KING'S Formula for the Mutation Load with Epistasis

Alexey S. Kondrashov* and James F. Crow^{†,1}

*Research Computer Center, Academy of Sciences, Pushchino, Moscow Region 142292, USSR, and

[†]Genetics Department, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received June 29, 1988

Revised copy accepted August 4, 1988

ABSTRACT

A formula by J. L. King gives the equilibrium mutation load as $L = 2\sum u_i(1 - q_i)/(\bar{z} - \bar{x})$ in which u_i is the mutation rate to deleterious alleles at the i th locus, q_i is the frequency of mutant alleles at this locus, $\bar{x}$ is the mean number of such mutant genes per individual before selection, $\bar{z}$ is the mean number in individuals eliminated by selection, and the summation is over all relevant loci. We show that this rule is inaccurate for intense selection and that a correct formula is $L = 2\sum u_i(1 - q_i) \bar{w}/(z - \bar{x}) = 2U \bar{w}/(\bar{z} - \bar{x}) = 2U/(\bar{z} - \bar{x} + 2U)$ in which U is the mean number of new mutations per haploid genome in the population and $\bar{w}$ is the mean relative fitness before selection. If $\bar{w}/(\bar{z} - \bar{x}) < 1/2$, the mutation load is less than the Haldane value ($U \leq L \leq 2U$) and can be considerably less. In a diploid asexual population, however, with independent occurrence of mutations, $L = 1 - e^{-2U}$ regardless of the mode of selection.

IN 1937 HALDANE showed that the equilibrium reduction in mean fitness caused by deleterious mutations, the mutation load (CROW 1958, 1970), is equal to the haploid mutation rate if the mutants are recessive and twice this if they are dominant or partially so. HALDANE's calculations assumed independent (i.e., multiplicative) contributions to fitness of individual loci. J. L. KING (1966) modified the rule to include departures from this assumption when there is epistasis. His mutation load formula reads

$$L = 2\sum u_i(1 - q_i)/(\bar{z} - \bar{x}) \quad (1)$$

in which u_i is the rate of mutation to all deleterious alleles at the i th locus, q_i is the equilibrium frequency of mutant alleles at this locus, $\bar{z}$ is the mean number of mutant alleles in individuals who are eliminated by selection, and $\bar{x}$ is the mean number of mutants in the population before selection. The summation is over all relevant loci. Since q_i is very small, this is approximately

$$L = 2\sum u_i(\bar{z} - \bar{x}) \quad (2)$$

in which $\sum u_i$ is the genomic (haploid) mutation rate. That is to say, the mutation load is the diploid mutation rate divided by the excess in the number of mutant genes in those individuals eliminated by selection. This formula has been used in problems related to mutational damage in *Drosophila* (CROW and SIMMONS 1983) and humans (CROW and DENNISTON 1981).

It is clear that the formula is not correct if the number eliminated by selection is a large fraction, since in this case $\bar{z}$ and $\bar{x}$ can be very similar (see Figure

FIGURE 1.—Illustration of the terms used in this article. The abscissa, x , is the number of mutant genes per individual and $f(x)$ is the frequency of individuals with this number. A proportion L , are removed by selection; $\bar{w} = 1 - L$ are retained. The mean number of mutant alleles per individual before selection is $\bar{x}$, among those eliminated by selection is $\bar{z}$, and among those retained is $\bar{y}$. The diagram is drawn as if selection were by truncation, but this assumption is not needed; another simple fitness function is shown by the dashed line.

1). If $2\sum u_i(1 - q_i) > \bar{z} - \bar{x}$, KING's formula implies an impossible mutation load greater than 1. We present here a corrected formula that is applicable to the full range of parameter values.

FORMULAE

Ignoring back mutations, the mean number of deleterious mutant genes per individual in the next generation is

$$\bar{x}' = \bar{y} + 2\sum u_i(1 - q_i). \quad (3)$$

The proportion of the population retained by selection, $\bar{w}$, is the mean relative fitness (Figure 1). The mean number of deleterious mutants per diploid individual in the population before selection is

$$\bar{x} = \bar{w}\bar{y} + (1 - \bar{w})\bar{z}. \quad (4)$$

¹ To whom reprint requests should be addressed.

The mutation load, L , is $(W_{\max} - \bar{W})/W_{\max}$ where $W_{\max}$ is the absolute fitness of the most fit genotype(s) and $\bar{W}$ the mean fitness (CROW 1958). Since $\bar{w} = \bar{W}/W_{\max}$, $L = 1 - \bar{w}$. At equilibrium, $\bar{x}' = \bar{x}$, so that combining (3) and (4) and eliminating $\bar{y}$ leads to

$$L = 2\sum u_i(1 - q_i) \bar{w}/(\bar{z} - \bar{x}) \\ = 2\sum u_i(1 - \bar{q})\bar{w}/(\bar{z} - \bar{x}) \quad (5)$$

in which $\bar{q} = \sum u_i q_i / \sum u_i$ is the mutation-rate-weighted average of the mutant allele frequencies. When $\bar{w} \rightarrow 1$, this agrees with KING's formula (1).

Equation 5 can also be written as

$$L = 2U\bar{w}/(\bar{z} - \bar{x}) = 2U/(\bar{z} - \bar{x} + 2U) \quad (6)$$

in which $U = \sum u_i (1 - q_i)$ is the total number of new deleterious mutations per haploid genome in the population being studied. $\sum u_i$ is the haploid rate of mutation from normal to mutant alleles, whereas U is the total number of mutations that occur; $U < \sum u_i$ because mutations from mutant alleles to other mutant alleles either do not occur or are not counted. If mutation rates are measured in standard laboratory experiments, (5) is appropriate, while (6) may be conceptually simpler. From (3), the equilibrium selection differential (FALCONER 1981) is equal to the diploid rate of occurrence of such mutations:

$$-(\bar{y} - \bar{x}) = 2U. \quad (7)$$

The minus sign is because the number of mutations is negatively related to fitness. The selection differential and the difference between the number of mutations in the eliminated and unselected group are related by $\bar{x} - \bar{y} = (\bar{z} - \bar{x})(1 - \bar{w})/\bar{w}$.

The ratio of the mutation load in (5) to that given by the KING formula (1) is $\bar{w}$. Thus, the improved accuracy becomes important if selection is intense. If $1/3$ of the population are selectively eliminated, the KING formula gives $3/2$ the correct value.

For selection in the haploid stage, the equations are still applicable but $2U$ is replaced by U .

For a single locus with recessive detrimental mutations, two mutant alleles are removed with each selective elimination while the mean number of mutants per individual before selection is $2q$. Thus $\bar{z} = 2$, $\bar{x} = 2q$, $\bar{q} = q$, $\sum u_i = u$, and (5) reduces to $L = u/(1 + u) \approx u$, as expected. For a rate dominant, $\bar{z} \approx 1$, $\bar{x} \approx 2q$, and $L \approx 2u$, also agreeing with HALDANE's (1937) results. As expressed by MULLER (1950), each selective elimination by death or infertility removes from one to two mutants depending on the dominance. But he went on to say that the load would be decreased if more than two mutants could be picked off at once. How can this happen? In natural populations the answer is synergistic epistasis, of which the extreme form is truncation selection.

As an illustration, suppose that $U = 0.5$ and $L = 0.3$. This implies, by (6), that $\bar{z} - \bar{x} = 2.33$; i.e., a

selectively eliminated individual has an average of 2.33 more mutations than a random member of the population before selection, a reasonable value with epistasis. The KING formula gives $L = 0.43$. If the mutants are eliminated independently, $L = 1 - e^{-2U}$, or 0.63 (KIMURA and MARUYAMA 1966).

The load equation can be interpreted in another way. The quantity $2U$ is the mean number of expressed lethal or sterile equivalents per individual in the equilibrium population, that is, the number of genetic deaths (MULLER 1950) if each allele exerted its effects separately (MORTON, CROW and MILLER 1956). Dividing this by $(\bar{z} - \bar{x})$ corrects for the fact that multiple mutants may be preferentially eliminated. Therefore, we can interpret $2U/(\bar{z} - \bar{x})$ as the effective number of genetic deaths (D) per individual (MULLER 1950). Substituting into (6) gives

$$L = D/(1 + D). \quad (8)$$

These formulas are quite general in that they make no assumption as to the equality of deleterious effects of individual mutations. This creates a complication for fitness functions, such as in Figure 1. There is no single-valued function of the allele number because, as also happens when environmental effects are present, different fitnesses can correspond to the same number of mutations.

In dealing with a particular selection model, one can make the simplifying assumption that all mutations have an equal effect. Alternatively, one may regard x not as the number of mutations in an individual, but as the decrease in fitness potential (MILKMAN 1978) caused by mutations in that individual. This can be done by assigning weights to each mutant proportional to its effect on fitness potential.

ASEXUAL POPULATIONS

In an asexual population the load is essentially the same as in a sexual population in which there is independent elimination of deleterious mutants. This was first shown in a pioneering paper by KIMURA and MARUYAMA (1966) and is easily seen by the following argument (CROW 1970). Assume that the proportion of the most fit genotype is p and its fitness is $W_{\max}$. Let M be the proportion of individuals with one or more deleterious alleles arising by mutation in the current generation. Then, since there is no Mendelian segregation and assuming no back mutation, the frequency of the fittest type in the next generation is $p' = p(1 - M)W_{\max}/\bar{W}$. At equilibrium, $p = p'$, and recalling that $L = (W_{\max} - \bar{W})/W_{\max}$, we immediately obtain

$$L = M. \quad (9)$$

If mutants occur independently and the population is diploid, $M = 1 - e^{-2U}$ and

$$L = 1 - e^{-2U}, \quad (10)$$

the same as with independent elimination of mutants in a sexual population.

Equation 5 remains correct for an asexual population. Equating the right parts of (10) and (6) gives an estimate of the excess number of mutant genes per individual eliminated by selection,

$$\bar{z} - \bar{x} = 2U\bar{w}/(1 - e^{-2U}). \quad (11)$$

For small mutation rates and weak selection, $\bar{z} - \bar{x} \approx 1$, as in the corresponding case for independent gene effects in a sexual population.

If recessiveness or epistasis is complete, (10) requires interpretation. For example, with complete recessivity, mutations eventually accumulate until there is only one normal allele at each locus. The formula is correct, but U is only about half as large. But complete recessiveness of any substantial fraction of deleterious mutations is unlikely. Back mutation also modifies the result.

DISCUSSION

The formula given here is an appreciable improvement over the KING formula when selection against mutant alleles is intense ($\bar{w} \ll 1$). This requires a high mutation rate. How likely is such a situation?

MUKAI (1964) and MUKAI *et al.* (1972) estimated the mutation rate of mildly deleterious viability mutations in *Drosophila* as 0.17 per second chromosome. Since the second chromosome is about $\frac{2}{5}$ of the total haploid genome, this leads to $U \approx 0.4$. This estimate depends on the ratio of a squared mean to a variance, and hence has a large sampling error even in MUKAI's heroic experiments; four independent experiments involving chromosome 2 gave values of 0.14, 0.30, 0.17 and 0.10. These estimates depend on the assumption that the deleterious effects of individual mutations are all the same, and to the extent that this is not true the calculated mutation rate is an underestimate. It could be a gross underestimate if there is a large class of very mildly deleterious mutations. Furthermore, the estimate does not include mutations affecting fertility but not viability, nor mutations whose effects are not manifest under laboratory conditions.

It might be suspected that the number of mutations in MUKAI's experiments was inflated by a high transposition rate, since the experiments involved repeated backcrossing of males carrying a chromosome from nature with laboratory females—a system that favors hybrid dysgenesis (KIDWELL, KIDWELL and SVED 1977; ENGELS 1988). The strongest evidence against any influence of hybrid dysgenesis comes from the rate of recessive lethal mutations, 0.0060 for the second chromosome (MUKAI *et al.* 1972), very close to previous estimates from laboratory strains (summarized by CROW and TEMIN 1964). Furthermore, in a study that probably *did* involve a transposable ele-

ment, MUKAI (1983) calculated a minimum mutation rate for mildly deleterious alleles of 1.95 per chromosome per generation, more than ten times the value in the earlier experiments. Curiously, the lethal rate appeared to be only about twice the normal rate, but this is based on only four mutations.

Nothing is known of the molecular basis of the viability-affecting mutations studied by MUKAI *et al.* Because of the high ratio of mildly deleterious mutations to isozyme mutations, MUKAI and COCKERHAM (1977) have suggested that most viability-depressing mutants are outside the coding regions. If this is true, the genomic rate of deleterious mutations in mammals, with their much larger amount of DNA, might be considerably larger. For all these reasons the value of U might be 1 or larger in many species, but the true value is not known for any.

Similar uncertainty clouds estimates of $\bar{w}$ and the extent of the load-reducing effects of epistasis. It has long been known to breeders that the most efficient form of selection is truncation, or rank-order, selection. Several authors have pointed out that the genetic load would be greatly reduced if a similar pattern were to occur in nature (KING 1967; MILKMAN 1967; SVED, REED and BODMER 1967). MILKMAN (1978) was the first to note that even a crude approximation to truncation selection has rather similar properties. For example, if the relative fitness of a genotype decreases uniformly from 1 at -2σ to zero at $+2\sigma$, where σ is the standard deviation of the number of mutants per individual (dashed line in Figure 1), the load-reducing effectiveness is 60% that of strict truncation selection (CROW and KIMURA 1979). Although strict truncation selection in nature seems improbable in the extreme, an approximation such as this example appears reasonable. Strong empirical evidence is lacking, however, and there is room for doubt that individuals with any number of mutants fewer than the critical number (*e.g.*, to the left of the dashed line in Figure 1) are equally fit. All that can be said with certainty is that, if the mutation rate is high, there must be some efficient load-reducing selective process.

To consider the consequences of a high mutation rate and an approximate truncation selection let $U = 1.0$ and $\bar{w} = 0.7$ ($L = 0.3$). From (6) we see that $\bar{z} - \bar{x} = 4.67$. From (7) the selection differential is 2.0 mutant alleles. This selection is less efficient than if it were by strict truncation. Suppose, as a crude approximation, that it is equivalent in mutation removal to truncation selection with $L = 0.1$. This corresponds to a selection differential for a normally distributed trait of about 0.195σ where σ is the standard deviation of the distribution. Using 2.0 as the selection differential, $\sigma = 2.0/0.195 \approx 10$ mutant alleles. If the mutant number has a Poisson distribution, approximated by the normal, the mean number of mutants per individual is about 100, which is 50 times the

zygotic mutation rate. This corresponds to a 50-generation average persistence of mutant alleles; *i.e.*, the mutant alleles are eliminated at a rate corresponding to a heterozygous selective disadvantage of 1/50 if eliminations were independent. This is in the range of estimates from experimental data, 0.01–0.03 (CROW 1979).

In this example, we have made many simplifying assumptions. But we believe the example illustrates that approximate truncation selection can reduce the mutation load so that a mutation rate of the order of 1 is tolerable and that this assumption is consistent with *Drosophila* data on the mean persistence of mildly deleterious mutations. What actually happens in nature remains to be determined. We need to know more about genomic deleterious mutation rates and the nature of the fitness function. In view of uncertainties about the parameter values, it is not clear whether our revision of the KING formula is a major improvement or only a minor refinement.

We have emphasized that quasi-truncation selection permits a sexual population to tolerate a much higher mutation rate than can an asexual one, giving one possible explanation for the prevalence of sex in nature. As here stated, this is a group selection argument (KONDRASHOV 1982, CROW and SIMMONS 1983), which is, however, sufficient to explain such things as the protection of a sexual population from invasion by an asexual clone. There are individual selection arguments as well; for example, it has been shown that with synergistic epistasis a modifier gene increasing the amount of recombination can increase (FELDMAN, CHRISTIANSEN and BROOKS 1980; KONDRASHOV 1984).

This article was written while A.S.K. was a visiting investigator at the University of Wisconsin, Madison. It is Contribution No. 2999 from the Laboratory of Genetics.

LITERATURE CITED

- CROW, J. F., 1958 Some possibilities for measuring selection intensities in man. *Hum. Biol.* **30**: 1–13.
- CROW, J. F., 1970 Genetic loads and the cost of natural selection. pp. 128–177. In: *Mathematical Topics in Population Genetics*, Edited by K.-I. KOJIMA. Springer-Verlag, New York.
- CROW, J. F., 1979 Minor viability mutants in *Drosophila*. *Genetics* **92** (Suppl.): s165–172.
- CROW, J. F., and C. DENNISTON, 1981 The mutation component of genetic damage. *Science* **212**: 888–893.
- CROW, J. F., and M. KIMURA, 1979 Efficiency of truncation selection. *Proc. Natl. Acad. Sci. USA* **76**: 396–399.
- CROW, J. F., and R. G. TEMIN, 1966 Evidence for the partial dominance of recessive lethal genes in natural populations of *Drosophila*. *Am. Nat.* **98**: 21–33.
- CROW, J. F., and M. J. SIMMONS, 1983 The mutation load in *Drosophila*. pp. 1–35. In: *The Genetics and Biology of Drosophila*, Edited by M. ASHBURNER, H. L. CARSON and J. N. THOMPSON. Academic Press, New York.
- ENGELS, W. J., 1988 P elements in *Drosophila*. In: *Mobile DNA*, Edited by D. E. BERG and M. M. HOWE. American Society for Microbiology, Washington, D.C. (in press).
- FALCONER, D., 1981 *Introduction to Quantitative Genetics*, Ed. 2. Longman, New York.
- FELDMAN, M. W., F. B. CHRISTIANSEN and L. D. BROOKS, 1980 Evolution of recombination in a constant environment. *Proc. Natl. Acad. Sci. USA* **77**: 4838–4841.
- HALDANE, J. B. S., 1937 The effect of variation on fitness. *Am. Nat.* **71**: 337–349.
- KIDWELL, M. G., J. F. KIDWELL and J. A. SVED, 1977 Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility, and male recombination. *Genetics* **86**: 813–833.
- KIMURA, M., and T. MARUYAMA, 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.
- KING, J. L., 1967 Continuously distributed factors affecting fitness. *Genetics* **55**: 483–492.
- KONDRASHOV, A. S., 1982 Selection against harmful mutations in large sexual and asexual populations. *Genet. Res.* **40**: 325–332.
- KONDRASHOV, A. S., 1984 Deleterious mutations as an evolutionary factor. I. The advantage of recombination. *Genet. Res.* **44**: 199–217.
- MILKMAN, R., 1967 Heterosis as a major cause of heterozygosity in nature. *Genetics* **55**: 493–495.
- MILKMAN, R., 1978 Selection differentials and selection coefficients. *Genetics* **88**: 391–403.
- MORTON, N. E., J. F. CROW and H. J. MULLER, 1956 An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Natl. Acad. Sci. USA* **42**: 855–863.
- MUKAI, T., 1964 The genetic structure of natural populations in *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**: 1–19.
- MUKAI, T., 1983 An extremely high rate of deleterious viability mutations in *Drosophila* possibly caused by transposons in non-coding regions. *Proc. Jpn. Acad. B* **59**: 316–319.
- MUKAI, T., and C. C. COCKERHAM, 1977 Spontaneous mutation rates at enzyme loci in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **74**: 2514–2517.
- MUKAI, T., S. I. CHIGUSA, L. E. METTLER and J. F. CROW, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**: 335–355.
- MULLER, H. J., 1950 Our load of mutations. *Am. J. Hum. Genet.* **2**: 111–176.
- SVED, J. A., T. E. REED and W. F. BODMER, 1967 The number of balanced polymorphisms that can be maintained in a natural population. *Genetics* **55**: 469–481.

Communicating editor: J. W. DRAKE