

SHORT PAPERS

The rate of mutant substitution in populations with overlapping generations*

By EDWARD POLLAK

Department of Statistics, Iowa State University, Ames, Iowa 50011, USA

(Received 12 August 1981 and in revised form 5 January 1982)

SUMMARY

We consider an age-structured population that is observed at times $t = 0, 1, 2, \dots$. It is assumed that for each t there is the same number of individuals of a particular sex and age group. Another assumption we make is that an offspring of a specified sex has at all times the same probability of having a parent of a particular age and sex. It is shown that the rate of substitution of neutral mutants is \bar{v}/L , where \bar{v} and L are respectively equal to the mean fraction of mutants among gametes succeeding in forming newborn individuals and the mean age of reproduction. This result also applies to monoecious populations. The substitution rate is also derived for advantageous mutants in a monoecious population. Once again, the mutation rate in the usual expression is replaced by \bar{v}/L . Implications of these results are discussed.

1. INTRODUCTION

One of the arguments that has been used in support of the neutral mutation theory is that the rate of mutant substitutions has been approximately constant and that this would only occur if the mutants that ultimately become fixed are at least approximately neutral. The mathematics upon which this reasoning is based is applicable to populations with discrete generations. The theory has been applied, however, to data from distantly related species, and such species have, in many cases, developed different life tables and fertility schedules since they have diverged from common ancestors. It is therefore of interest to derive a theory for the rate of mutant substitution in a population with overlapping generations.

Enough is now known about the theory for age-structured populations with an unchanging demographic structure to obtain the substitution rate for neutral mutations, as well as a general impression of what it should be when the mutants are advantageous. These rates will be derived in sections 2 and 3 and discussed in section 4.

2. THE DERIVATION FOR NEUTRAL ALLELES

We consider a population that is observed at times $t = 0, 1, 2, \dots$. An individual will be said to be in age group i at time t if it is between i and $i + 1$ units of age at that time, where $i = 1$ among the newborn. It shall be assumed that, for each value of t ,

* Journal paper no. J-10369 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project 1669. Partial support by National Institutes of Health, Grant GM 13827.

N_i^m = the number of males in age group i , $i = 1, \dots, a_m$,
 N_j^f = the number of females in age group j , $j = 1, \dots, a_f$,
 $\frac{1}{2}p_i^m$ = P (a gene in a male comes from a father in age group i),
 $\frac{1}{2}p_j^m$ = P (a gene in a male comes from a mother in age group j),
 $\frac{1}{2}p_i^f$ = P (a gene in a female comes from a father in age group i),
 $\frac{1}{2}p_j^f$ = P (a gene in a female comes from a mother in age group j).

Then if v_i and v_j are respectively equal to the probabilities that a gamete from a father in age group i and from a mother in age group j contain a new mutant, the expected numbers of mutants entering the population at any particular time in males and females of age group 1 are

$$M_m = 2N_1^m \frac{1}{2} \left(\sum_{i=1}^{a_m} p_i^m v_i + \sum_{j=1}^{a_f} p_j^m v_j \right) \tag{1}$$

$$= 2N_1^m \bar{v}_m$$

and

$$M_f = 2N_1^f \frac{1}{2} \left(\sum_{i=1}^{a_m} p_i^f v_i + \sum_{j=1}^{a_f} p_j^f v_j \right) \tag{2}$$

$$= 2N_1^f \bar{v}_f.$$

Now a substitution takes place when a mutant that was originally present in one individual becomes fixed in a population. Following Kimura & Ohta (1971), we define the rate of substitution per unit time to be

$$k = \lim_{T \rightarrow \infty} n(T)/T, \tag{3}$$

where $n(T)$ is the cumulative number of mutants that have been fixed in a population during a time period of length T .

The expression $n(T)$ may be written as

$$n(T) = \sum_{\tau=1}^T \sum_{t_1=1}^{\tau} n_{1t_1}(\tau) + \sum_{\tau=1}^T \sum_{t_1=1}^{\tau} n_{2t_1}(\tau), \tag{4}$$

where $n_{1t_1}(\tau)$ and $n_{2t_1}(\tau)$ are respectively equal to the numbers of mutants fixed at time τ that originated t_1 units of time earlier in newborn males and females. Thus, we have from (1) and (2) that

$$E[n_{1t_1}(\tau)] = 2N_1^m \bar{v}_m (U_{1t_1} - U_{1, t_1-1}),$$

$$E[n_{2t_1}(\tau)] = 2N_1^f \bar{v}_f (U_{2t_1} - U_{2, t_1-1}),$$

where, for example, U_{1t_1} is equal to the probability that a mutant originating in a newborn male becomes fixed within t_1 units of time. Hence

$$E \left[\sum_{t_1=1}^{\tau} n_{1t_1}(\tau) \right] = 2N_1^m \bar{v}_m \sum_{t_1=1}^{\tau} (U_{1t_1} - U_{1, t_1-1})$$

$$= 2N_1^m \bar{v}_m U_{1\tau}$$

and, similarly,

$$E \left[\sum_{t_1=1}^{\tau} n_{2t_1}(\tau) \right] = 2N_1^f \bar{v}_f U_{2\tau}.$$

It therefore follows from the law of large numbers that

$$n(T) \sim T[2N_1^m \bar{v}_m U_1 + 2N_1^f \bar{v}_f U_2], \quad T \rightarrow \infty,$$

where U_1 and U_2 are the probabilities of ultimate fixation of mutants originating in

newborn males and females. Therefore (3) takes the form

$$k = 2N_1^m \bar{v}_m U_1 + 2N_1^f \bar{v}_f U_2 \tag{5}$$

whether mutants are neutral or not.

It has been shown by Emigh & Pollak (1979) that, if reproduction is not periodic and there is no selection,

$$U_1 = \frac{1}{2LN_1^m}, \tag{6}$$

$$U_2 = \frac{1}{4LN_1^f}, \tag{7}$$

where

$$L = \text{the average of the mean ages of parents} \\ = \frac{1}{4} \left[\sum_{i=1}^{a_m} i(p_i^m + p_i^f) + \sum_{j=1}^{a_f} j(p_j^m + p_j^f) \right].$$

Therefore, if there is a dioecious population, aperiodic reproduction and all mutants are neutral, we find, by substituting (6) and (7) in (5), that

$$k = \frac{1}{2L} (\bar{v}_m + \bar{v}_f) = \frac{\bar{v}}{L}. \tag{8}$$

The same sort of reasoning as we have just used may also be applied to a monoecious population. Thus, we assume that at all times 0, 1, 2, . . . ,

$$N_i = \text{the number of individuals in age group } i$$

and

$$p_i = P \text{ (a gene in a newborn individual comes from a parent in age group } i).$$

Then if v_i denotes the probability that a gamete from a parent in age group i contains a new mutant, the expected number of mutants entering the population at any particular time is

$$M = 2N_1 \sum_i p_i v_i = 2N_1 \bar{v}.$$

Expressions (3) and (4) may be replaced by

$$k = \lim_{T \rightarrow \infty} \left[\frac{\sum_{\tau=1}^T \sum_{t_1=1}^{\tau} n_{t_1}(\tau)}{T} \right],$$

where $n_{t_1}(\tau)$ is equal to the number of mutants fixed at time τ that originated t_1 units of time earlier in newborn individuals. Hence

$$E \left[\sum_{t_1=1}^{\tau} n_{t_1}(\tau) \right] = 2N_1 \bar{v} U_{\tau},$$

where U_{τ} is equal to the probability that a mutant becomes fixed within τ units of time. The law of large numbers then leads to

$$k = 2N_1 \bar{v} U, \tag{9}$$

where U is equal to the probability that a mutant that is originally present in one individual is ultimately fixed.

It has been shown by Emigh & Pollak (1979) that, if there is no selection and reproduction is aperiodic,

$$U = \frac{1}{2N_1 L}, \tag{10}$$

where $L = \sum_i ip_i$. Therefore, if only neutral genes are substituted, (9) and (10) imply that

$$k = \frac{\bar{v}}{L}. \quad (11)$$

3. THE RATE OF SUBSTITUTION FOR ADVANTAGEOUS MUTANTS IN A MONOECIOUS POPULATION

If there is a monoecious population in which mutants have a selective advantage, it has been shown by Charlesworth & Williamson (1975) and Pollak (1976) that

$$U \doteq \frac{2(\bar{c}-1)}{\sigma_c^2} \quad (12)$$

where \bar{c} and σ_c^2 are respectively equal to the mean and variance of the number of heterozygous children produced by a rare heterozygote in its lifetime. Because we are assuming that the number of individuals in each particular age group does not change with time, it is clear that the mean number of offspring of individuals carrying an advantageous mutant must decrease as the frequency of such individuals increases in the population. Nevertheless, it is reasonable to assume that (12) is still applicable if \bar{c} is taken to denote the ratio of the mean number of offspring produced by a heterozygote in its lifetime to the mean number of offspring produced in the lifetime of an individual of the type that was present before the mutant was introduced. Examination of the expression for the fixation probability of a mutant in a finite haploid population, derived by Emigh (1979), indicates that this should be true if there is random union of gametes, no selection for fecundity and intermediate viability of heterozygotes.

Also, it follows from Pollak (1979, expression (6.8)) that $\sigma_c^2 = LN_1/N_e$. Thus, (12) can be rewritten as

$$U \doteq \frac{2N_e s}{LN_1}, \quad (13)$$

where s and $2s$ are respectively equal to the selective advantages of heterozygotes and homozygotes that carry an advantageous mutant. It follows from (9) and (13) that

$$k \doteq 4N_e s \frac{\bar{v}}{L}. \quad (14)$$

4. DISCUSSION

If $L = 1$, so that generations are discrete, expressions (8), (11) and (14) agree with results given by Kimura (1968), King & Jukes (1969) and Kimura & Ohta (1971). Charlesworth (1980, p. 105), has given an expression for k with $L > 1$ that agrees with (8) if u is taken to be 'the rate of neutral mutation per time interval' and the rate of mutation in a germ cell increases linearly with age. In this case the rate of mutation per generation is equal to Lu , so that $k = u$, as given by Charlesworth.

If k is multiplied by L , the generation interval, it follows from (8), (11) and (14) that k_g , the mean number of substitutions per generation, is given by

$$k_g = \bar{v} \quad (15)$$

for neutral mutants and

$$k_g = 4N_e s \bar{v} \quad (16)$$

for mutants with a selective advantage s in monoecious populations. Hence \bar{v} and k are respectively consistent with the 'potential mutation rate per generation' and the 'substitution rate per year' of Ohta & Kimura (1971), although they do not define \bar{v} or L precisely.

Thus, if mutants are neutral, it is the substitution rate per generation that is equal to the mutation rate, and not the substitution rate per year. This means that, if two species which exist now had a common ancestral form in the distant past, and the species have since that time developed very different values of L , the substitution rate per unit time would only under special circumstances be constant in both lineages. Suppose, for example, that species 1 and 2 are distantly related and the potential mutation rate per generation and the generation interval for species i are \bar{v}_i and L_i . Then

$$k_1 = \frac{\bar{v}_1}{L_1} = \frac{\bar{v}_2}{L_2} = k_2 \quad (17)$$

only if $\bar{v}_1/\bar{v}_2 = L_1/L_2$, so that the potential mutation rates are proportional to the generation intervals.

King (1972) has asserted that \bar{v}/L , the mutation rate per unit time, is indeed constant in any one species and that such quantities are very similar in different vertebrate species. It has, however, been pointed out by Crow (1972) that classical studies of single locus mutation rates in humans, mice and *Drosophila* yield figures that are rather similar if time is measured in generations. He thus concludes that either mutation rates differ between substitutions that are neutral and those that are severely deleterious or that substitution rate measures are wrong.

Kohne, Chiscon & Hoyer (1972) are also in disagreement with King (1972), because they claim that data on primates suggest the rate of nucleotide sequence change seems to have been fastest in species with short generation times. Fitch & Langley (1976) point out that their data, as well as those of other authors contributing to the same symposium volume, suggest that the rate of substitution among the primates has slowed down. This may in part be due to larger values of L among primates than among, say, mice, rats and chickens, for (17) does not hold if \bar{v} does not increase in proportion to L when different species are compared. Vogel, Kopun & Rathenberg (1976) argue that the probability that a gamete carries a mutant increases with the number of cell divisions leading to its formation. They calculate that the generation interval is about 75 times as long among humans as among mice, whereas the numbers of cell divisions leading to spermatogenesis differ by a factor of only about 10. Thus, other things being equal, \bar{v}/L should be much smaller among species with life cycles like humans than those with life cycles like mice.

Therefore, even if there is no selection, the substitution rates per unit time need not be the same for different species. If they are the same, selection cannot be completely ruled out because we could have $k_1 = k_2$ if $N_e s \bar{v}$ is a constant multiple of L among the species. Perhaps, however, this is more implausible than (17).

Clearly, if data on substitution rates are to be used as evidence for or against neutrality of the alleles involved, estimates must be available for at least generation intervals as well as mutation rates. In addition, comparisons should be most informative for species with rather similar values of L . Even then, the collection of appropriate data may prove to be a formidable task.

My calculations are, strictly speaking, applicable only to a population with an unchanging size and demographic structure. The results may therefore furnish good approximations if a species has long been well adapted to its ecological niche and has not fluctuated greatly in numbers. Perhaps similar results can also be derived if the population size fluctuates in regular cycles, but I do not have proof of this.

REFERENCES

- CHARLESWORTH, B. (1980). *Evolution in Age-structure Populations*. Cambridge, London, New York, New Rochelle, Melbourne, Sydney: Cambridge University Press.

- CHARLESWORTH, B. & WILLIAMSON, J. A. (1975). The probability of survival of a mutant gene in an age-structured population and implications for the evolution of life histories. *Genetical Research* **26**, 1–10.
- CROW, J. F. (1972). Darwinian and non-Darwinian evolution. *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability* vol. V, pp. 1–22. Berkeley: University of California Press.
- EMIGH, T. H. (1979). The dynamics of finite haploid populations with overlapping generations. II. The diffusion approximation. *Genetics* **92**, 339–351.
- EMIGH, T. H. & POLLAK, E. (1979). Fixation probabilities and effective population numbers in diploid populations with overlapping generations. *Theoretical Population Biology* **15**, 86–107.
- FITCH, W. M. & LANGLEY, C. H. (1976). Evolutionary rates in proteins: neutral mutations and the molecular clock. In *Molecular Anthropology* (ed. M. Goodman, R. E. Tashian and J. H. Tashian), pp. 197–219. New York and London: Plenum Press.
- KIMURA, M. (1968). Evolutionary rate at the molecular level. *Nature* **217**, 624–626.
- KIMURA, M. & OHTA, T. (1971). On the rate of molecular evolution. *Journal of Molecular Evolution* **1**, 1–17.
- KING, J. L. (1972). The role of mutation in evolution. *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, vol. v, pp. 69–100.
- KING, J. L. & JUKES, T. H. (1969). Non-Darwinian evolution. *Science* **164**, 788–798.
- KOHNE, D. E., CHISCON, J. A. & HOYER, B. H. (1972). Evolution of mammalian DNA. *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, vol. v, pp. 193–210. Berkeley: University of California Press.
- OHTA, T. & KIMURA, M. (1971). On the constancy of the evolutionary rate of cistrons. *Journal of Molecular Evolution* **1**, 18–25.
- POLLAK, E. (1976). A stochastic theory for rare genes in large populations with overlapping generations. *Theoretical Population Biology* **10**, 109–126.
- POLLAK, E. (1979). Effective population numbers and mean times to extinction in monoecious populations with overlapping generations. *Mathematical Biosciences* **46**, 87–106.
- VOGEL, F., KOPUN, M. & RATHENBERG, R. (1976). Mutation and molecular evolution. In *Molecular Anthropology* (ed. M. Goodman, R. E. Tashian and J. H. Tashian), pp. 13–33. New York and London: Plenum Press.