

The Fundamental Theorem of Natural Selection in Ewens' Sense (Case of Many Loci)

ANNE-MARIE CASTILLOUX AND SABIN LESSARD

*Département de mathématiques et de statistique, Université de Montréal,
C.P. 6128, Succursale Centre-ville, Montréal (Québec), Canada H3C 3J7*

Received August 24, 1994

We correct an equation characterizing the additive allelic effects on fitness determined at many loci without the assumption of random mating given in Ewens (1989, 1992) and we show that the additive genetic variance in fitness divided by the mean fitness is equal to a partial change in the mean fitness from one generation to the next as stated in Ewens (1989). © 1995 Academic Press, Inc.

I. INTRODUCTION

Fisher's (1930) Fundamental Theorem of Natural Selection (FTNS) states that "*the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time*" but "*the rigour of the demonstration requires that the terms employed should be used strictly as defined*" (Fisher, 1958, pp. 37–38).

Price (1972) points out that, if the mean fitness is defined in a way suggested in Fisher's works, the FTNS actually concerns a partial rate of change in the mean fitness. Ewens (1989) claims that the FTNS should not be limited to continuous-time models and interprets the FTNS in the case of discrete nonoverlapping generations as follows: *a partial change in the mean fitness of a population is exactly equal to the ratio of the additive genetic variance in fitness over the mean fitness*. The partial change in the mean fitness is obtained by replacing the exact genotypic values of fitness by the associated additive genotypic values and by keeping these values constant from one generation to the next. Some of the main arguments in support of Ewens' interpretation include the following: the FTNS should be exact and it should hold without any assumption of random mating.

A further contribution of Ewens (1989) is to show that the FTNS in its new interpretation holds in the case of fitness determined at many loci. But,

in Eq. (26) therein, which characterizes the additive (average) allelic effects on fitness in this case, a term was inadvertently omitted. This equation is corrected in a follow-up paper (Eq. (26) in Ewens (1992)). But a term is still missing in this corrected equation.

In this paper, we give the complete equation characterizing the additive allelic effects on fitness determined at many loci without the assumption of random mating. Then we confirm that the FTNS in Ewens' (1989) sense is fully valid in this general case. This is pertinent since "*the new interpretation leans heavily on the concept of the average effect of an allele*" (Ewens, 1992, p. 345).

2. MODEL AND NOTATION

Let $A_1^{(k)}, \dots, A_{s_k}^{(k)}$ be the s_k alleles of a gene at locus k for $k = 1, \dots, n$. The genotypes in this n locus system for a diploid population are made of two gametes, i and j , of the form

$$i = (i_1, \dots, i_n),$$

$$j = (j_1, \dots, j_n),$$

with $1 \leq i_k \leq s_k$ and $1 \leq j_k \leq s_k$, for $k = 1, \dots, n$. The parameter i_k represents the allele $A_{i_k}^{(k)}$ and the parameter j_k represents the allele $A_{j_k}^{(k)}$ for $k = 1, \dots, n$. At the beginning of the parental generation, the nonordered genotype $\{i, j\}$ has a frequency represented by $2g_{ij}$ if $i \neq j$ and by g_{ii} if $i = j$, in such a way that the ordered genotype (i, j) has a frequency g_{ij} for all i, j . We can agree that the genotype (i, j) is the result of a union of gamete i originating from the mother and of gamete j originating from the father. Moreover, we have the symmetry condition $g_{ij} = g_{ji}$ for all i, j in a population assumed without sex differences.

If p_{ki} designates the frequency of $A_{i_k}^{(k)}$ at locus k for i_k fixed, then we have

$$p_{ki} = \sum_j \sum_{i \ni i_k} g_{ij}, \tag{1}$$

where the second summation is taken over all gametes i containing the fixed i_k at the k th component, for $i_k = 1, \dots, s_k$ and $k = 1, \dots, n$.

Now let w_{ij} be the fitness of genotype (i, j) . By symmetry, we have $w_{ij} = w_{ji}$ for all i, j . This means that the fitness associated with a pair of gametes does not depend on the origin (maternal or paternal) of the gametes. However, the fitness parameters can depend on the composition

of the population. The mean fitness in the population is given in all cases by

$$\bar{w} = \sum_{i,j} w_{ij} g_{ij}.$$

3. CHANGE IN MEAN FITNESS

Let us consider the change in the mean fitness from one generation to the next. This change can be caused by a change in the frequency of genotype (i, j) , denoted Δg_{ij} , or by a change in the fitness of genotype (i, j) , denoted Δw_{ij} , or by a change in these two quantities simultaneously, and this for every genotype (i, j) of the population. Actually, we have

$$\Delta \bar{w} = \sum_{i,j} w_{ij} (\Delta g_{ij}) + \sum_{i,j} (\Delta w_{ij}) g_{ij} + \sum_{i,j} (\Delta w_{ij}) (\Delta g_{ij}). \quad (2)$$

We write the fitness w_{ij} in the form

$$w_{ij} = \bar{w} + \sum_k (\alpha_{kik} + \alpha_{kj_k}) + \gamma_{ij},$$

with the condition

$$\sum_{ik} \alpha_{kik} p_{kik} = 0, \quad \text{for } k = 1, \dots, n, \quad (3)$$

by using the least-squares method (see Section 5). The parameter α_{kik} represents the additive effect of allele $A_{ik}^{(k)}$, while γ_{ij} measures the interactions between the alleles carried by i and j . The first term on the right-hand side of (2) becomes

$$\sum_{i,j} \left[\bar{w} + \sum_k (\alpha_{kik} + \alpha_{kj_k}) \right] \Delta g_{ij} + \sum_{i,j} \gamma_{ij} \Delta g_{ij}.$$

Let $w_{ij}^{(A)}$ be the additive genotypic value, that is,

$$w_{ij}^{(A)} = \bar{w} + \sum_k (\alpha_{kik} + \alpha_{kj_k}).$$

Following Price (1972) and Ewens (1989), we consider the change in the mean fitness only through the changes in the genotypic frequencies and by replacing the genotypic values by the corresponding additive genotypic

values. Therefore, we consider the following partial change in the mean fitness:

$$\Delta_{\text{part}} \bar{w} = \sum_{i,j} \left[\bar{w} + \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right] \Delta g_{ij} = \sum_{i,j} w_{ij}^{(A)} \Delta g_{ij}. \tag{4}$$

We show below that this partial change in the mean fitness of a population is σ_A^2/\bar{w} , where σ_A^2 is the additive genetic variance in fitness.

4. CHANGES IN ALLELIC FREQUENCIES

The change in the frequency of genotype (i, j) between the time of conception and the time of maturity is

$$\tilde{g}_{ij} - g_{ij} = \frac{g_{ij}(w_{ij} - \bar{w})}{\bar{w}},$$

where \tilde{g}_{ij} represents the frequency of genotype (i, j) at maturity, for every i and j . Then the change in the frequency of allele $A_{ik}^{(k)}$ between the time of conception and the time of maturity is

$$\tilde{p}_{ki_k} - p_{ki_k} = \sum_j \sum_{i \ni i_k} (\tilde{g}_{ij} - g_{ij}) = \sum_j \sum_{i \ni i_k} \frac{g_{ij}(w_{ij} - \bar{w})}{\bar{w}}, \tag{5}$$

for $i_k = 1, \dots, s_k$ and $k = 1, \dots, n$. Under most mating patterns, \tilde{p}_{ki_k} is also the frequency of $A_{ik}^{(k)}$ at the beginning of the next generation, denoted p'_{ki_k} (in the n locus case as well as in the one locus case, since the gene frequencies are not directly affected by recombination). Throughout we will make the assumption

$$\Delta p_{ki_k} = p'_{ki_k} - p_{ki_k} = \tilde{p}_{ki_k} - p_{ki_k}, \tag{6}$$

for $i_k = 1, \dots, s_k$ and $k = 1, \dots, n$.

5. ADDITIVE ALLELIC EFFECTS

The additive allelic effects are obtained by the least-squares method, under condition (3). Therefore, let us define

$$S = \sum_{i,j} g_{ij} \left[w_{ij} - \bar{w} - \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right]^2 + \sum_k \lambda_k \sum_{i_k} \alpha_{ki_k} p_{ki_k}, \tag{7}$$

whose partial derivatives are given by

$$\frac{\partial S}{\partial \alpha_{mi_m}} = -4 \sum_j \sum_{i \ni i_m} g_{ij} \left[w_{ij} - \bar{w} - \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right] + \lambda_m p_{mi_m},$$

for $i_m = 1, \dots, s_m$ and $m = 1, \dots, n$. By equating the partial derivatives to 0 and by summing on i_m for each m , we find

$$\begin{aligned} \lambda_m &= 4 \sum_{i,j} g_{ij} \left[w_{ij} - \bar{w} - \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right] \\ &= -8 \sum_k \sum_{i_k} \alpha_{ki_k} p_{ki_k} \\ &= 0, \end{aligned}$$

for $m = 1, \dots, n$. Now according to (5) and (6), we have

$$\sum_j \sum_{i \ni i_m} g_{ij} \left[w_{ij} - \bar{w} - \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right] = \bar{w}(\Delta p_{mi_m}) - \sum_j \sum_{i \ni i_m} g_{ij} \sum_k (\alpha_{ki_k} + \alpha_{kj_k}).$$

Then, the equality

$$\frac{\partial S}{\partial \alpha_{mi_m}} = 0,$$

for $i_m = 1, \dots, s_m$ and $m = 1, \dots, n$ is equivalent to

$$\bar{w}(\Delta p_{mi_m}) = \sum_j \sum_{i \ni i_m} g_{ij} \sum_k (\alpha_{ki_k} + \alpha_{kj_k}), \quad (8)$$

for $i_m = 1, \dots, s_m$ and $m = 1, \dots, n$. Moreover, we have

$$\begin{aligned} &\sum_j \sum_{i \ni i_m} g_{ij} \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \\ &= \alpha_{mi_m} p_{mi_m} + \sum_{k \neq m} \sum_{i_k} \alpha_{ki_k} P_{mi_m, ki_k} + \sum_k \sum_{j_k} \alpha_{kj_k} Q_{mi_m, kj_k}, \end{aligned}$$

where

$$P_{mi_m, ki_k} = \sum_j \sum_{i \ni i_k, i_m} g_{ij}$$

and

$$Q_{mi_m, kj_k} = \sum_{j \ni j_k} \sum_{i \ni i_m} g_{ij}.$$

Therefore, the additive allelic effects are implicitly defined by the relation

$$\bar{w}(\Delta p_{mi_m}) = \alpha_{mi_m} p_{mi_m} + \sum_{k \neq m} \sum_{i_k} \alpha_{ki_k} P_{mi_m, ki_k} + \sum_k \sum_{j_k} \alpha_{kj_k} Q_{mi_m, kj_k}, \quad (9)$$

for $m = 1, \dots, n$, where the quantity P_{mi_m, ki_k} is the frequency of all gametes i that carry $A_{i_m}^{(m)}$ and $A_{i_k}^{(k)}$ at loci m and k , respectively, while Q_{mi_m, kj_k} represents the frequency of all genotypes that carry $A_{i_m}^{(m)}$ and $A_{j_k}^{(k)}$ at loci m and k , respectively, on different gametes, i and j , respectively. Note that P_{mi_m, ki_k} concerns pairs of genes on the same gamete while Q_{mi_m, kj_k} concerns pairs of genes on two united gametes.

6. DECOMPOSITION OF GENETIC VARIANCE

Let σ_T^2 be the total genetic variance of fitness. Therefore, we have

$$\begin{aligned} \sigma_T^2 &= \sum_{i,j} g_{ij} (w_{ij} - \bar{w})^2 \\ &= \sum_{i,j} g_{ij} \left[w_{ij} - \bar{w} - \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right]^2 + \sum_{i,j} g_{ij} \left[\sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right]^2 \\ &\quad + 2 \sum_{i,j} g_{ij} \left[w_{ij} - \bar{w} - \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right] \left[\sum_l (\alpha_{li_l} + \alpha_{lj_l}) \right]. \end{aligned}$$

But, by using (5), (6), and (8), we find

$$\begin{aligned} &\sum_{i,j} g_{ij} \left[w_{ij} - \bar{w} - \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right] \left[\sum_l (\alpha_{li_l} + \alpha_{lj_l}) \right] \\ &= 2 \sum_l \sum_{i_l} \alpha_{li_l} \sum_j \sum_{i \ni i_l} g_{ij} (w_{ij} - \bar{w}) - 2 \sum_k \sum_{i_k} \alpha_{ki_k} \sum_j \sum_{i \ni i_k} g_{ij} \sum_l (\alpha_{li_l} + \alpha_{lj_l}) \\ &= 2\bar{w} \sum_l \sum_{i_l} \alpha_{li_l} (\Delta p_{li_l}) - 2\bar{w} \sum_k \sum_{i_k} \alpha_{ki_k} (\Delta p_{ki_k}) \\ &= 0. \end{aligned}$$

Then, we have

$$\sigma_T^2 = \sigma_R^2 + \sigma_A^2, \quad (10)$$

where

$$\sigma_R^2 = \sum_{i,j} g_{ij} \left[w_{ij} - \bar{w} - \sum_k (\alpha_{ki_k} + \alpha_{kj_k}) \right]^2$$

and

$$\sigma_A^2 = \sum_{i,j} g_{ij} \left[\sum_k (\alpha_{kik} + \alpha_{kj_k}) \right]^2.$$

The quantity σ_R^2 is the variance of the interaction effects, also called the residual variance, and σ_A^2 is the variance of the additive allelic effects, called the additive genetic variance. Moreover, by using (8), we have

$$\begin{aligned} \sigma_A^2 &= \sum_{i,j} g_{ij} \left[\sum_k (\alpha_{kik} + \alpha_{kj_k}) \right]^2 \\ &= 2 \sum_k \sum_{i_k} \alpha_{kik} \sum_j \sum_{i \ni i_k} g_{ij} (\alpha_{li_i} + \alpha_{lj_i}) \\ &= 2\bar{w} \sum_k \sum_{i_k} \alpha_{kik} (\Delta p_{kik}). \end{aligned} \tag{11}$$

7. PARTIAL CHANGE IN MEAN FITNESS

Finally, if Δg_{ij} is the change in the frequency of genotype (i, j) between the mother generation and the daughter generation both at conception, then the partial change in the mean fitness between these two generations as defined previously is

$$\begin{aligned} \Delta_{\text{part}} \bar{w} &= \sum_{i,j} \Delta g_{ij} \left[\bar{w} + \sum_k (\alpha_{kik} + \alpha_{kj_k}) \right] \\ &= 2 \sum_k \sum_{i_k} \alpha_{kik} \sum_j \sum_{i \ni i_k} \Delta g_{ij} \\ &= 2 \sum_k \sum_{i_k} \alpha_{kik} (\Delta p_{kik}), \end{aligned} \tag{12}$$

where Δp_{kik} is the change in the frequency of allele $A_{i_k}^{(k)}$ between the mother generation and the daughter generation both at conception. The above expressions for σ_A^2 and $\Delta_{\text{part}} \bar{w}$ allow us to conclude that

$$\Delta_{\text{part}} \bar{w} = \frac{\sigma_A^2}{\bar{w}}. \tag{13}$$

Therefore, we get the result stated in Ewens (1989) in the case of many loci without any assumption of random mating.

8. DISCUSSION

Equation (9) characterizes the additive allelic effects on fitness determined at many loci without any assumption of random mating. If we write it in the form

$$\begin{aligned} \bar{w}(\Delta p_{mi_m}) = & \alpha_{mi_m} p_{mi_m} + \sum_{k \neq m} \sum_{i_k} \alpha_{ki_k} P_{mi_m, ki_k} + \sum_{j_m} \alpha_{mj_m} Q_{mi_m, mj_m} \\ & + \sum_{k \neq m} \sum_{j_k} \alpha_{kj_k} Q_{mi_m, kj_k}, \end{aligned} \tag{14}$$

for $m = 1, \dots, n$, then it is the last two terms which are missing in Eq. (26) of Ewens (1989), while only the last term is missing in the corrected Eq. (26) of Ewens (1992). Note that, in the case of random mating, which is equivalent to random union of gametes, we have

$$Q_{mi_m, kj_k} = p_{mi_m} p_{kj_k}$$

for $i_m = 1, \dots, s_m$ and $j_k = 1, \dots, s_k$ for $m, k = 1, \dots, n$, and then

$$\sum_k \sum_{j_k} \alpha_{kj_k} Q_{mi_m, kj_k} = p_{mi_m} \sum_k \sum_{j_k} \alpha_{kj_k} p_{kj_k} = 0.$$

In this case, Eq. (9) becomes

$$\bar{w}(\Delta p_{mi_m}) = \alpha_{mi_m} p_{mi_m} + \sum_{k \neq m} \sum_{i_k} \alpha_{ki_k} P_{mi_m, ki_k} \tag{15}$$

for $m = 1, \dots, n$. This happens to be the incomplete Eq. (26) in Ewens (1989) with reference to Ewens (1979, p. 218). But Ewens was not assuming random mating and the correspondence of this incomplete equation to the equation in the random mating case must be attributed to typographic errors or to algebraic errors or to a combination of both.

Note that Eq. (9) is equivalent to Eq. (8), from which it has been deduced, but Eq. (8) leads more directly to the expression (11) for the additive genetic variance σ_A^2 which is used to get Ewens' result (13). Unfortunately, Eq. (8) or a corresponding equation is not given in Ewens (1989). But, as we have shown, the main conclusion of this paper remains fully valid.

Equation (14) can be arranged in a different way. Combining the terms associated to genes at the same pairs of distinct loci on a same gamete and on two united gametes leads to a corrected matrix \mathbf{Q} of pairwise two-locus gametic frequencies given in Ewens (1992), whose (mi_m, ki_k) element for any i_m, i_k and any $k \neq m$ represents the frequency of all gametes carrying $A_{i_m}^{(m)}$ at locus m and $A_{i_k}^{(k)}$ at locus k , or would do so if recombination

occurred between m and k loci in every genotype (Ewens, personal communication). Formally, this element is

$$P_{mi_m, ki_k} + Q_{mi_m, ki_k}.$$

We propose the following interpretation, which makes more precise the previous one: this element is twice the probability that two genes chosen at random and independently, one at locus m and one at locus k , in a same individual chosen at random, be $A_{i_m}^{(m)}$ at locus m and $A_{i_k}^{(k)}$ at locus k . This is so, since the two genes chosen at random will be on the same gamete with probability $\frac{1}{2}$ and on different united gametes with probability $\frac{1}{2}$. In the first event, the probability is P_{mi_m, ki_k} and, in the second, it is Q_{mi_m, ki_k} . It is interesting to note that the actual recombination rate between the two loci does not play any role in the definition of the matrix of pairwise two-locus gametic frequencies.

Moreover, the remaining terms in (14) concern genes at the same locus m for $m = 1, \dots, n$ and yield a block diagonal matrix, represented by $\mathbf{D} + \mathbf{P}$ in Ewens (1992), whose (mi_m, mj_m) element for any i_m, j_m and any m is

$$\delta_{i_m, j_m} P_{mi_m} + Q_{mi_m, mj_m},$$

where $\delta_{i_m, j_m} = 1$ if $j_m = i_m$ and 0 otherwise. We notice that this element is twice the probability that two genes chosen at random and independently, one after the other at a same locus m , in a same individual chosen at random, be $A_{i_m}^{(m)}$ and $A_{j_m}^{(m)}$ in this order. This interpretation is analogous to the previous one.

To sum up, the allelic effects are characterized by the equation

$$\bar{w}(\Delta p_{mi_m}) = \sum_k \sum_{i_k} \alpha_{ki_k} V_{mi_m, ki_k},$$

for $m = 1, \dots, n$, where V_{mi_m, ki_k} for any i_m, i_k and any m, k is twice the probability that two genes chosen at random and independently, the first one at locus m and the second one at locus k , in a same individual chosen at random, be $A_{i_m}^{(m)}$ and $A_{i_k}^{(k)}$ in this order. In matrix notation, we have

$$\bar{w}(\Delta p) = \mathbf{V}\alpha.$$

The matrix \mathbf{V} coincides to the corrected matrix \mathbf{Q} off the block diagonal, where the elements are associated to pairs of genes at different loci, while the block diagonal elements, which are those of the matrix $\mathbf{D} + \mathbf{P}$ associated to pairs of genes at a same locus, share the same interpretation as the others.

ACKNOWLEDGMENTS

The authors are most grateful to Professor Warren J. Ewens for having called their attention to his 1992 paper and for several other helpful comments on a first draft of the present paper which are reported in part in the discussion section. This research was supported in part by a CRSNG grant from the Government of Canada and a FCAR grant from the Government of Quebec.

REFERENCES

- EWENS, W. J. 1979. "Mathematical Population Genetics," Springer-Verlag, Berlin/New York.
- EWENS, W. J. 1989. An interpretation and proof of the fundamental theorem of natural selection, *Theor. Popul. Biol.* **36**, 167-180.
- EWENS, W. J. 1992. An optimizing principle of natural selection in evolutionary population genetics, *Theor. Popul. Biol.* **42**, 333-346.
- FISHER, R. A. 1930. "The Genetical Theory of Natural Selection," Clarendon Press, Oxford; 2nd edition, 1958, Dover, New York.
- PRICE, G. R. 1972. Fisher's "fundamental theorem" made clear, *Am. Hum. Genet.* **36**, 129-140.