Genet. Res., Camb. (1989), **53**, *pp.* 141–145 *Printed in Great Britain*

The mutational load of a multigene family with uniform members

TOMOKO OHTA
National Institute of Genetics, Mishima 411, Japan

(*Received 25 May 1988 and in revised form 26 September 1988*)

Summary

The mutational load of a multigene family with uniform members was studied by computer simulations. Two models of selection, truncation and exponential fitness, were examined, by using a simple model of gene conversion. It was found that the load is much smaller than the Haldane–Muller prediction under the truncation selection, and that it becomes approximately equal to the value calculated by the formula, $nv(1-q)/(m-nq)$, where n is the copy number, v is the rate of detrimental mutation per gene copy, m is the truncation point in terms of the number of detrimental genes eliminated, and q is the equilibrium frequency of detrimental mutation. However the equilibrium frequency cannot be analytically obtained. For the exponential fitness model, the load is close to the Haldane–Muller value. When there is no gene conversion, the load becomes larger than the cases with conversion both for the truncation and the exponential fitness models. Thus, gene conversion or other mechanisms that are responsible for contraction–expansion of mutants on chromosomes helps eliminating deleterious mutations occurring in multigene families.

1. Introduction

All species suffer genetic load due to mutations of detrimental effect (Muller, 1950). This is the burden for keeping genetic information already accumulated. The famous Haldane–Muller principle of genetic load states that the genetic death is equal to the detrimental mutation rate in haploids, and twice the mutation rate in diploids for semi-dominant mutations (Muller, 1950; Haldane, 1937, 1957). These values were obtained by assuming equilibrium between mutation and selective elimination.

When genes are redundant, this balance is not so simple. For example, when almost identical genes are tandemly arranged several hundred times as in ribosomal RNA genes, does the load become several hundred times the mutation rate per single copy? Each gene cannot behave independently, and some kind of epistasis should be present. Natural selection is thought to recognize properties of the whole family rather than those of each member (e.g. see Ohta, 1980; Dover, 1982). Of course, the situation is different when gene members are diverse, and their functions are differentiated. For gene families with uniform members, it is likely that the mutational load becomes much smaller than the total mutation rate.

A relevant previous study relating to this problem is that of Kimura & Maruyama (1966). They analysed the effect of epistasis on the mutational load under ordinary multi-locus model. Especially noteworthy is their finding that the mutational load may become very small when epistasis is truncation type. An interesting logic is that, when m detrimental mutations are eliminated through one genetic death, the load would become only $1/m$ times as large as the value predicted by the Haldane–Muller principle. The analysis of Kimura & Maruyama (1966) suggests that when m is large, the above relationship does not quite hold. King (1966) also presented a formula for the mutational load under truncation selection as the ratio of diploid genomic mutation rate to the excess of number of mutant genes in the class eliminated by selection. Recently, Kondrashov & Crow (1988) improved the formula.

In multigene families, detrimental mutants may increase or decrease on chromosomes by various molecular interaction mechanisms that are responsible for concerted evolution (Ohta, 1980, 1983*a*; Dover, 1982; Arnheim, 1983). Ohta & Dover (1984) studied the process of gene substitution in multigene families, and showed that the process is much influenced by gene conversion. In our problem, the frequency distribution of mutants is also likely to be much influenced, and the load may also depend upon

molecular interactions. The purpose of the present note is to show that the mutational load may be much smaller than the total mutation rate of a multigene family. In order to show this, extensive Monte Carlo simulations were performed.

2. Model and simulations

To make simulations easier, a most simple model of gene conversion (Ohta, 1982) is used without incorporating unequal crossing-over and duplicative transposition. This model can account for the concerted evolution, even when unequal crossing-over occurs, by replacing corresponding parameter values (Ohta, 1983*a*, *b*). The more general formulation by incorporating symmetric and inter-chromosomal conversion (Nagylaki, 1984*a*, *b*; Kaplan & Hudson, 1987) would be too complicated for the present purpose. The gene conversion model can be studied precisely, since neither shift of position of loci on the chromosome nor change in copy number occur. The present study concentrates on the range of realistic values of parameters for gene families with uniform members. Thus, the conversion rate was taken to be rather high as compared with the rate of interchromosomal recombination. See Matsuo & Yamazaki (1988) and Seperack *et al.* (1988) for evidence suggesting such range of parameter values for gene families with uniform members.

The simulated population consists of $2N$ chromosomes, each containing n tandemly arranged genes. With a constant rate, λ, per gene per generation, a gene is converted by one of the remaining $(n-1)$ genes on the same chromosome. The rate of detrimental mutations is v per gene per generation, and irreversible mutation is assumed.

Mutations at the individual loci are independent, and normal genes are stored by $+1$, and mutated genes, by -1. The rate of interchromosomal recombination is β between adjacent genes. One generation of simulation consists of mutation, gene conversion, interchromosomal recombination, sampling and selection. Gametes are sampled, and selection is carried out by haploid.

Two models of selection were studied: Truncation and exponential fitness function. As for the truncation selection, chromosomes which have m or more mutated genes are assumed to be lethal. Exponential fitness function is as follows,

$$w_i = \exp(-sl_i), \tag{1}$$

where s is the selection coefficient, and l_i is the number of mutated genes on the ith chromosome.

Monte Carlo experiments were continued either $20N$ or $30N$ generations, and genetic death and frequency of detrimental mutants were counted in the two intervals: I, $(10N+1 \sim 20N)$ generations; II, $(20N+1 \sim 30N)$ generations. These periods are thought to represent equilibrium. Sampling and selection were done by chromosomes, so the simulated population was essentially haploid. However, interchromosomal crossing-over was incorporated by random mating before sampling. Realistic values of products such as Nv, $N\lambda$ and $N\beta$ were chosen so that the results are applicable. As controls, experiments for the case of no conversion ($N\lambda = 0$) were also carried out.

3. Results

Table 1 gives results on equilibrium frequency of mutant and genetic load for the truncation selection model. Two levels of truncation point were chosen: $m = n/4+1$ and $n/2+1$, where n is the copy number. Copy number was 40, 80, or 160, each with three levels of interchromosomal crossing-over, and four levels of conversion rate. If the Haldane–Muller principle holds, the load would become 0·05 for $n = 40$, 0·1 for $n = 80$ and 0·2 when $n = 160$, since $v = 0{\cdot}00125$ and simulated population is haploid. It can be seen that the load is much smaller than these values.

The previous theory by King (1966) and by Kondrashov & Crow (1988) gives the relationship between the load and the equilibrium frequency of mutants. The load [eqn (5) of Kondrashov & Crow] is approximately expressed by the following formula in our notation,

$$L \approx nv(1-q)/(m-nq), \tag{2}$$

where q is the equilibrium frequency of mutants.

Unfortunately, the equilibrium frequency cannot be analytically obtained for the moment. It can be seen from the table that the above relationship holds for the present case.

It is particularly noted that the load is much reduced by gene conversion: when $\lambda = 0$, the load is only slightly smaller than the Haldane–Muller prediction, but as λ gets larger, the load becomes less and less. The formula (2) was developed for ordinary detrimental mutations at single-copy loci, and their frequencies are very low. In such cases, the number of mutants per individual obeys roughly the Poisson distribution, whereas the situation is very different in our model. The frequency of mutants is high in the present model where genes are redundant, and does not obey Poisson distribution. Here gene conversion is clearly effective for reducing both genetic load and equilibrium gene frequency.

Mutant genes may increase or decrease by gene conversion on chromosomes, and such expansion–contraction makes selection more efficient via increasing variance of mutant's number among individuals. This effect may be compared with that of gene conversion on mutant behaviour of multigene families that are going to substitute previous alleles (Ohta & Dover, 1984). Present results clearly show that the contraction–expansion of mutant genes is important not only for substitution but also for elimination of

https://doi.org/10.1017/S0016672300028020 Published online by Cambridge University Press

Table 1. *Results of simulations on genetic load and equilibrium gene frequency under truncation selection*

Truncation point (m)	λ	$2Nn\beta$	Time interval[a]	$n = 40$ Equi. freq.	$n = 40$ Load Eqn (2)	$n = 40$ Load Obs.	$n = 80$ Equi. freq.	$n = 80$ Load Eqn (2)	$n = 80$ Load Obs.	$n = 160$ Equi. freq.	$n = 160$ Load Eqn (2)	$n = 160$ Load Obs.
$\frac{n}{4}+1$	0·0	0	I	0·250	0·038	0·038	0·250	0·075	0·072	0·250	0·150	0·143
			II	0·250	0·038	0·037	0·250	0·075	0·075	0·250	0·150	0·141
		10	I	0·244	0·030	0·032	0·241	0·044	0·045	0·246	0·090	0·085
			II	0·233	0·023	0·023	0·242	0·045	0·044	0·246	0·090	0·087
		20	I	0·212	0·016	0·013	0·234	0·034	0·032	0·241	0·062	0·059
			II	0·225	0·019	0·018	0·236	0·036	0·037	0·245	0·086	0·084
	0·025	0	I	0·135	0·008	0·011	0·154	0·010	0·010	0·189	0·015	0·017
			II	0·123	0·007	0·007	0·128	0·008	0·006	0·199	0·017	0·020
		10	I	0·142	0·008	0·009	0·173	0·012	0·013	0·192	0·016	0·013
			II	0·137	0·008	0·009	0·163	0·010	0·012	0·194	0·016	0·018
		20	I	0·147	0·008	0·012	0·178	0·012	0·018	0·199	0·018	0·020
			II	0·118	0·007	0·006	0·167	0·011	0·012	0·200	0·018	0·020
	0·05	0	I	0·094	0·006	0·009	0·131	0·008	0·009	0·161	0·011	0·012
			II	0·106	0·006	0·103	0·137	0·009	0·011	0·166	0·012	0·014
		10	I	0·105	0·007	0·006	0·135	0·009	0·008	0·176	0·013	0·011
			II	0·105	0·007	0·006	0·151	0·009	0·012	0·177	0·013	0·012
		20	I	0·108	0·007	0·006	0·149	0·009	0·009	0·186	0·015	0·018
			II	0·109	0·007	0·005	0·157	0·010	0·012	0·183	0·014	0·015
	0·1	0	I	0·056	0·005	0·006	0·101	0·007	0·009	0·133	0·009	0·011
			II	0·057	0·005	0·008	0·097	0·007	0·010	0·146	0·010	0·015
		10	I	0·083	0·006	0·009	0·132	0·008	0·012	0·171	0·012	0·018
			II	0·086	0·006	0·008	0·132	0·008	0·011	0·170	0·012	0·017
		20	I	0·110	0·007	0·012	0·129	0·008	0·008	0·173	0·012	0·014
			II	0·101	0·006	0·009	0·140	0·009	0·011	0·183	0·014	0·024
$\frac{n}{2}+1$	0·0	0	I	0·459	0·010	0·005	0·496	0·038	0·043	0·499	0·086	0·087
			II	0·500	0·025	0·027	0·500	0·050	0·051	0·500	0·100	0·098
		10	I	0·487	0·017	0·018	0·468	0·015	0·017	0·491	0·043	0·049
			II	0·489	0·018	0·019	0·494	0·035	0·033	0·495	0·055	0·051
		20	I	0·456	0·010	0·009	0·484	0·021	0·024	0·479	0·024	0·022
			II	0·483	0·015	0·016	0·489	0·027	0·030	0·491	0·043	0·042
	0·025	0	I	0·273	0·004	0·007	0·269	0·004	0·003	0·380	0·006	0·007
			II	0·256	0·003	0·004	0·316	0·004	0·005	0·418	0·008	0·012
		10	I	0·288	0·004	0·002	0·372	0·006	0·006	0·412	0·008	0·007
			II	0·326	0·004	0·005	0·387	0·006	0·009	0·395	0·007	0·007
		20	I	0·271	0·004	0·001	0·386	0·006	0·007	0·415	0·008	0·007
			II	0·336	0·004	0·006	0·359	0·005	0·003	0·415	0·008	0·008
	0·05	0	I	0·201	0·003	0·005	0·268	0·004	0·006	0·366	0·006	0·009
			II	0·224	0·003	0·006	0·279	0·004	0·006	0·374	0·006	0·007
		10	I	0·287	0·004	0·006	0·342	0·005	0·006	0·397	0·007	0·007
			II	0·287	0·004	0·005	0·299	0·004	0·002	0·399	0·007	0·008
		20	I	0·262	0·004	0·003	0·335	0·005	0·006	0·396	0·007	0·008
			II	0·280	0·004	0·005	0·361	0·005	0·007	0·385	0·006	0·005
	0·1	0	I	0·144	0·003	0·003	0·216	0·003	0·003	0·238	0·004	0·003
			II	0·191	0·003	0·005	0·304	0·004	0·009	0·320	0·005	0·007
		10	I	0·229	0·003	0·004	0·310	0·004	0·006	0·371	0·006	0·008
			II	0·238	0·003	0·005	0·307	0·004	0·005	0·361	0·006	0·007
		20	I	0·208	0·003	0·002	0·286	0·004	0·003	0·361	0·006	0·005
			II	0·276	0·003	0·008	0·335	0·005	0·006	0·390	0·007	0·009

[a] I, $(10N+1)$th $-20N$th generations; II, $(20+1)$th $-30N$th generations.

Genetic load was calculated by the formula, $d/(1+d)$, where d is the number of genetic deaths per one surviving individual (Dr Motoo Kimura, personal communication). The values are averages for the two intervals as indicated. Other parameters are $2N = 100$ and $v = 0{\cdot}001\,25$.

https://doi.org/10.1017/S0016672300028020 Published online by Cambridge University Press

detrimental mutants. If there is neither contraction–expansion, nor interchromosomal recombination, increase of mutant frequency is solely due to mutation, and this is balanced by selection. Then chromosomes are saturated by mutants up to the truncation point, and the rate of mutations is $(n-m)v$, that is balanced by selective elimination. Therefore the load would be $(n-m)v/\{1+(n-m)v\}$. This situation is shown by the top line of Table 1. By incorporating interchromosomal recombination, the load is reduced, but gene conversion has a much larger effect on the load.

The results of the exponential fitness model are given in Table 2. It is very clear from the table that the load is much larger than that of the truncation model. In fact, the load is close to the Haldane–Muller prediction when gene conversion occurs ($\lambda > 0$). When $\lambda = 0$, the load is even larger than the Haldane–Muller value. The reason is again the small variance of the number of mutants among individuals when there is no contraction–expansion of mutant genes on chromosomes. It is noted that the frequency of mutants is considerably higher than the ordinarily predicted value under mutation-selection balance for $\lambda = 0$; $v/s = 0{\cdot}125$ or $0{\cdot}025$ in the present simulations. In the present simulation, the mutations are irreversible, and they would eventually accumulate when selection is weak. Thus, the mutant frequency examined is not truly at equilibrium. Except for the case of $\lambda = 0$, however, both the mutant frequency and the load are not different between the two time intervals, I and II. Only in the case of no conversion ($\lambda = 0$), slight increases of these values are observed in the interval II as compared with the values of the interval

Table 2. *Results of simulations on genetic load and equilibrium gene frequency under exponential fitness function*

				$n = 40$		$n = 80$		$n = 160$	
Selection coefficient	λ	$2Nn\beta$	Time interval[a]	Equi. freq.	Obs. load	Equi. freq.	Obs. load	Equi. freq.	Obs. load
0·01	0·00	0	I	0·509	0·184	0·451	0·307	0·457	0·522
			II	0·559	0·199	0·530	0·346	0·605	0·623
		10	I	0·392	0·143	0·296	0·211	0·333	0·413
			II	0·452	0·165	0·512	0·332	0·538	0·576
		20	I	0·378	0·141	0·339	0·237	0·433	0·500
			II	0·414	0·153	0·416	0·281	0·528	0·575
	0·025	0	I	0·049	0·020	0·091	0·071	0·086	0·129
			II	0·079	0·029	0·088	0·070	0·069	0·105
		10	I	0·070	0·028	0·122	0·093	0·081	0·125
			II	0·114	0·044	0·107	0·083	0·082	0·126
		20	I	0·119	0·046	0·080	0·061	0·105	0·154
			II	0·077	0·030	0·104	0·081	0·107	0·159
	0·05	0	I	0·058	0·023	0·045	0·036	0·058	0·090
			II	0·047	0·019	0·045	0·038	0·044	0·069
		10	I	0·093	0·038	0·081	0·062	0·092	0·137
			II	0·065	0·026	0·075	0·058	0·064	0·099
		20	I	0·078	0·031	0·089	0·071	0·075	0·115
			II	0·073	0·029	0·081	0·064	0·086	0·129
	0·1	0	I	0·034	0·014	0·037	0·030	0·040	0·062
			II	0·043	0·018	0·040	0·033	0·039	0·062
		10	I	0·054	0·022	0·064	0·051	0·062	0·095
			II	0·046	0·017	0·067	0·053	0·073	0·112
		20	I	0·095	0·037	0·075	0·059	0·092	0·140
			II	0·087	0·034	0·083	0·063	0·073	0·114
0·05	0·00	0	I	0·085	0·158	0·130	0·392	0·160	0·735
			II	0·233	0·371	0·224	0·597	0·326	0·929
		10	I	0·027	0·051	0·057	0·208	0·122	0·626
			II	0·031	0·060	0·053	0·191	0·120	0·616
		20	I	0·029	0·058	0·028	0·109	0·051	0·339
			II	0·025	0·051	0·042	0·157	0·068	0·422
	0·025	0	I	0·020	0·040	0·020	0·080	0·019	0·149
			II	0·018	0·035	0·020	0·078	0·021	0·158
		10	I	0·019	0·038	0·021	0·082	0·020	0·150
			II	0·022	0·045	0·022	0·088	0·021	0·165
		20	I	0·025	0·048	0·023	0·087	0·023	0·171
			II	0·025	0·051	0·020	0·080	0·022	0·167

[a] I, $(10N+1)$th–$20N$th generations; II, $(20+1)$th–$30N$th generations.
The values are averages for the two intervals as indicated. Other parameters are $2N = 100$ and $v = 0{\cdot}001\,25$.

https://doi.org/10.1017/S0016672300028020 Published online by Cambridge University Press

I. This case reduces to the ordinary population genetic model, and with the present selection intensity ($2Ns = 1$ and 5), mutations are expected to accumulate with time. The rate of accumulation appears to be not quite as expected from the single locus theory because of linkage.

When conversion is incorporated, selection becomes more effective, and mutant frequencies stop increasing beyond their equilibrium values. This may be somewhat similar to bringing the effective population size larger by allowing genes to 'migrate' between loci. Therefore selection becomes more effective, and mutation–selection balance is almost true equilibrium. When conversion occurs, i.e. $\lambda > 0$, both the frequency and the load become slightly smaller than the values of the Haldane–Muller theory. Thus the results again show that gene conversion is important for effectively eliminating detrimental mutations from the multigene family.

4. Discussion

The present simulation incorporated only gene conversion as mechanisms of the contraction–expansion of mutants. Unequal crossing-over and DNA replicative slippage would have the same effect on genetic load as gene conversion, so long as copy number remains stable. Of course, the situation would be more complicated. In general, it is thought that selective elimination of detrimental mutations is coupled with selective regulation of copy number, because natural selection would recognize the total amount of normal gene product. A remarkable example is the 5S-ribosomal RNA gene family of *Xenopus*, in which a functional gene and a pseudogene are arranged alternately (Fedoroff, 1979). In this gene family, a unit consisting of a normal gene and a pseudogene has spread on the chromosome presumably by unequal crossing-over rather rapidly as compared with change by mutation or gene conversion. Therefore, the frequency of detrimental mutants is 1/2 in this family. Also, the copy number appears to be considerably different between the closely related species of *Xenopus* for this multigene family (Hood *et al.* 1975). Thus, the rate of unequal crossing-over is thought to be high.

In spite of such complicated examples, evolutionary consideration is mostly in agreement with a simple model of approximate equilibrium with respect to copy number and genetic variation of multigene families (Ohta, 1980, 1983*a*; Dover, 1982; Arnheim, 1983). Then the present study on genetic load would be applicable to multigene families with uniform members.

Our study clearly shows that the truncation selection is effective for reducing the mutational load together with gene conversion. The truncation selection is realistic here, since natural selection would recognize the amount of normal gene products. However for multigene families with diverse copy members such as genes of immunoglobulin or cytochrome P-450, the exponential fitness model would be more appropriate, since individual gene product has important functions non-replaceable by others.

I thank Dr M. Kimura for his stimulating discussions. I also thank Dr J. F. Crow for his many valuable comments on the manuscript, particularly for pointing out an error on calculating the load, and two anonymous referees for their many useful suggestions. This work is supported by a grant-in-aid from the Ministry of Education, Science and Culture of Japan. Contribution no. 1774 from the National Institute of Genetics, Mishima, 411 Japan.

References

Arnheim, N. (1983). Concerted evolution of multigene families. In *Evolution of Genes and Proteins* (ed. M. Nei and R. K. Koehn), pp. 38–61. Sunderland, Mass.: Sinauer.

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

Fedoroff, N. V. (1979). On spacers. *Cell* **16**, 697–710.

Haldane, J. B. S. (1937). The effect of variation on fitness. *American Naturalist* **71**, 337–349.

Haldane, J. B. S. (1957). The cost of natural selection. *Journal of Genetics* **55**, 511–524.

Hood, L., Campbell, J. H. & Elgin, S. C. R. (1975). The organization, expression, and evolution of antibody genes and other multigene families. *Annual Review of Genetics* **9**, 305–353.

Kaplan, N. L. & Hudson, R. R. (1987). On the divergence of genes in a multigene family. *Theoretical Population Biology* **31**, 178–194.

Kimura, M. & Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351.

King, J. L. (1966). The gene interaction component of the genetic load. *Genetics* **53**, 403–413.

Kondrashov, A. S. & Crow, J. F. (1988). King's formula for the mutation load with epistasis. *Genetics* (in the press).

Matsuo, Y. & Yamazaki, T. (1988). Nucleotide variation and divergence in the histone multigene family in *Drosophila melanogaster*. *Genetics* (in the press).

Muller, H. J. (1950). Our load of mutations. *American Journal of Human Genetics* **2**, 111–176.

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

Nagylaki, T. (1984*b*). Evolution of multigene families under interchromosomal gene conversion. *Proceedings of the National Academy of Sciences, USA* **81**, 3796–3800.

Ohta, T. (1980). *Evolution and Variation of Multigene Families*. Lecture Notes in Biomathematics, vol. 37. Berlin, New York: Springer-Verlag.

Ohta, T. (1982). Allelic and nonallelic homology of a supergene family. *Proceedings of the National Academy of Sciences, USA* **79**, 3251–3254.

Ohta, T. (1983*a*). On the evolution of multigene families. *Theoretical Population Biology* **23**, 216-240.

Ohta, T. (1983*b*). Time until fixation of a mutant belonging to a multigene family. *Genetical Research* **41**, 47–55.

Ohta, T. & Dover, G. A. (1984). The cohesive population genetics of molecular drive. *Genetics* **108**, 501–521.

Seperack, P., Slatkin, M. & Arnheim, N. (1988). Linkage disequilibrium in human ribosomal genes: implications for multigene family evolution. *Genetics* **119**, 943–949.