Evolution and Variation of Multigene Families* (Springer, Berlin).
20. Ohta, T. (1983) *Theor. Popul. Biol.* **23,** 216–240.
21. Ohta, T. (1984) *J. Mol. Evol.* **20,** 274–280.
22. Gojobori, T. & Nei, M. (1984) *Mol. Biol. Evol.* **1,** 195–212.
23. Shapira, S. K. & Finnerty, V. G. (1986) *J. Mol. Evol.* **23,** 159–167.
24. Spofford, J. (1969) *Am. Nat.* **103,** 407–432.
25. Maynard-Smith, J. (1978) *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, England).
26. Fisher, R. A. (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford).
27. Muller, H. J. (1932) *Am. Nat.* **66,** 118–138.
28. Marchalonis, J. J. & Schonfeld, S. A. (1970) *Biochim. Biophys. Acta* **221,** 604–611.
29. Arden, B., Klotz, J. L., Siu, G. & Hood, L. E. (1985) *Nature (London)* **316,** 783–787.
30. Steinmetz, M., Winoto, A., Minard, K. & Hood, L. (1982) *Cell* **28,** 489–498.
31. Barsh, G. S., Seeburg, P. H. & Gelinas, R. E. (1983) *Nucleic Acids Res.* **11,** 3939–3959.
32. den Dunnen, J. T., Moormann, R. J. M., Lubsen, N. H. & Schoenmakers, J. G. G. (1986) *J. Mol. Biol.* **189,** 37–46.
33. Luo, C.-C., Li, W.-H., Moore, M. N. & Chan, L. (1986) *J. Mol. Biol.* **187,** 325–340.
34. Lecanidou, R., Rodakis, G. C., Eickbush, T. H. & Kafatos, F. C. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 6514–6518.
35. Hamer, D. H. (1986) *Annu. Rev. Biochem.* **55,** 913–951.
36. Ghazal, P., Clark, A. J. & Bishop, J. O. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4182–4185.
37. Li, W.-H. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Koehn, R. K. (Sinauer, Sunderland, MA), pp. 14–37.
38. Hunkapiller, T. & Hood, L. (1986) *Nature (London)* **323,** 15–16.
39. Williams, G. C. (1975) *Sex and Evolution* (Princeton Univ. Press, Princeton, NJ).