

BRIEF COMMUNICATIONS

Evolution, 57(11), 2003, pp. 2644–2646

ON THE PERSISTENCE AND PERVASIVENESS OF A NEW MUTATION

AURORA GARCÍA-DORADO,¹ ARMANDO CABALLERO,² AND JAMES F. CROW³

¹Departamento de Genética, Facultad de Ciencias Biológicas, Universidad Complutense, 28040 Madrid, Spain

²Departamento de Bioquímica, Genética e Inmunología, Facultad de Ciencias, Universidad de Vigo, 36200 Vigo, Spain

³Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706

Abstract.—It has frequently been assumed that the persistence of a deleterious mutation (the average number of generations before its loss) and its pervasiveness (the average number of individuals carrying the gene before its loss) are equal. This is true for a particular simple, widely used infinite model, but this agreement is not general. If $hs \gg 1/(4N_e)$, where hs is the selective disadvantage of mutant heterozygotes and N_e is the effective population number, the contribution of homozygous mutants can be neglected and the simple approximate formula $1/hs$ gives the mean pervasiveness. But the expected persistence is usually much smaller, $2(\log_e(1/2hs) + 1 - \gamma)$ where $\gamma = 0.5772$. For neutral mutations, the total number of heterozygotes until fixation or loss is often the quantity of interest, and its expected value is $2N_e$, with remarkable generality for various population structures. In contrast, the number of generations until fixation or loss, $2(N_e/N)(1 + \log_e 2N)$, is much smaller than the total number of heterozygotes. In general the number of generations is less than the number of individuals.

Key words.—Deleterious mutations, finite population, fitness, partial dominance, time to extinction.

Received March 14, 2003. Accepted June 18, 2003.

In 1950 Muller defined the *persistence* of a new, partially dominant, deleterious mutation as the average number of generations that the mutant gene remains in the population before elimination. Morton et al. (1956) regarded this as the reciprocal of the reduction in heterozygous fitness, as indeed it is for simple models. Later, it was suggested (Crow 1979) that it is often more meaningful to ask for the number of individuals affected by a mutation before it is eliminated. We shall call this quantity the *pervasiveness*. It has often been assumed that, if homozygotes are ignored, the two definitions lead to the same quantity, although one is measured in generations and the other in numbers of individuals. The purpose of this note is to point out that this is not generally true and the pervasiveness is often much larger than the persistence. In fact pervasiveness can never be smaller because there must be at least one individual carrying the mutant gene in each generation of persistence. Much of what we report here is not new, but we believe it has been underappreciated.

A Simple Model in Which the Two Quantities Are Equal

Consider an infinite population with a new mutation whose heterozygous deleterious effect (hs) is large enough that selection against homozygotes can be ignored. Assume that a mutant individual will pass a single copy to the next generation with probability $1 - hs$, or none with probability hs . Then, the probability that mutant gene is transmitted through generation t but not through generation $t + 1$ is $(1 - hs)^t hs$. Thus, the total number of individuals that are expected to carry the mutant during this time (the pervasiveness) is

$$\bar{p} = \sum_{t=0}^{\infty} [(t + 1)hs(1 - hs)^t] = 1/hs. \quad (1)$$

This is also the expected number of generations until the mutant's extinction (the persistence), including generation 0, and it is the reciprocal of its probability of being eliminated at any particular generation (Morton et al. 1956). Thus, in

this case there is no numerical difference between pervasiveness and persistence.

A More Realistic Model

The assumption that either a single copy or none is passed to the next generation is unrealistic. Even for an infinite population, it is more reasonable to assume that the number of copies, x , that each gene transmits to the next generation is a random variable, whose average is the gene's fitness. If, for a new mutant, x is Poisson distributed with mean $1 - hs$, the probability that a single copy mutation is lost in the next generation is $\exp[-(1 - hs)]$, which can be much larger than hs , especially for small hs . Furthermore, the number of copies transmitted while the mutant persists can be considerably larger than one. Therefore, the persistence time is expected to be smaller than the pervasiveness.

Still considering an infinite population, assume that at time $t = 0$ the mutant had an initial frequency q_0 , small enough that selection against homozygotes can be ignored. Under both the simple model and the more realistic one, the gene frequency accumulated over all future generations is

$$\sum_{t=0}^{\infty} q_t = \sum_{t=0}^{\infty} q_0(1 - hs)^t = q_0/hs. \quad (2)$$

Because the ratio of the cumulated frequency to the initial frequency ($\sum_{t=0}^{\infty} q_t/q_0 = 1/hs$) is the same as the ratio (overall number of copies contributed from $t = 0$ onwards)/(initial number of copies), the pervasiveness of a single mutation continues to be $1/hs$. Note that, since we are assuming that q_0 is small enough that homozygotes can be ignored, $1/hs$ is also the number of individuals that will be heterozygous for a single mutation, which we call the *heterozygote pervasiveness*.

For a finite population, Li and Nei (1972), using a diffusion approximation, showed that when selection against heterozygotes is the leading force ($hs \gg 1/2N_e$, where N_e is the

TABLE 1. Numerical values for heterozygote pervasiveness ($1/hs$), overall pervasiveness $(1/hs)(1 + 1/8N_ehs)$, and persistence by diffusion and branching methods for various values of hs . In column 3 N_ehs is assumed to be 10.

hs	$\bar{p} = 1/hs$	$\bar{p} = 1/hs(1 + 1/8N_ehs)$	\bar{t} (diffusion)	\bar{t} (branching)
0.001	1,000	1012.5	13.28	12.64
0.01	100	101.2	8.67	8.13
0.05	20	20.2	5.45	5.16
0.1	10	10	4.06	4.00

effective population size), the heterozygote pervasiveness is approximately the same as in an infinite population, $1/hs$, whereas the overall pervasiveness, including homozygotes, is approximately

$$\frac{1}{hs} \left(1 + \frac{1}{8N_ehs} \right)$$

in terms of the number of affected individuals, and

$$\frac{1}{hs} \left(1 + \frac{1}{4N_ehs} \right)$$

in terms of number of gene copies. The last term is for homozygotes. This expression also shows forcefully that in finite populations $1/hs$ approximates the pervasiveness only when $hs \gg 1/4N_e$ so that homozygotes can be ignored.

In contrast, under the same assumption ($N_e \gg 1/4hs$ and assuming $N_e = N$, where N is the actual population size), the persistence time is approximately

$$\bar{t} = 2 \left[\log_e \left(\frac{1}{hs} \right) + 1 - \gamma \right] \quad (3)$$

where γ (≈ 0.5772) is Euler's constant (Kimura and Ohta 1969). For smaller effective size, the persistence can also be approximated by a diffusion model as the average number of segregating sites when there is a single new mutation per generation (Kimura 1969).

For an infinite population, the persistence can also be obtained in a more exact way, using branching process theory. The probability of loss up to generation t (l_t) can be computed by numerical iteration as

$$l_t = \exp[-(1 - l_{t-1})(1 - hs)]. \quad (4)$$

with initial value l_0 (e.g., Gale 1990, Ch. 4). Therefore the persistence can be computed as

$$\bar{t} = \sum_{t=0}^{\infty} (1 - l_t). \quad (5)$$

Both the diffusion and the branching process results can be applied to a finite population where the actual population size, N , is larger than the effective size, N_e (see Kimura and Ohta 1969, Gale 1990, Ch. 4; Caballero et al. 1996).

Table 1 shows, for different heterozygous effects of a single new mutant, the pervasiveness \bar{p} in an infinite population, in a finite one where $N_e = 10/hs$, and the persistence (\bar{t}) calculated by expressions (3) and (4–5). It can be seen that homozygotes contribute very little to the pervasiveness. Regarding persistence, diffusion approximation results can be obtained without numerical iteration (from 3), and are in good

agreement with branching process values. It is clear that the pervasiveness can be much larger than the persistence.

Neutral Mutations

The cumulated frequency of heterozygotes over all future generations caused by a neutral gene in a finite population is

$$\sum H_t = \sum_{t=0}^{\infty} H_0 \left(1 - \frac{1}{2N_e} \right)^t = 2N_e H_0 \quad (6)$$

where H_0 is the initial heterozygosity. Thus, in analogy to Eq. (2), $2N_e$ is the number of individuals that will be heterozygous for a new mutant before its fixation or extinction (for a simple alternative approach, see Crow 1972). At mutation-drift equilibrium, this is also the ratio of the equilibrium heterozygosity [$H_e = H_e(1 - 1/2N_e) + H_m$] to that introduced by mutation per generation (H_m), that is $H_e/H_m = 2N_e$. This is also the ratio of the equilibrium genetic variance to the mutational variance (Lynch and Hill 1986).

The probability of eventual loss is $1 - 1/2N$ and the average time until this happens is $2(N_e/N)\log_e(2N)$ generations (for diffusion approximation, see Kimura and Ohta 1969). As fixation takes on the average $4N_e$ generations (Kimura and Ohta 1969) the average time the gene segregates in the population is approximately

$$t \approx \frac{1}{2N} 4N_e + \left(1 - \frac{1}{2N} \right) \frac{N_e}{N} 2 \log_e(2N) \\ \approx 2 \frac{N_e}{N} [1 + \log_e(2N)], \quad (7)$$

which can be much smaller than the number of heterozygotes $2N_e$. In the 1920s Fisher (1930, p. 91), with his usual ingenuity, found that the time to fixation or extinction for $N_e = N$ is $2[g + \log_e 2N] + 0.200645$. According to Ewens (1979, p. 19) for large N this is accurate to at least five decimal places.

Maruyama (1971) has shown that, remarkably, the value $2N_e$ does not require random mating; it also holds for a structured population. Furthermore, of the $2N_e$ total heterozygotes, one-third are those on the way to fixation and two-thirds on the way to loss, again regardless of population structure (Maruyama 1972). Thus, the expression $2N_e$ has great generality.

In conclusion, the mean number of generations that a neutral or deleterious mutation remains segregating in the population can be very much less than the expected number of individuals carrying the mutation before it is fixed or lost.

ACKNOWLEDGMENTS

We thank M. Lynch and an anonymous reviewer for helpful comments on the manuscript. This work was supported by grants BMC2002-00476 and BOS2000-0896 (Ministerio de Ciencia y Tecnología of Spain), 64102C124 (Universidade de Vigo, Spain), and PGIDT01PX130104PN (Xunta de Galicia, Spain).

LITERATURE CITED

Caballero, A., M. Wei, and W. G. Hill. 1996. Survival rates of mutant genes under artificial selection using individual and family information. *J. Genet.* 75:63–80.

- Crow, J. F. 1972. The dilemma of nearly neutral mutations: how important are they for evolution and human welfare. *J. Hered.* 63:306–316.
- . 1979. Minor viability mutants in *Drosophila*. *Genetics* 92: s165–s172.
- Ewens, W. J. 1979. *Mathematical population genetics*. Springer-Verlag, Berlin.
- Fisher, R. A. 1930. *The genetical theory of natural selection*. Oxford Univ. Press, Oxford, U.K.
- Gale, J. S. 1990. *Theoretical population genetics*. Unwin Hyman, London.
- Kimura, M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61:893–903.
- Kimura, M., and T. Ohta. 1969. The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* 63:701–709.
- Li, W.-H., and M. Nei. 1972. Total number of individuals affected by a single deleterious mutation in a finite population. *Am. J. Hum. Genet.* 24:667–679.
- Lynch, M., and W. G. Hill. 1986. Phenotypic evolution by neutral mutation. *Evolution* 40:915–935.
- Maruyama, T. 1971. An invariant property of a structured population. *Genet. Res.* 18:81–84.
- Maruyama, T. 1972. Some invariant properties of a geographically structured finite population: distribution of heterozygotes under irreversible mutation. *Genet. Res.* 20:141–149.
- 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. U.S.A.* 42:855–863.
- Muller, H. J. 1950. Our load of mutations. *Am. J. Hum. Genet.* 2: 111–176.

Corresponding Editor: S. Elena