Genet. Res., Camb. (1975), **25**, *pp.* 137–143
Printed in Great Britain

Profiles of electrophoretic alleles in natural populations

By A. H. D. BROWN, D. R. MARSHALL and L. ALBRECHT
Division of Plant Industry, CSIRO, Canberra, Australia

(*Received 12 December 1974*)

SUMMARY

The charge state model of Ohta & Kimura (1973) for the number of electrophoretically detectable alleles in a finite population, is extended to include mutations of both one and two charge changes. The effective number of alleles (n_e) is increased only slightly by this extension. Electrophoretic profiles of neutral variants are shown on average to be leptokurtic and have their odd central moments equal to zero. The expected frequency distribution of pairs of gametes which differ by 1, 2, 3, ... charge units can be obtained as the sum of the appropriate terms from two geometric series.

1. INTRODUCTION

That electrophoretic surveys can detect only a fraction of the genetic differences for proteins within populations has been clearly recognized right from the pioneering estimates of Lewontin & Hubby (1966). These authors claimed that many amino acid substitutions might occur in a protein, without making a detectable difference in the net charge. Direct evidence to substantiate this claim comes from a survey of variation in xanthine dehydrogenase within and between various species of the *Drosophila virilis* group (Bernstein, Throckmorton & Hubby, 1973). The effect of the limitations in resolving power in routine electrophoretic procedures upon the results obtained from surveys of natural populations has received considerable attention recently (Ohta & Kimura, 1973, 1974; King, 1974). An important conclusion to emerge from these analyses is that the limitations tend to explain some important discrepancies between the neutral mutation theory of molecular variants (Kimura & Ohta, 1971) and the observed properties of allozyme polymorphisms in natural populations. For example, Ohta & Kimura's (1973) analysis of a model of detectable mutations as single-step charge changes shows that the effective number of alleles (n_e) detected by electrophoresis can be much less than predicted by their original formula which assumed absolute resolution.

In this paper, we extend their model to include both single-step and two-step charge changes. Further details of the model and a discussion of its validity are given in Marshall & Brown (1974). We envisage a population of effective size N_e diploid individuals, with total mutation rate u per gamete per generation. These mutations are assumed to be selectively neutral. Of the mutations, a constant fraction, β, yields an amino acid substitution which results in a charge change of $+1$ units,

https://doi.org/10.1017/S0016672300015536 Published online by Cambridge University Press

γ result in $+2$ units, and that further fractions β and γ correspond to changes of -1 and -2 units respectively. The probability (α) that a mutation remains undetected by electrophoresis is thus $1-2\beta-2\gamma$. At equilibrium the recruitment of new variants by mutation is counterbalanced by their loss due to random genetic drift. Consider the distribution of charge states about some arbitrary zero charge in such a population and let p_i denote the total frequency of all the genes coding for a protein which carry a net charge of i units. The ordered sequence of random variables $\{p_i\}$ is termed the electrophoretic profile of the population.

2. CENTRAL MOMENTS OF THE EQUILIBRIUM ELECTROPHORETIC PROFILES

We will denote the charge carried by a randomly chosen gamete by the integer random variable I. At equilibrium, the following approximate formulae hold for the central moments of the charge distribution:

$$\begin{aligned}\mu_2 &= E[I-E(I)]^2 \simeq \theta(\beta+4\gamma),\\ \mu_3 &= E[I-E(I)]^3 \simeq 0,\\ \mu_4 &= E[I-E(I)]^4 \simeq 3[\theta(\beta+4\gamma)]^2+\theta(\beta+16\gamma)/4,\end{aligned} \tag{1}$$

where $\theta = 4N_e u$. Each of these formulae is derived by the same method, exemplified here by the derivation of the fourth moment.

The increment to the quantity $[K = \Sigma p_i(i-\bar{I})^4]$ in any one generation which arises from mutation is

$$\begin{aligned}\Delta K(m) &= \sum_i [(1-u+u\alpha)p_i + u\beta(p_{i+1}+p_{i-1}) + u\gamma(p_{i+2}+p_{i-2})](i-\bar{I})^4\\ &\quad - \sum_i p_i(i-\bar{I})^4\\ &= 12uV(\beta+4\gamma)+2u(\beta+16\gamma),\end{aligned}$$

where $\bar{I} = \Sigma i p_i$ and $V = \Sigma p_i(i-\bar{I})^2$. The effect of random sampling is to induce small changes in the $\{p_i\}$ and consequently change the fourth central moment by an amount

$$\Delta K(d) = \sum_i (p_i+\delta_i)[i-\bar{I}']^4 - \sum_i p_i[i-\bar{I}]^4,$$

where $\bar{I}' = \Sigma i(p_i+\delta_i)$. The first summation can be expanded as the binomial

$$\sum_i (p_i+\delta_i)[i-\bar{I}']^4 = \sum_i (p_i+\delta_i)[(i-\bar{I}) - \sum_j j\delta_j]^4.$$

The expectation of the expansion is obtained by noting that

$$E_\delta[\delta_i] = 0; \quad E_\delta[\delta_i\delta_j] \simeq -p_ip_j/2N_e \quad (i \neq j); \quad E_\delta[\delta_i^2] \simeq p_i(1-p_i)/2N_e$$

and the expectation of all higher powers of δ_i are of the order of $(N_e)^{-2}$ and are therefore neglected (Kimura, 1955). Hence, all terms beyond the first three of the expansion are assumed to be negligible:

$$E_\delta[\Delta K(d)] \simeq [3V^2-2K]/N_e.$$

At equilibrium, the total change in K is zero if

$$\Delta K(d)+\Delta K(m) = 0,$$

https://doi.org/10.1017/S0016672300015536 Published online by Cambridge University Press

so that the fourth moment of the equilibrium charge distribution is given approximately by

$$E[I-E(I)]^4 \simeq 3[\theta(\beta+4\gamma)]^2+\theta(\beta+16\gamma)/4.$$

The general recurrence formula for the even central moments is

$$\mu_{2n} \simeq (2n-1)\mu_2\mu_{2n-2}/2+\theta(\beta+4^n\gamma)/2n + \theta\sum_{i=1}^{n-1}(2n-1)!\,(\beta+4^i\gamma)\mu_{2n-2i}/(2i)!\,(2n-2i)! \quad (n=2,3,\ldots)$$

and that for the odd central moments is

$$\mu_{2n+1} \simeq n\mu_2\mu_{2n-1}+\theta\sum_{i=1}^{n-1}(2n)!\,(\beta+4^i\gamma)\mu_{2n+1-2i}/[(2i)!\,(2n+1-2i)!] \quad (n=2,3,\ldots) \simeq 0.$$

Thus, at equilibrium, electrophoretic profiles tend to show the following characteristics:

(i) All odd order central moments approach zero.

(ii) They are leptokurtic [$\mu_4\mu_2^{-2} > 3$]. The degree of leptokurtosis decreases as θ increases.

3. EQUILIBRIUM FREQUENCIES OF CLASSES OF HETEROZYGOTES

Following Ohta & Kimura (1973), we can define a set of expected values of transformations of the allele frequencies

$$C_j = E[\sum_i p_i p_{i+j}] \quad (j=0,1,2,\ldots).$$

In a random mating population, the quantity C_0 would represent the frequency of apparent homozygotes, and $2C_j$, the hypothetical frequencies of heterozygotes for two alleles which differ by j units of charge. Note that

$$C_0+2\sum_{j=1}^{\infty}C_j = 1;$$

and that, by conceiving the various classes of heterozygotes as 'hypothetical', the analysis below is not restricted to randomly mating populations.

Recurrence relationships between the C_j for an equilibrium population are obtainable by the method given above. For example, the effect of genetic drift on C_0 is

$$\Delta C_0(d) = E_\delta[\Sigma(p_i+\delta_i)^2]-\Sigma p_i^2 = (1-C_0)/2N_e.$$

The effect of mutation is

$$\Delta C_0(m) = \Sigma[(1-u+\alpha u)p_i+u\beta(p_{i+1}+p_{i-1})+u\gamma(p_{i+2}+p_{i-2})]^2-\Sigma p_i^2 \simeq 4\mu[\beta(C_1-C_0)+\gamma(C_2-C_0)],$$

when terms in u^2 are omitted. Thus at equilibrium the following relationship holds approximately:

$$C_0 \simeq (1+2\theta\beta C_1+2\theta\gamma C_2)/(1+2\theta\beta+2\theta\gamma).$$

https://doi.org/10.1017/S0016672300015536 Published online by Cambridge University Press

In like manner, the full set of recurrence relations can be obtained:

$$(1+2\theta\beta+2\theta\gamma)C_0 = 1+2\theta\beta C_1+2\theta\gamma C_2, \tag{2a}$$

$$(1+2\theta\beta+2\theta\gamma)C_1 = \theta\gamma C_1+\theta\beta C_0+\theta\beta C_2+\theta\gamma C_3, \tag{2b}$$

$$(1+2\theta\beta+2\theta\gamma)C_j = \theta\gamma C_{j-2}+\theta\beta C_{j-1}+\theta\beta C_{j+1}+\theta\gamma C_{j+2} \quad (j = 2,3,4,\ldots). \tag{2c}$$

These recurrence relations are analogous to Ohta & Kimura's (1973) differential equations (4) and (5). Solution of the set with the side conditions that

$$C_0+2\sum_{j=1}^{\infty} C_j = 1$$

and $0 < C_j < 1$ follows from the following procedure.

Due to the form of the relation (2c) it is possible to write (Li, 1955) that

$$C_j = a_1\lambda_1^j+a_2\lambda_2^j+a_3\lambda_3^j+a_4\lambda_4^j.$$

To obtain the λ_i we substitute $C_j = a\lambda^j$ into formula 2(c):

$$[1+2\theta\beta+2\theta\gamma] = \theta\beta[\lambda+1/\lambda]+\theta\gamma[\lambda+1/\lambda]^2-2\theta\gamma.$$

Now substitute $\kappa = \lambda+1/\lambda$, and solve for κ

$$\kappa = [-\theta\beta \pm \sqrt{\{(\theta\beta)^2+4\theta\gamma(1+2\theta\beta+4\theta\gamma)\}}]/2\theta\gamma \quad (\theta\gamma \neq 0),$$

which, together with

$$\lambda = [\kappa \pm \sqrt{\{\kappa-4\}}]/2,$$

define the four possible roots ($\theta\gamma \neq 0$):

$$\lambda_1 = [S-\theta\beta-\sqrt{\{(S-\theta\beta)^2-(4\theta\gamma)^2\}}]/4\theta\gamma,$$
$$\lambda_2 = [S-\theta\beta+\sqrt{\{(S-\theta\beta)^2-(4\theta\gamma)^2\}}]/4\theta\gamma,$$
$$\lambda_3 = [-S-\theta\beta-\sqrt{\{(-S-\theta\beta)^2-(4\theta\gamma)^2\}}]/4\theta\gamma,$$
$$\lambda_4 = [-S-\theta\beta+\sqrt{\{(-S-\theta\beta)^2-(4\theta\gamma)^2\}}]/4\theta\gamma,$$

where S is the positive square root of $[(\theta\beta+4\theta\gamma)^2+4\theta\gamma]$. It can be shown that $\lambda_3 < -1 < \lambda_4 < 0 < \lambda_1 < 1 < \lambda_2$, and that $\lambda_4 = \lambda_3^{-1}$ and $\lambda_1 = \lambda_2^{-1}$. In order that $0 < C_j < 1$ for all j, it follows that $a_2 = a_3 = 0$. The eigenvalues λ_2 and λ_3 are appropriate for the symmetric quantities C_{-j}, which are numerically equal to C_j. The values for a_1 and a_4 can be derived from the two simultaneous equations

$$a_1[1+2\theta\beta(1-\lambda_1)+2\theta\gamma(1-\lambda_1^2)]+a_4[1+2\theta\beta(1-\lambda_4)+2\theta\gamma(1-\lambda_4^2)] = 1,$$

which follows from relation (2a), and

$$C_0+2\sum_{j=1}^{\infty} C_j = 1$$

or

$$a_1(1+\lambda_1)/(1-\lambda_1)+a_4(1+\lambda_4)/(1-\lambda_4) = 1.$$

Thus, the expected values of the classes of heterozygotes (C_j) can now be found as the sum of terms from two geometric series $C_j = a_1\lambda_1^j+a_4\lambda_4^j$.

https://doi.org/10.1017/S0016672300015536 Published online by Cambridge University Press

4. SIMULATION EXPERIMENTS

The formulae derived above are admittedly approximate, so the process was simulated in order to check their accuracy. We followed the procedures outlined by Ohta & Kimura (1974) with the exception that mutations of two charge changes were incorporated. The values for $\beta = 0{\cdot}12$ and $\gamma = 0{\cdot}01$ have been derived elsewhere (Marshall & Brown, 1974). In each experiment the mean value for each of various parameters were computed from more than 300 consecutive generations after apparent equilibrium. The effective population size in all experiments was $N_e = 100$. There were 20 replicate experiments of each of four mutation rates ($u = 0{\cdot}0025, 0{\cdot}01,$ $0{\cdot}04$ and $0{\cdot}1$). The overall means and their standard errors were then computed from the 20 replicates. Table 1 lists the observed moments with their theoretical expectations according to formulae (1). The agreement between observed and expected is satisfactory, despite the considerable sampling variation to which these statistics are subject. Table 2 shows the good agreement between the observed means of C_0, C_1, C_2, C_3, and C_4 and their expected values.

Table 1. *Comparison of the moments of the charge distribution in simulated populations with their theoretical approximate values*

$N_e u$	0·25	1·0	4·0	10·0
$E[I-E(I)]^2$	0·16	0·64	2·56	6·4
Observed	0·12	0·71	1·84*	6·6
S.E.	0·03	0·13	0·20	1·1
$E[I-E(I)]^3$	0	0	0	0
Observed	0·02	0·10	0·73	−0·2
S.E.	0·02	0·17	0·48	2·7
$E[I-E(I)]^4$	0·15	1·5	20·8	126
Observed	0·15	2·4	13·9	176
S.E.	0·05	0·7	3·6	50

S.E. is the standard error of the observed mean of 20 replicates.

5. DISCUSSION

The importance of models which take account of the electrophoretic behaviour of mutant proteins has only recently been realized. Using a model of mutation causing single charge changes, Ohta & Kimura (1973) argued that the paucity of electrophoretically detected alleles which Ayala (1972) noted, is explicable in large measure. However, this explanation could be challenged because charge changes of two steps (such as when a basic amino acid is substituted for an acidic one) can occur and thereby generate further detectable variability not included in the simpler model.

When Ohta & Kimura's model is extended to encompass both one and two charge changes, and the appropriate values of their relative occurrence substituted ($\beta/\gamma \simeq 12$), we find that the anticipated level of homozygosity is only marginally

https://doi.org/10.1017/S0016672300015536 Published online by Cambridge University Press

decreased. This is shown in Table 2 by comparing our theoretical value for C_0 with that of the Ohta & Kimura model in which the rates of one and two charge changes are combined $[1+4\theta(\beta+\gamma)]^{-\frac{1}{2}}$. Therefore their main conclusion is robust to this criticism.

Table 2. *Comparison of the observed and theoretical values of the C_j*

	$N_e u$	0·25	1·0	4·0	10·0
C_0	Theoretical	0·810	0·561	0·315	0·202
	Observed	0·842	0·561	0·328	0·200
	S.E.	0·032	0·030	0·010	0·010
C_1	Theoretical	0·079	0·147	0·155	0·129
	Observed	0·068	0·137	0·159	0·128
	S.E.	0·013	0·007	0·005	0·007
C_2	Theoretical	0·014	0·049	0·086	0·088
	Observed	0·011	0·049	0·087	0·087
	S.E.	0·004	0·007	0·005	0·005
C_3	Theoretical	0·002	0·016	0·046	0·059
	Observed	0·001	0·022	0·050	0·059
	S.E.	0·001	0·005	0·005	0·002
C_4	Theoretical	0·0003	0·005	0·025	0·040
	Observed	0·000	0·010	0·026	0·039
	S.E.	0·000	0·005	0·005	0·002
$[1+4\theta(\beta+\gamma)]^{-\frac{1}{2}}$		0·811	0·570	0·328	0·214

In addition, these models possess two properties which are important because they hold irrespective of the value of θ. These are:

(1) The odd order central moments of electrophoretic profiles tend to zero. This is a necessary but not sufficient condition that they tend to be symmetric, and offers a ready explanation of why the modal allele frequency should be intermediate in electrophoretic mobility.

(2) The expected frequencies of various classes of hypothetical heterozygotes form a series which is obtainable as the sum of two geometric series, and when γ is small relative to β, as a single geometric distribution.

The profiles of natural populations may be expected to show discrepancies from these properties. Should these discrepancies systematically exceed the fluctuations from genetic drift, then we may conclude that either the assumptions concerning the mutation process are inapplicable, or that selection plays a role in determining such profiles. This selection may be of the stabilizing or balancing mode or both.

We wish to thank Drs A. Grassia, B. D. H. Latter, P. A. P. Moran and B. S. Weir for critically reading the manuscript and making useful comments.

https://doi.org/10.1017/S0016672300015536 Published online by Cambridge University Press

REFERENCES

AYALA, F. J. (1972). Darwinian versus non-Darwinian evolution in natural populations of *Drosophila*. *Proceedings 6th Berkeley Symposium Mathematical Statistics and Probability*, v, 211–236.

BERNSTEIN, S. C., THROCKMORTON, L. H. & HUBBY, J. L. (1973). Still more genetic variability in natural populations. *Proceedings of National Academy Sciences U.S.A.* **70**, 3928–3931.

KIMURA, M. (1955). Random genetic drift in multiallelic locus. *Evolution* **9**, 419–435.

KIMURA, M. & OHTA, T. (1971). Protein polymorphism as a phase of molecular evolution. *Nature* **229**, 467–469.

KING, J. L. (1974). Isoallele frequencies in very large populations. *Genetics* **76**, 607–613.

LEWONTIN, R. C. & HUBBY, J. L. (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595–609.

LI, C. C. (1955). *Population Genetics*. University of Chicago Press.

MARSHALL, D. R. & BROWN, A. H. D. (1974). The charge-state model of protein polymorphism in natural populations. *Journal of Molecular Evolution* (in the Press).

OHTA, T. & KIMURA, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**, 201–204.

OHTA, T. & KIMURA, M. (1974). Simulation studies on electrophoretically detectable genetic variability in a finite population. *Genetics* **76**, 615–624.

https://doi.org/10.1017/S0016672300015536 Published online by Cambridge University Press