Is Knowing the Age-Order of Alleles in a Sample Useful in Testing for Selective Neutrality?

Simon Tavaré* W. J. Ewens[†,‡] and Paul Joyce[§]

**Department of Mathematics, University of Utah, Salt Lake City, Utah 84112; [†]Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104; [‡]Department of Mathematics, Monash University, Clayton, Victoria 3168, Australia; and [§]Department of Statistics, University of Washington, Seattle, Washington 98195*

Manuscript received April 13, 1988
Accepted for publication March 13, 1989

ABSTRACT

The most powerful, and most frequently used, test of selective neutrality, based on data consisting of observed allelic frequencies in a sample of genes at some locus, is the procedure of G. A. Watterson. This procedure uses the sample homozygosity F^* as the test statistic, and in effect leads to rejection of the hypothesis of selective neutrality if the observed value of F^* differs significantly from neutral theory expectations. The homozygosity statistic is invariant under relabeling of the alleles and thus cannot use any further information on the alleles which might be available. We present results which suggest that information concerning the age order of the alleles cannot be used to provide a more powerful testing procedure than that of Watterson.

WE investigate whether information on the age order of alleles in a sample of genes from a stationary population may be used to obtain a test of selective neutrality more powerful than the WATTERSON (1977) testing procedure.

Suppose that a sample of n genes at some locus A yields n_1 genes of allelic type A_1, n_2 of allelic type A_2, $\cdots$, n_k of allelic type A_k. To test for selective neutrality of the alleles at this locus, WATTERSON proposed a testing procedure which uses as test statistic the sample homozygosity F^*, defined by

$$F^* = \sum_{i=1}^{k} \left(\frac{n_i}{n}\right)^2. \tag{1}$$

Significance points for F^* were found by WATTERSON (1978) and ANDERSON (1978) [see Appendix C in EWENS (1979)] from extensive Monte Carlo simulations, using as starting point the joint distribution of $n_1, \cdots, n_k$ conditional on n and k, under the assumption of a neutral infinitely-many-alleles model. This distribution is most easily presented by giving arbitrary labels $A_1, \cdots, A_k$ to the k alleles observed, and then calculating the probability that there are n_1 genes of the allele labeled $A_1, \cdots, n_k$ of the allele labeled A_k. This probability is

$$P(n_1, \cdots, n_k \mid k, n, \text{neutrality}) = n!/(k! \mid S_n^k \mid n_1 n_2 \cdots n_k), \tag{2}$$

where S_n^k is a Stirling number of the first kind. Note that the distribution in (2) in symmetric, so that, conditional on k, each allele frequency has expectation n/k.

A statistic equivalent to F^* is the variance-like measure F, defined as

$$F = \sum_{i=1}^{k} \left(n_i - \frac{n}{k}\right)^2. \tag{3}$$

F is a linear function of F^* and its significance points are the same linear function of those of F^*. For our purposes, it is more convenient to use F rather than F^*, and we do so from now on.

It might initially be thought [see, for example, WRIGHT (1978, p. 303)] that under selective neutrality the numbers $n_1, \cdots, n_k$ should be approximately equal, so that small values of F would suggest selective neutrality and large values some form of selection. This is not, however, true. If, for example, $n = 200$ and $k = 6$, a typical configuration of neutral allele numbers is, approximately,

i	1	2	3	4	5	6
n_i	90	55	30	15	7	3

(4)

The reason why configurations such as (4) arise under selective neutrality is that alleles enter the population at various points in the past, and an allele which entered some considerable time ago has had more chance to reach a high frequency than an allele which has only recently entered the population.

This raises the possibility that if the *age order* of the alleles in the sample is known (perhaps by a reasonable inference from the DNA sequences of the alleles), then a test of neutrality taking this age-ordering into account would be more powerful than the WATTERSON (1977, 1978) test which, because of the symmetric

way in which $n_1, \cdots, n_k$ enter F^* (and equivalently F), cannot use age order information.

Suppose then that age order information is available, and that we denote by $n_{(1)}$ the number of genes of the oldest allele in the sample, $\cdots$, $n_{(k)}$ the number of genes of the youngest allele. Then $(n_{(1)}, \cdots, n_{(k)})$ is simply a rearrangement of $(n_1, \cdots, n_k)$. DONNELLY and TAVARÉ (1986) have shown that the joint distribution of $n_{(1)}, \cdots, n_{(k)}$, given k, n and neutrality, is

$$P(n_{(1)}, \cdots, n_{(k)} \mid k, n, \text{neutrality}) = n!/(|S_n^k| \, n_{(k)}(n_{(k)} + n_{(k-1)}) \cdots (n_{(k)} + \cdots + n_{(1)})). \quad (5)$$

This formula, of course, bears a close resemblance to that in (2). In the case $n = 200$, $k = 6$, we can use (5) to find the mean values of $n_{(1)}, \cdots, n_{(k)}$: the theory of section (i) of the APPENDIX shows that these mean values are

i	1	2	3	4	5	6
$E(n_{(i)})$	94.19	52.95	28.53	14.55	6.88	2.90

(6)

This set of values indicates why, under selective neutrality, we would expect a sample configuration such as (4), since we would expect about 94 genes of the oldest allele in the sample, about 53 of the second oldest, and so on. With (6) in mind, we would clearly accept the hypothesis of selective neutrality, given the sample

i	1	2	3	4	5	6
$n_{(i)}$	90	55	30	15	7	3

(7)

which is the sample in (4) with a particular age order for the alleles observed. On the other hand, we would not be inclined to accept the hypothesis of selective neutrality given the sample

i	1	2	3	4	5	6
$n_{(i)}$	7	15	3	30	90	55

(8)

since although in this sample the allele frequencies are a rearrangement of those in (7), most allele frequencies differ considerably from their neutral theory mean values. On the other hand, the WATTERSON statistic F^* takes the same value for (7) and (8), as does F, and hence must make the same decision on accepting or rejecting neutrality for the two data sets. This makes it plausible that we can improve on the WATTERSON test by using age order information. In particular, the test statistic G defined by

$$G = \sum_{i=1}^{k} (n_{(i)} - E(n_{(i)}))^2, \quad (9)$$

which is analogous to F, but which (unlike F) should take small values under selective neutrality, might be expected to lead to a more powerful test of neutrality than does F. Thus for the "neutral-like" data set (7), $G = 24.15$, while for the 'non-neutral-like' data set (8), $G = 19{,}556.13$. Further, we obtain easily from (3) and (9)

$$E(F) = E(G) + \sum_{i=1}^{k} (E(n_{(i)}) - n/k)^2, \quad (10)$$

and we may interpret this equation in the "analysis of variance" sense as showing that of the total expected variation in the n_i values, a portion $\sum (E(n_{(i)}) - n/k)^2$ is explained by age-order information in the testing procedure. For example, for the case $n = 200$, $k = 6$ described above, the methods given in section (i) of the APPENDIX may be used to show that

$$E(F) = 12{,}171.33, \qquad E(G) = 6{,}081.13,$$
$$\sum_{i=1}^{k} (E(n_{(i)}) - n/k)^2 = 6{,}090.20$$

under selective neutrality. This may be interpreted as showing that approximately 50% of the value of F can be explained by age-order information. We would hope then that the use of G would yield a more powerful test of neutrality than would use of F. We now examine whether this is so.

THEORY FOR AGE-ORDERED ALLELES

If we are to use (9) as a test statistic, we must first calculate $E(n_{(i)})$. This is done, in principle, by using the probability distribution (5). In practice, we use a direct and simple approach, using recurrence relations, as described in section (i) of the APPENDIX. We will also consider test statistics which use the variance σ_i^2 of $n_{(i)}$, and in section (i) of the APPENDIX we give a recurrence relation from which σ_i^2 may also be calculated. The expected values in (6) are found using these recurrence relations.

The test statistic (1), or equivalently (3), was derived by WATTERSON using Neyman-Pearson statistical theory for the case where age-order information is not available. This theory leads to a powerful test of neutrality for this case. Unfortunately, the theory required when age-order information is available, in particular the analog of the distribution (5) when selection exists, is not yet available, so our choice of test statistic is a subjective one. Thus while the statistic G, defined in (9), appears a reasonable choice for the exploitation of age-order information, we have no guarantee that it leads to a powerful test of selective neutrality. Given that a sample of n genes results in k different alleles, each of the quantities $n_{(1)}, \cdots, n_{(k)}$ could be used as a test statistic. However we noted that the values of $n_{(1)}$ and $n_{(k)}$ were more influenced by selection than were the values of the remaining $n_{(i)}$, and we therefore investigated, as test statistics, the quantities $L = n_{(k)}$ and $M = n_{(1)}$. We also considered a number of other statistics which use all the $n_{(i)}$, and

we examined the power properties of all of them. These statistics are:

$$H = \sum (n_{(i)} - E(n_{(i)}))^2/\sigma_i^2;$$
$$I = \sum |n_{(i)} - E(n_{(i)})|;$$
$$J = \sum |n_{(i)} - E(n_{(i)})|/\sigma_i; \qquad (11)$$
$$K = \sum n_{(i)}/E(n_{(i)});$$
$$L = \text{number of genes of youngest allele;}$$
$$M = \text{number of genes of oldest allele.}$$

In order to find the power properties of the test of neutrality using each of these test statistics, we must first find their respective neutral theory significance points. It is clear that large values of H, I, J, K and L tend to reject neutrality, as do small values of M. The significance points must be found, as with F and F^*, by extensive Monte Carlo simulation. This simulation requires drawing random vectors $(n_{(1)}, \cdots, n_{(k)})$ from the distribution in (5) for given n and k values. The method for doing this is given in section (ii) of the APPENDIX. Estimated significance points for $n = 200$ and $k = 6$ are given in section (iii) of the APPENDIX.

We next estimate the probability with which the hypothesis of selective neutrality is rejected when selection exists, using these significance points, for each of the statistics, F, G, $\cdots$, M. The selective scheme of greatest interest is that where each homozygote has fitness 1, and each heterozygote fitness $1 + s$ (s positive), with new alleles arising by mutation according to the infinitely-many-alleles model. The hypothesis of selective neutrality is rejected in favor of heterozygote advantage when the observed value of F is too small. Unfortunately, the mathematical form of the probability distribution analogous to (5) when this form of selection exists is not known, and we therefore obtained empirical properties about power curves (that is, the probability with which the hypothesis of neutrality is rejected as a function of s) by a large-scale simulation, which is described below.

We decided first that the population from which samples would be taken would comprise 250 diploid individuals (and so 500 genes at the locus of interest). The population would reproduce according to the standard Wright-Fisher model, allowing first for the fact that heterozygotes have fitness $1 + s$, and second that, with probability u each daughter generation gene is a new mutant. For any given choice of s and u, the population was to evolve for 25,000 generations. The first 4,000 generations were to be discarded in order to overcome initial effects, so that observations were to be taken only in the final 21,000 generations.

In the sample of n genes which we decided to take from each generation, we planned to observe the values of the test statistics in (11), and then record, for each test statistic and for each observed value of k, whether or not it exceeded its neutral theory 95% significance point. We then planned to calculate the fraction of generations in which each of the statistics exceeded the appropriate neutral theory percentage point: these fractions would provide unbiased estimates of the probability of rejecting the hypothesis of neutrality for the chosen s and u values.

Because of the high autocorrelation from one generation to another there is no obvious and immediate way to calculate standard errors of these estimates. To allow for this, 32 independent replicates of the 25,000 generation evolutionary process were planned. These were to be used to compute a grand average estimate of the probability of rejecting the neutrality hypothesis, together with a standard error of this estimate.

Our preliminary simulations showed that the values of s which are of interest for power curve properties are those values between 0 and 0.2. The mean of the number k of alleles in a sample increases with s and also with u, the mutation rate. Since our aim is to estimate power curves for a fixed value of k, we therefore chose the mutation rate in such a way that the mean number of alleles observed was essentially the same for all s values. We found that to do this, the following s and u values were appropriate:

s	0.02	0.04	0.08	0.12	0.16	0.20
$u \times 10^3$	0.46	0.44	0.40	0.35	0.29	0.22

With these choices we found that in approximately 95% of the generations considered, the value of k was 4, 5, 6, 7 or 8, and in particular in approximately 35% of all generations considered the value of k was 6.

RESULTS

Essentially similar conclusions were obtained for all values of k noted, so we report here in detail only the case $n = 200$, $k = 6$. Table 1 gives the fraction of times that the null hypothesis ($s = 0$) of selective neutrality was rejected, for a variety of values of s, for each of the statistics F, G, $\cdots$, M. Figure 1 displays these values graphically.

The conclusion to be drawn from Figure 1 is unexpected. The most powerful test of neutrality is the WATTERSON test, which does not use age order information at all. Of those testing methods that do, the most powerful derives from the statistic K. Perhaps most surprising, the test statistic G has the most undesirable property that, using it, the probability of accepting the null hypothesis of selective neutrality increases steadily as the degree of selection increases!

There are two conclusions to be explained. The first is that F should lead to the most powerful test, and the second the very poor power properties of G, I and M. The first conclusion presumably arises be-

TABLE 1

Empirical power estimates

Statistic	Value of s					
	.02	.04	.08	.12	.16	.20
F	.20	.40	.74	.91	.97	.99
	(.09)	(.09)	(.08)	(.05)	(.03)	(.02)
G	.03	.02	.007	.004	.001	.001
	(.05)	(.03)	(.02)	(.016)	(.010)	(.006)
H	.15	.25	.43	.60	.71	.80
	(.11)	(.11)	(.13)	(.13)	(.14)	(.10)
I	.06	.07	.07	.08	.08	.09
	(.07)	(.07)	(.07)	(.07)	(.08)	(.07)
J	.16	.26	.44	.61	.73	.82
	(.11)	(.11)	(.13)	(.13)	(.14)	(.10)
K	.16	.31	.57	.76	.86	.93
	(.07)	(.08)	(.09)	(.08)	(.07)	(.05)
L	.01	.16	.33	.51	.64	.74
	(.06)	(.07)	(.08)	(.09)	(.12)	(.08)
M	.07	.07	.06	.05	.05	.04
	(.05)	(.04)	(.04)	(.04)	(.04)	(.03)

Empirical estimates of the probability of rejecting the null hypothesis of selective neutrality ($s = 0$), using the statistics F, G, H, I, J, K, L and M. Standard errors in parentheses. See text for details.

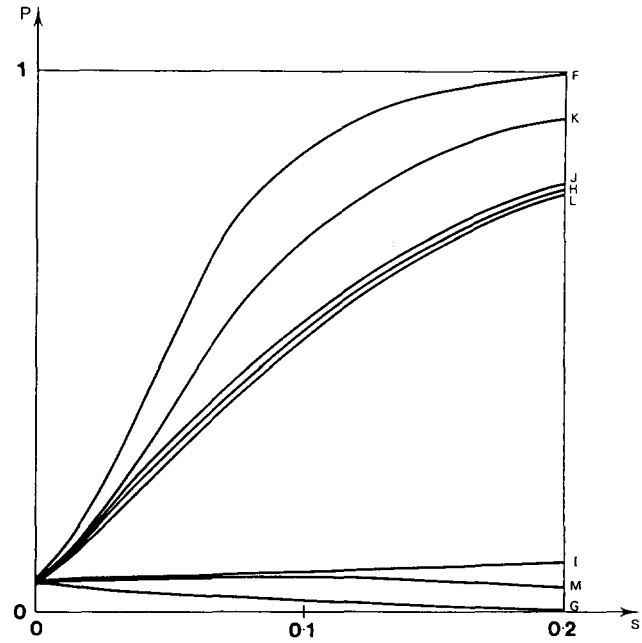

FIGURE 1.—A comparison of the empirical probability P of rejecting neutrality for different values of the selection parameter s for the test statistics F, G, $\cdots$, M.

cause the Neyman-Pearson likelihood ratio test leads to the choice of F as test statistic, even when age-order information is available. This would occur if the likelihood ratio reduced to a function of $\sum n_{(i)}^2/n^2$ only, since this function is identical to F^* and is a linear function of F. The theory required to check this conjecture, however, appears quite difficult.

The poor power properties of G are more readily explained. In the case $n = 200$, $k = 6$, the 95% significance point of G is approximately 14135. In other words, the probability that G exceeds this value under selective neutrality is approximately 5%. Consider now an extreme form of heterotic selection (for example such as occurs with self-sterility alleles), when all alleles present in a sample tend to be present in approximately equal frequencies. In the present case, the number of genes of each of the 6 alleles present in the sample would be approximately 200/6. In the extreme case where each $n_{(i)} = 200/6$, the value of G is only 6,090, well below the 95% significance level and indeed very close to the *neutral* theory mean value of G. Further, the variance in G under extreme forms of heterotic selection is quite small, so that the observed values of G vary little from the neutral theory mean value and very seldom reach the (neutral) 95% significance point. Indeed, the stronger the selection, the more likely it is that the value of G will not reach this significance point, and the extent to which this is so is shown by Table 1 and Figure 1. A similar conclusion applies to the statistic I.

A similar phenomenon arises for the statistic M, the number of genes of the oldest allelic type. Here the 95% significance point is 10, values of M less than this cut-off being significant at the 5% level. However although, when heterotic selection exists, the mean frequency of the oldest allele does decrease considerably from its neutral theory mean of 94.19, the variance also decreases considerably, and the net result is that the probability that the frequency of this allele is less than 10 actually decreases as the intensity of selection increases. The extent to which this is so is also illustrated in Table 1 and Figure 1.

Although the array of test statistics in (11) does not exhaust all possibilities, they do cover all cases which appear reasonably natural, and it is therefore plausible that age-order information cannot be used to improve on the WATTERSON test statistic F.

S.T. was supported in part by grants DMS 86-08857 and 88-03284 from the National Science Foundation. W.J.E. was supported in part by grant GM 21135 from the National Institutes of Health.

LITERATURE CITED

ANDERSON, R., 1978 Unpublished M.Sc. thesis, Monash University, Australia.

DONNELLY, P., 1986 Partition structures, Pólya urns, the Ewens sampling formula and the ages of alleles. Theor. Popul. Biol. **30:** 271–288.

DONNELLY, P., and S. TAVARÉ, 1986 The ages of alleles and a coalescent. Adv. Appl. Prob. **18:** 1–19.

EWENS, W. J., 1972 The sampling theory of selectively neutral alleles. Theor. Popul. Biol. **3:** 87–112.

EWENS, W. J., 1979 *Mathematical Population Genetics.* Springer Verlag, New York.

FUERST, P. A., R. CHAKRABORTY and M. NEI, 1977 Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. Genetics **86:** 455–483.

HOPPE, F. M., 1984 Pólya-like urns and the Ewens sampling formula. J. Math. Biol. **20:** 91–99.

HOPPE, F. M., 1987 The sampling theory of neutral alleles and an urn model in population genetics. J. Math. Biol. **25:** 123–159.

JOYCE, P., and S. TAVARÉ, 1987 Cycles, permutations and the structure of the Yule process with immigration. Stochast. Proc. Appl. **25:** 309–314.

WATTERSON, G. A., 1977 Heterosis or neutrality? Genetics **85:** 789–814.

WATTERSON, G. A., 1978 The homozygosity test of neutrality. Genetics **88:** 405–417.

WATTERSON, G. A., 1984 Estimating the divergence time of two species. Statistics Research Report 94, Monash University, Australia.

WRIGHT, S., 1978 *Evolution and the Genetics of Populations*, Vol. 4. The University of Chicago Press, Chicago.

Communicating editor: E. THOMPSON

APPENDIX

(i) Recursion equations for moments of $n_{(m)}$: Here we derive a recursive scheme for computing the moments of the $n_{(m)}$ for the neutral distribution defined in Equation 5. The method is based on the popular "genealogical urn" representation for generating neutral samples; compare HOPPE (1984, 1987), WATTERSON (1984), DONNELLY (1986), JOYCE and TAVARÉ (1987) for example. According to this theory, observations having the (unconditional) distribution of $n_{(1)}$, $n_{(2)}, \cdots$ may be generated sequentially. Given the (age-ordered) configuration of the first $n-1$ genes, the nth gene to enter the sample is a mutant (and so a novel allele) with probability $\theta/(\theta+n-1)$, or a copy of one of the first $n-1$ genes, each with probability $1/(\theta+n-1)$. Let η_n denote the (random) number of alleles in a sample of n genes. The distribution of η_n was given by EWENS (1972) as:

$$P(\eta_n = k) := u_0(n, k) = \frac{\theta^k |S_n^k|}{\theta_{(n)}}, \quad k = 1, \cdots, n, \tag{A1}$$

where $x_{(n)} = x(x+1)\cdots(x+n-1)$, and $x_{(0)} = 1$. Let ξ_n denote the number of genes of the mth oldest allele in a sample of size n, and define

$$u_r(n, k) = E(\xi_n^r I\{\eta_n = k\}), \tag{A2}$$

for $m \le k \le n$; $r = 0, 1, \cdots$, where we have defined

$$I\{A\} = \begin{cases} 1 & \text{if } A \text{ occurs;} \\ 0 & \text{if } A \text{ does not occur.} \end{cases}$$

When $r = 0$, $u_0(n, k)$ is given by (1); for $n \ge k$, these elements may be computed recursively via the equation

$$u_0(n, k) = \frac{\theta}{\theta+n-1} u_0(n-1, k-1) + \frac{n-1}{\theta+n-1} u_0(n-1, k).$$

To derive an equation for the $u_r(n, k)$, choose and fix m, and define three events A_n, B_n and C_n as follows: A_n is the event that the nth gene is a mutant, B_n is the event that the nth gene is in the mth allelic class and is not a mutant, and C_n is the event that the nth gene is not in the mth allelic class and is not a mutant. It follows that

$$\xi_n^r I\{\eta_n = k\} = \xi_n^r I\{\eta_n = k, A_n\} + \xi_n^r I\{\eta_n = k, B_n\} + \xi_n^r I\{\eta_n = k, C_n\}. \tag{A3}$$

There are three cases to consider:

Case (a) $k > m > 1$: The structure of the genealogical urn shows that

$$\begin{aligned} \xi_n^r I\{\eta_n = k, A_n\} &= \xi_{n-1}^r I\{\eta_{n-1} = k-1\} I\{A_n\}; \\ \xi_n^r I\{\eta_n = k, B_n\} &= (1+\xi_{n-1})^r I\{\eta_{n-1} = k\} I\{B_n\}; \\ \xi_n^r I\{\eta_n = k, C_n\} &= \xi_{n-1}^r I\{\eta_{n-1} = k\} I\{C_n\}. \end{aligned} \tag{A4}$$

Now condition on ξ_{n-1}, η_{n-1}, and use the fact that

$$P(A_n \mid \eta_{n-1}, \xi_{n-1}) = \frac{\theta}{\theta+n-1},$$

$$P(B_n \mid \eta_{n-1}, \xi_{n-1}) = \frac{\xi_{n-1}}{\theta+n-1},$$

$$P(C_n \mid \eta_{n-1}, \xi_{n-1}) = \frac{n-1-\xi_{n-1}}{\theta+n-1},$$

to obtain

$$\begin{aligned} E(\xi_n^r I\{\eta_n = k\}) &= EE(\xi_n^r I\{\eta_n = k\} \mid \xi_{n-1}, \eta_{n-1}) \\ &= E\Bigg(\xi_{n-1}^r I\{\eta_{n-1} = k-1\} \frac{\theta}{\theta+n-1} \\ &\quad + (1+\xi_{n-1})^r I\{\eta_{n-1} = k\} \frac{\xi_{n-1}}{\theta+n-1} \\ &\quad + \xi_{n-1}^r I\{\eta_{n-1} = k\} \frac{n-1-\xi_{n-1}}{\theta+n-1}\Bigg). \end{aligned} \tag{A5}$$

When $r = 1$, we may simplify (A5) to obtain

$$E(\xi_n I\{\eta_n = k\}) = E\left(\frac{\theta}{\theta+n-1} \xi_{n-1} I\{\eta_{n-1} = k-1\} + \frac{n}{\theta+n-1} \xi_{n-1} I\{\eta_{n-1} = k\}\right).$$

Hence for $k > m$, we have

$$u_1(n, k) = \frac{\theta}{\theta+n-1} u_1(n-1, k-1) + \frac{n}{\theta+n-1} u_1(n-1, k). \tag{A6}$$

When $r = 2$, and $k > m$, an analogous argument

gives

$$u_2(n, k) = \frac{\theta}{\theta + n - 1} u_2(n - 1, k - 1) + \frac{n + 1}{\theta + n - 1} u_2(n - 1, k) + \frac{1}{\theta + n - 1} u_1(n - 1, k). \quad \text{(A7)}$$

Case (b) $m = k > 1$: In this case, Equation A4 is to be replaced by

$$\xi_n^r I\{\eta_n = m, A_n\} = I\{\eta_{n-1} = m - 1\} I\{A_n\};$$

$$\xi_n^r I\{\eta_n = m, B_n\} = (1 + \xi_{n-1})^r I\{\eta_{n-1} = m\} I\{B_n\};$$

$$\xi_n^r I\{\eta_n = m, C_n\} = \xi_{n-1}^r I\{\eta_{n-1} = m\} I\{C_n\}.$$

Repeating the argument that leads to (A5) then shows that

$$u_1(n, m) = \frac{\theta}{\theta + n - 1} u_0(n - 1, m - 1) + \frac{n}{\theta + n - 1} u_1(n - 1, m), \quad \text{(A8)}$$

while

$$u_2(n, m) = \frac{\theta}{\theta + n - 1} u_0(n - 1, m - 1) + \frac{n + 1}{\theta + n - 1} u_2(n - 1, m) + \frac{1}{\theta + n - 1} u_1(n - 1, m). \quad \text{(A9)}$$

In the last four recurrences, $u_r(\cdot, 0) = 0$ and $u_r(j, j + 1) = 0$ for all j.

Case (c) $m = 1$: If $k > 1$, Equation A4, and hence (A6) and (A7), hold. If $k = 1$, then Equation A4 is replaced by

$$\xi_n^r I\{\eta_n = 1, A_n\} = 0,$$

$$\xi_n^r I\{\eta_n = 1, A_n^c\} = n^r I\{\eta_{n-1} = 1\} I\{A_n^c\},$$

so that

$$u_r(n, 1) = n^r \frac{n - 1}{\theta + n - 1} u_0(n - 1, 1) = n^r u_0(n, 1), \qquad r = 1, 2.$$

For a given value of m, these equations can be solved numerically, beginning with $n = m = k$, and $u_r(m, m) = u_0(m, m)$, then $n = m + 1$, $k = m, m + 1$; $n = m + 2$, $k = m, m + 1, m + 2$ and so on until the desired maximum values of n and k are reached. Finally, we calculate

$$En_{(m)}^r = \frac{u_r(n, m)}{u_0(n, m)}, \qquad r = 1, 2. \quad \text{(A10)}$$

Despite the explicit appearance of θ in these recursions, the sufficiency of k for θ [EWENS (1972)] guarantees that the moments calculated in (A10) are indeed independent of θ.

(ii) Simulation method for neutral samples: In this appendix, we describe the simulation method used to derive percentage points for the null (selectively neutral) distribution of the test statistics described in Equation 11. The idea is related to that of STEWART in the appendix of FUERST, CHAKRABORTY and NEI (1977). Recall from Equation 5 of the text that under the neutrality hypothesis the joint distribution of $n_{(1)}, \cdots, n_{(k)}$ given k is given by

$$P(n_{(1)}, \cdots, n_{(k)} \mid k) = n!/ (|S_n^k| \, n_{(k)}(n_{(k)} + n_{(k-1)}) \cdots (n_{(k)} + \cdots + n_{(1)})), \quad \text{(A11)}$$

for $n_{(1)} + \cdots + n_{(k)} = n$. It follows from this (compare DONNELLY and TAVARÉ (1986, Equation 6.1)) that the joint conditional distribution of $n_{(r+1)}, \cdots, n_{(k)}$ given $k, n_{(1)}, \cdots, n_{(r)}$ is

$$P(n_{(r+1)}, \cdots, n_{(k)} \mid k, n_{(1)}, \cdots, n_{(r)}) = \frac{(n - n_{(1)} - \cdots - n_{(r)})!}{|S_{n-n_{(1)}-\cdots-n_{(r)}}^{k-r}| \, n_{(k)}(n_{(k)} + n_{(k-1)}) \cdots (n_{(k)} + \cdots + n_{(r+1)})}. \quad \text{(A12)}$$

Comparing (A11) and (A12), it is clear that given k and the frequencies of the r oldest alleles, the remaining $k - r$ age-ordered allele frequencies behave like a sample of size $n - n_{(1)} - \cdots - n_{(r)}$. Given n and k, the probability $p(j \mid n, k)$ that the oldest allele has j representatives is given by

$$p(j \mid n, k) = \frac{(n - 1)!}{|S_n^k|} \frac{|S_{n-j}^{k-1}|}{(n - j)!}, \qquad j = 1, \cdots, n - k + 1. \quad \text{(A13)}$$

We can therefore generate an age-ordered sample recursively using (A12) and (A13). First, simulate an observation $n_{(1)}$ from $p(\cdot \mid n, k)$. Then simulate an observation $n_{(2)}$ from $p(\cdot \mid n - n_{(1)}, k - 1)$, $\cdots$, an observation $n_{(k-1)}$ from $p(\cdot \mid n - n_{(1)} - \cdots - n_{(k-2)}, 2)$. Finally, set $n_{(k)} = n - n_{(1)} - \cdots - n_{(k-1)}$.

We can simulate observations from the distribution (A13) for particular n and k values once an efficient algorithm for computing such probabilities is found. We proceeded as follows: define

$$v(n, k) = |S_n^k| / n!.$$

It follows that

$$p(j \mid n, k) = \frac{v(n - j, k - 1)}{n v(n, k)}, \qquad j = 1, \cdots, n - k + 1, \quad \text{(A14)}$$

and that the $v(n, k)$ satisfy the recursion

$$v(n, k) = \left(1 - \frac{1}{n}\right)v(n-1, k) + \frac{1}{n}v(n-1, k-1)$$

for $k = 1, \cdots, n$; $n = 2, 3, \cdots$, with initial conditions

$$v(1, 1) = 1.0, \qquad v(i, i+1) = 0.0, \qquad i = 1, 2, \cdots$$

For a given value of n and k, only the $v(m, j)$ values for $j \leq k$, $m \leq n$ have to be found. The look-up table of these values is computed just once at the beginning of the simulation, and in conjunction with (A14) it provides all the probabilities we need.

(iii) Estimated significance points for test statistics: Estimated 95% significance points of the test statistics F, G, H, I, J, K, L and M when $n = 200$, $k = 6$ were found using the method described in section (ii) of this APPENDIX. The values reported here are the average of 5 runs, each using 25,000 simulated vectors. Values of G, H, I, J, K and L above the significance point shown reject neutrality, while values of F and M below the significance point reject neutrality.

F	3831
G	16072
H	17.22
I	215.18
J	7.40
K	11.26
L	10
M	10