

A compact proof of Fisher's Fundamental Theorem for multiple loci

NARAYAN BEHERA

Developmental Biology and Genetics Laboratory, Indian Institute of Science,
Bangalore 560 012, India

MS received 2 January; revised 18 August 1995

Abstract. A systematic method is formulated to carry out theoretical analysis in a multilocus multiallele genetic system. As a special application, the Fundamental Theorem of Natural Selection is proved (in the continuous time model) for a multilocus multiallele system if all pairwise linkage disequilibria are zero.

Keywords. Evolution; natural selection; allele frequency; variance in fitness; integral variational principle.

1. Introduction

R. A. Fisher (1930) enunciated his Fundamental Theorem of Natural Selection as 'the rate of increase of mean fitness of a population at any time is equal to its genetic (that is additive genetic) variance in fitness at that time'. He dealt with a continuous population and was not clear whether the theorem could be applied to discrete generations. Kimura (1958) was the first to add explicit terms into the theorem to account for environmental change (or other causes of changes in the genotypic fitnesses). His treatment has been widely regarded as clarifying and extending Fisher's Fundamental Theorem (FFT). Price's (1972) analysis of FFT recognizes the fact that the theorem refers to the partial change. Finally Ewens's clearer formulation of the theorem led to its wider acceptance (Ewens 1989). Ewens believes that Fisher was interested in fitness changes only through changes in the gene frequency and not through changes in the expected value of the fitness. The theorem, although exact, measures only one component of fitness change. So the modern interpretation of FFT (following Ewens and Price) is: 'the rate of increase in the mean fitness of any population at any time ascribable to natural selection acting through changes in gene frequencies is exactly equal to its genic variance in the fitness at that time'. Ewens's new formulation does not require any specific mating scheme (Fisher's result is true for random mating only) and holds good in both discrete and continuous time. It is clear that the calculation of the partial change is exact and involves no approximations. This supports the argument that Fisher himself viewed the theorem as exact (for some background discussion, see Edwards 1994).

A systematic and compact way of carrying out theoretical analysis for a multilocus system has not been developed in the literature. I present here a systematic method. As a special case of the application of the method, I derive the following result: when all the pairwise linkage disequilibria are zero and gametic frequencies can

be written as a product of allelic frequencies (under random mating), a compact proof of Fisher's Fundamental Theorem for multiple loci (in the continuous time model) is possible.

2. Fisher's Fundamental Theorem for the multilocus multiallele case

For a compact proof of FFT in the general multilocus multiallele case, consider a diploid genotype $\Pi_n(A_{n_i}A_{n_j})$. Here, n stands for different loci, and each locus has l_n alleles, so that

$$\begin{aligned} i1 &= 1, \dots, l_1 & j1 &= 1, \dots, l_1 \\ &\vdots & &\vdots \\ &\text{and} & & \\ &\vdots & &\vdots \\ in &= 1, \dots, l_n & jn &= 1, \dots, l_n. \end{aligned}$$

Let the genotypic value be $\bar{a} + a_{i1j1\dots injn}$, where \bar{a} is the population mean and $a_{i1j1\dots injn}$ is the average excess of genotype $\Pi_n(A_{n_i}A_{n_j})$. Let the fitness of the above genotype be

$$m_{i1j1\dots injn} = \bar{m} + a_{i1j1\dots injn},$$

where \bar{m} is the mean fitness. Let the genic value for the above genotype be $\bar{a} + \sum_n (\alpha_{n_i} + \alpha_{n_j})$, where α_{n_i} is the genic value for A_{n_i} . We have

$$\sum_{i1} p_{i1} = \dots = \sum_{jn} p_{jn} = 1$$

and

$$\sum_{i1, \dots, in} \left(\prod_n p_{n_i} \right) = \sum_{j1, \dots, jn} \left(\prod_n p_{n_j} \right) = 1, \quad (1)$$

where p_{n_i} is the allelic frequency at the n th locus.

The sum of deviations from the mean is zero. This implies

$$\sum_{i1, \dots, in} \left[\left(\prod_n p_{n_i} p_{n_j} \right) \left(\sum_n \alpha_{n_i} + \alpha_{n_j} \right) \right] = 0$$

or

$$2 \sum_{i1, \dots, in} \left[\left(\prod_n p_{n_i} \right) \left(\sum_n \alpha_{n_i} \right) \right] \left(\sum_{j1, \dots, jn} \prod_n p_{n_j} \right) = 0$$

or

$$\left(\sum_{i1} p_{1_i} \alpha_{1_i} \right) \left(\sum_{i2, \dots, in} \prod_{n \neq 1} p_{n_i} \right) + \left(\sum_{i2} p_{2_i} \alpha_{2_i} \right) \left(\sum_{i1, i3, \dots, in} \prod_{n \neq 2} p_{n_i} \right) + \dots = 0$$

or

$$\sum_{i1, \dots, in} \sum_n (p_{n_i} \alpha_{n_i}) = 0. \quad (2)$$

In equation (2) each term is separately zero as in the case of a single-locus problem (Crow and Kimura 1970).

The genic variance V_g is given by

$$\begin{aligned}
 V_g &= \sum_{i1, \dots, in; j1, \dots, jn} \left(\prod_n p_{n_{i_n}} p_{n_{j_n}} \right) \left[\sum_n (\alpha_{n_{i_n}} + \alpha_{n_{j_n}}) \right]^2 \\
 &= \sum_{i1, \dots, in; j1, \dots, jn} \left(\prod_n p_{n_{i_n}} p_{n_{j_n}} \right) \left[\sum_n (\alpha_{n_{i_n}} + \alpha_{n_{j_n}})^2 \right] \\
 &\quad + 2 \sum_{i1, \dots, in; j1, \dots, jn} \left(\prod_n p_{n_{i_n}} p_{n_{j_n}} \right) \sum_{n < n'} \{ (\alpha_{n_{i_n}} + \alpha_{n_{j_n}}) (\alpha_{n'_{i'}} + \alpha_{n'_{j'}}) \} \\
 &= 2 \sum_{i1, \dots, in} \left(\prod_n p_{n_{i_n}} \right) \left(\sum_n \alpha_{n_{i_n}}^2 \right) \left(\sum_{j1, \dots, jn} \prod_n p_{n_{j_n}} \right) \\
 &\quad + 2 \left[\sum_{i1, \dots, in} \prod_n (p_{n_{i_n}} \alpha_{n_{i_n}}) \right] \left[\sum_{j1, \dots, jn} \prod_n p_{n_{j_n}} \alpha_{n_{j_n}} \right] \\
 &\quad + \sum_{i1, \dots, in; j1, \dots, jn} \sum_{n < n'} \left[(p_{n_{i_n}} \alpha_{n_{i_n}}) (p_{n'_{i'}} \alpha_{n'_{i'}}) \right] \prod_{n'' \neq n, n'} p_{n''_{i''}} \prod_n p_{n_{j_n}} \\
 &\quad + \sum_{i1, \dots, in; j1, \dots, jn} \sum_{n < n'} \left[(p_{n_{i_n}} \alpha_{n_{i_n}}) (p_{n'_{j'}} \alpha_{n'_{j'}}) \right] \prod_{n'' \neq n} p_{n''_{i''}} \prod_{n'' \neq n'} p_{n''_{j''}} \\
 &\quad + \sum_{i1, \dots, in; j1, \dots, jn} \sum_{n < n'} \left[(p_{n_{j_n}} \alpha_{n_{j_n}}) (p_{n'_{i'}} \alpha_{n'_{i'}}) \right] \prod_{n'' \neq n} p_{n''_{j''}} \prod_{n'' \neq n'} p_{n''_{i''}} \\
 &\quad + \sum_{i1, \dots, in; j1, \dots, jn} \sum_{n < n'} \left[(p_{n_{j_n}} \alpha_{n_{j_n}}) (p_{n'_{j'}} \alpha_{n'_{j'}}) \right] \prod_{n'' \neq n, n'} p_{n''_{j''}} \prod_n p_{n_{i_n}}, \tag{3}
 \end{aligned}$$

where $n, n', n'' \in \mathbb{Z}$, the set of positive integers.

Making use of equations (1) and (2), we get

$$\begin{aligned}
 V_g &= 2 \sum_{i1} p_{1_{i1}} \alpha_{1_{i1}}^2 \sum_{i2, \dots, in} \prod_{n \neq 1} p_{n_{i_n}} + 2 \sum_{i2} p_{2_{i2}} \alpha_{2_{i2}}^2 \sum_{i1, i3, \dots, in} \prod_{n \neq 2} p_{n_{i_n}} + \dots \\
 &= 2 \sum_{i1, \dots, in} \sum_n p_{n_{i_n}} \alpha_{n_{i_n}}^2 = \sum_n V_{g_n}, \tag{4}
 \end{aligned}$$

where V_{g_n} is the genic variance of the n th locus.

We minimize the weighted average of the squared differences between genic and genotypic values to determine the $\alpha_{n_{i_n}}$ from the least-square method.

Let

$$Q = \sum_{i1, \dots, in; j1, \dots, jn} \left(\prod_n p_{n_{i_n}} p_{n_{j_n}} \right) \left[a_{i1, j1, \dots, in, jn} - \left(\sum_n (\alpha_{n_{i_n}} + \alpha_{n_{j_n}}) \right) \right]^2.$$

Therefore

$$\frac{\partial Q}{\partial \alpha_{n_m}} = -2(1 + \delta_{m,jn}) \sum_{i_1, \dots, i_m; j_1, \dots, j_n} \left(\prod_n p_{n_m} p_{n_j} \right) \times$$

$$\left[a_{i_1 j_1, \dots, i_m j_m} - \sum_n (\alpha_{n_m} + \alpha_{n_j}) \right] = 0. \quad (5)$$

Hence

$$- \sum_{i_n} p_{n_m} a_{i_n} + \sum_{i_1, \dots, i_m} \left(\prod_n p_{n_m} \right) \left(\sum_n \alpha_{n_m} \right) \sum_{j_1, \dots, j_n} \prod_n p_{n_j}$$

$$+ \sum_{j_1, \dots, j_n} \left(\prod_n p_{n_j} \right) \left(\sum_n \alpha_{n_j} \right) \sum_{i_1, \dots, i_m} \prod_n p_{n_m} = 0, \quad (6)$$

where

$$a_{i_n} = \sum_{i_1, \dots, i_{(n-1)}; j_1, \dots, j_n} p_{i_1} p_{i_2} \dots p_{i_{n-1}} a_{i_1 j_1, \dots, i_n j_n}.$$

Therefore

$$- \sum_{i_n} p_{n_m} a_{i_n} + \sum_{i_1} p_{i_1} \alpha_{i_1} \left(\sum_{i_2, \dots, i_m} \prod_{n \neq 1} p_{n_m} \right) + \sum_{i_2} p_{i_2} \alpha_{i_2} \left(\sum_{i_1, i_3, \dots, i_m} \prod_{n \neq 2} p_{n_m} \right) + \dots$$

$$+ \sum_{j_1} p_{j_1} \alpha_{j_1} \left(\sum_{j_2, \dots, j_n} \prod_{n \neq 1} p_{n_j} \right) + \sum_{j_2} p_{j_2} \alpha_{j_2} \left(\sum_{j_1, j_3, \dots, j_n} \prod_{n \neq 2} p_{n_j} \right) + \dots = 0. \quad (7)$$

In the above equation each term other than the first one is separately zero because of equations (1) and (2). We keep only one term suitable to our purpose. Therefore

$$- \sum_{i_n} p_{n_m} (a_{i_n} - \alpha_{n_m}) = 0.$$

Hence

$$a_{i_n} = \alpha_{n_m}. \quad (8)$$

Similarly

$$a_{j_n} = \alpha_{n_j}.$$

These equations are true for any n .

$$\bar{m} = \sum_{i_1, \dots, i_m; j_1, \dots, j_n} \left(\prod_n p_{n_m} p_{n_j} \right) m_{i_1 j_1, \dots, i_m j_n}$$

or

$$\dot{\bar{m}} = 2 \sum_{i_1, \dots, i_m} \sum_n \dot{p}_{n_m} m_{i_n} = 2 \sum_{i_1, \dots, i_m} \sum_n p_{n_m} (m_{i_n} - \bar{m}) m_{i_n}$$

$$= 2 \left[\sum_{i_1, \dots, i_m} \sum_n p_{n_m} (m_{i_n} - \bar{m}) m_{i_n} - \sum_{i_1, \dots, i_m} \sum_n p_{n_m} \bar{m} (m_{i_n} - \bar{m}) \right] \quad (9)$$

as the second term is zero. Therefore

$$\begin{aligned}
 \dot{\bar{m}} &= 2 \sum_{i_1, \dots, i_n} \sum_n p_{n_{i_n}} (m_{i_n} - \bar{m})^2 \\
 &= 2 \sum_{i_1, \dots, i_n} \sum_n p_{n_{i_n}} \left(\sum_{j_n} p_{n_{j_n}} m_{i_n j_n} - \bar{m} \right)^2 \\
 &= 2 \sum_{i_1, \dots, i_n} \sum_n p_{n_{i_n}} \left[\sum_{j_n} p_{n_{j_n}} (\bar{m} + a_{i_n j_n}) - \bar{m} \right]^2 \\
 &= 2 \sum_{i_1, \dots, i_n} \sum_n p_{n_{i_n}} \left(\sum_{j_n} p_{n_{j_n}} a_{i_n j_n} \right)^2 = 2 \sum_{i_1, \dots, i_n} \sum_n p_{n_{i_n}} a_{i_n}^2 \\
 &= 2 \sum_{i_1, \dots, i_n} \sum_n p_{n_{i_n}} \alpha_{n_{i_n}}^2 = \sum_n V_{\xi_n} \tag{10}
 \end{aligned}$$

(we have used equations (4) and (8)).

This proves Fisher's Fundamental Theorem for a multilocus multiallele genetic system without linkage and epistasis.

3. Discussion

Price (1972) is probably correct in saying that Fisher viewed the partial changes in mean fitness as those due to single-locus gene-frequency changes. These changes formed the real substance of evolution according to Fisher [in contrast, for example, to the evolution of co-adapted gene complexes favoured by Wright (1988)]. This may be the central role of FFT. Nagylaki (1976) has shown that under most circumstances the terms additional to the genic variance are of a smaller order than the genic variance; the major exception occurs when the population is near an equilibrium. However, a quite general analysis, involving any number of loci, any number of alleles at each locus, and arbitrary mating schemes appears to be difficult. Nagylaki (1991) believed that the biological significance of the partial change in the mean fitness remains to be demonstrated. Ewens (1989) also recorded a negative assessment of the theorem as a biological statement. Even if FFT is inexact and incomplete, it captures the essence of the way selection works, and encapsulates a great deal of evolutionary insight in a simple expression (Crow 1990).

The compact and systematic method developed in section 2 can be used to simplify theoretical analyses for multiple loci (with multiple alleles at each locus), which otherwise would have been too cumbersome. This method has been successfully applied to prove Svirezhev's integral variational principle (Svirezhev 1972) for a multilocus multiallele genetic system (Behera, in press).

Acknowledgements

I gratefully acknowledge the valuable comments of an anonymous referee. I also acknowledge a research associateship from the Council of Scientific and Industrial Research, New Delhi.

References

- Behera N. 1995 Variational principles in evolution. *Bull. Math. Biol.* (in press)
- Crow J. F. and Kimura M. 1970 *An introduction to population genetic theory* (New York: Harper and Row) pp. 117–118
- Crow J. F. 1990 Fisher's contribution to genetics and evolution. *Theoret. Pop. Biol.* 38: 263–275
- Edwards A. W. F. 1994 The Fundamental Theorem of Natural Selection. *Biol. Rev. Cambridge Philos. Soc.* 69: 443–474.
- Ewens W. J. 1989 An interpretation and proof of the Fundamental Theorem of Natural Selection. *Theoret. Pop. Biol.* 36: 167–180
- Fisher R. A. 1930 *The genetical theory of natural selection* (Oxford: Clarendon Press)
- Kimura M. 1958 On the change of population fitness by natural selection. *Heredity* 12: 145–167
- Nagylaki T. 1976 The evolution of one- and two-locus systems. *Genetics* 83: 583–600
- Nagylaki T. 1991 Error bounds for the fundamental and secondary theorems of natural selection. *Proc. Natl. Acad. Sci. USA* 88: 2402–2406
- Price G. R. 1972 Fisher's 'Fundamental Theorem' made clear. *Ann. Hum. Genet.* 36: 129–140
- Svirezhev Y. M. 1972 Optimum principles in genetics. In *Studies on theoretical genetics* (ed.) V. A. Ratner (Novosibirsk: USSR Academy of Sciences) pp. 85–102
- Wright S. 1988 Surfaces of selective value revisited. *Am. Nat.* 131: 115–123