Gametic selection and the selection component analysis

By JOSEPH H. NADEAU,^{1,3} KLAUS DIETZ² AND ROBERT H. TAMARIN¹

¹ Biology Department, Boston University, Boston MA 02215, ² Institut für Medizinische Biometrie, Hallstattstraße 6, Eberhard-Karls-Universität, 7400 Tübingen, FRG, ³ Present address: Jackson Laboratory, Bar Harbor, ME 04609

(Received 28 January 1980 and in revised form 5 November 1980)

SUMMARY

The selection component analysis developed by Christiansen & Frydenberg (1973, 1976) is a method to study the components of natural selection and is based on an analysis of population samples which include mothers and their progeny (mother-offspring combinations). Because only one progeny per mother-offspring combination is analysed, gametic selection and the reproductive components of selection in males are poorly characterized. We investigated the information which can be gained by analysing all progeny in each combination and showed that analysis of entire litters enables detection of gametic selection. Testing and estimation procedures are derived for this purpose. Sexual selection in males and mate preferences remain poorly characterized because the only information about the reproductive components in males is contained in the sample of male gametes and is insufficient to reconstruct the breeding structure of the male population. The format and interpretation of a selection component analysis is presented which takes these results into account.

1. INTRODUCTION

The direct measurement of selection in natural population could resolve the controversy regarding both the selective value of genetic variation and the patterns of selection characterizing hatural populations. The absence of accurate, locus-specific estimates reflects the complexities of measuring the modes, magnitudes, and units of natural selection (Lewontin, 1974). Accurate estimates require measurement of all fitness components and the age-specific fertility and mortality schedules. One of the most promising methods to resolve some of these problems is the selection component analysis (SCA).

The SCA developed by Christiansen & Frydenberg (1973, 1976) is based on population samples which include mothers and their progeny (mother-offspring combinations) and has been used to study polymorphisms in several species (Christiansen, Frydenberg & Simonsen, 1973; Christiansen, Frydenberg, Glydenholm & Simonsen, 1974; Baccus *et al.* 1977; Baccus, Joule & Kimberling, 1980). Evidence for selection was observed in each study, but because only one offspring per mother-offspring combination was analysed, tests of gametic and sexual selection in males and mate preferences were confounded (Christiansen & Frydenberg, 1973). In this paper, we study the information which can be gained by including entire litters in the analysis.

2. ANALYSIS

(i) Format. Table 1 defines the notation used in most of the analysis. The notation and the relationships between the observed and expected frequencies are retained from the Christiansen-Frydenberg SCA (1973) to facilitate the incorpora-

Table 1. Definition of terms

(The notation of Christiansen & Frydenberg (1973) is retained. i, j = 1, 2, 3 for A_1A_1, A_1A_2, A_2A_2 and *i* denotes adults and *j* denotes zygotes. When reference is to allelic frequency, 1 denotes the allele shared by genotypes 1 and 2, and 2 the allele shared by genotypes 2 and 3.)

$\mathbf{Expected}$	Observed	Definition
Ø,	F_{i}	Mothers
σ_i	S_i	Non-pregnant females
α_{si}	Aei	Adults. $s = \delta$ or φ and is deleted if irrelevant
β_{ii}	B _{si}	Breeding adults
E sj	Z_{sj}	Zygotes
ρ	p	The frequency of allele 1 among gametes of hetero- zygotes. p' and p'' denote the degree of distortion in males and females, respectively
	k	k = 1,, 7 for litters with the following ordered geno- typic compositions, $1 - (00+)$, $2 - (0+0)$, $3 - (0++)$, 4 - (+00), $5 - (++0)$, $6 - (+0+)$, and $7 - (+++)$, where the presence or absence of progeny from a litter is denoted by + or 0, respectively, and where the first, second, and third positions are reserved for progeny of genotypes A_1A_1, A_2A_2, A_3A_3 , respectively
Yik	C _{ik}	Mother-offspring combination. For example, C_{12} denotes a combination where the mother is A_1A_1 and all progeny are of genotype A_1A_2 . This term is analogous to the Christiansen-Frydenberg term C_{ij}
E _{j(Y(k)}	$Z_{j(c_{ik})}$	Progeny of genotype j in mother-offspring combination C_{ik} . For example, $Z_{1(e_{ik})} = 6$ indicates that $6 A_1 A_1$ progeny were observed in litter(s) composed of genotypes $A_1 A_1$ and $A_1 A_2$ born by mother(s) of genotype $A_1 A_1$

tion of our results into their model. The principal additions are litter size parameters and the notation to index unique mother-offspring combinations. The following equalities are also included:

$$\sum_{j(\gamma_{ik})} \epsilon_{sj(\gamma_{ik})} = \epsilon_{s0}; \qquad (1)$$

$$\sum_{c_{ik}} Z_{sj(c_{ik})} = Z_{sj}; \qquad (2)$$

$$\sum_{j} Z_{sj} = Z_{s0}.$$
 (3)

and

The maximum-likelihood (MLE) of the zygotic frequency parameter is given by

$$\hat{\epsilon}_{j(\gamma_{ik})} = Z_{j(c_{ik})}/Z_0. \tag{4}$$

Table 2 defines the format of the SCA.

The selection components are defined as follows:

(a) gametic selection – the distortion of segregation in heterozygotes so that one allele is preferentially represented among progeny at conception;

(b) sexual selection – breeding individuals are a non-random sample of adults with respect to genotype;

(c) fecundity selection - litter size is dependent on parental genotype;

(d) zygotic selection - survival is dependent on the genotype of the zygote.

The principal conditions of the analysis are that it applies to diploid, dioecious organisms and that no selection has occurred between conception and progeny sampling.

(ii) Gametic selection in males. An unambiguous method to detect gametic selection in males exists if we confine our analysis to those mother-offspring combinations where paternal genotype is known with certainty. This is the case when the mother is homozygous and two genotypic classes occur in the litter. For example, AA and AB progeny by an AA mother must have been sired by an AB male. While we assume here that there is only one sire per mother-offspring combination, the tests are also applicable to situations where multiple insemination occurs. The occurrence of multiple insemination affects the interpretation but not the validity of the proposed tests (see below). If there is no gametic selection, both progeny classes should occur in equal frequency. The null hypothesis of no gemetic selection in males, p' = 0.5, is given by

$$\epsilon_{1(\gamma_{15})} + \epsilon_{2(\gamma_{33})} = \epsilon_{2(\gamma_{15})} + \epsilon_{3(\gamma_{33})}.$$
⁽⁵⁾

The requirement that only those litters are considered where $n > t \ge 1$ means that the MLE of the degree of segregation distortion in males, p', must be based on a doubly truncated binomial distribution. The MLE for single litters is given by the maximum of

$$L(p') = \frac{\binom{n}{t} (p')^{n-t} (1-p')^t}{1-(p')^n - (1-p')^n},$$
(6)

where t = 1, ..., n-1 (Johnson & Kotz, 1969). Assuming a constant probability of sampling paternal allele 1 among all litters, the MLE of p' for more than one litter is given by the maximum of

$$L(p') = \prod_{d=1}^{D} \frac{\binom{n_d}{t_d} (p')^{n_d - t_d} (1 - p')^{n_d}}{1 - (p')^{n_d} - (1 - p')^{n_d}},$$
(7)

where d refers to the dth of D litters, d = 1, ..., D. Programs for the calculation of the MLE and variance are available on request.

In addition there exists a maximum degree of segregation distortion which may be detected when paternal genotype is unspecified. Because both paternal alleles

	<u> </u>	males	A_1	• .	A_2		•					A_3	•	
	Non-	females	S_1	• •	S.	•	•	•			•	S	•	
ent analysis	y	A_2A_2	0	00	0	01	Z ^{3(c₂₁₎}	0 1	Z ³ (c ₂₃)	Z ^{3 (c} 27)	Z-3 (c26)	0	Z _{3 (cgs)}	$Z_{3}(c_{31})$
format for the selection compone	Number of progen	A_1A_2	0	$Z_{2}^{2(e_{15})}$	(31 ₂)2 _ 0	$Z_{2}^{(c_{33})}$	0	Z ₂ (c ₂₅)	Z ² (c23)	$Z_{2(c_{27})}$	0	$Z_{2^{\{b_{32}\}}}$	$\overset{Z_{2}(c_{33})}{\overset{\circ}{\phantom}}$	Ð
		A_1A_1	$Z_{1(e_{14})}$	$Z_{1(e_{1b})}$	$Z_{1(e_{n_i})}$	0	0	Z _{1(c25)}	0 1	Z1 (c27)	$Z_{1}(c_{26})$	0	0	0
Table 2. Data	Mother-	combination	C 14	2°0	0 ¹⁸	\widetilde{C}_{22}	c ²¹	C ^{se}	s S C	22 2	C.26	\tilde{C}_{32}	en e	C ₃₁
	Progeny	$A_1A_1A_1A_1A_2A_2A_3$	+	+ + +	 . +	+	+ •	• • + •	+ ·	∔ · + + ·	+ +	+	+ ·	+
		Mother	F_1		F_2							F_3		

J. H. NADEAU AND OTHERS

must occur at least once in each litter for the litter to be included in the analysis of gametic selection, the maximum degree of distortion is given by the maximum of

$$L(p'_{\max}) = \frac{n(p')^{n-1}(1-p')}{1-(p')^n - (1-p')^n}$$
(8)

and for all litters by the maximum of

$$L(p'_{\max}) = \prod_{d=1}^{D} \frac{n_d(p')^{n_d-1} (1-p')}{1-(p')^{n_d} - (1-p')^{n_d}}.$$
(9)

(iii) Gametic selection in females. The test given by Christiansen & Frydenberg (1973) is based on the null hypothesis that one half the progeny of heterozygous females are heterozygotes. However, the null hypothesis depends on both ρ' and β_{3i} and will not be rejected unless the frequency of male gamete *1* transmitted through heterozygous females is not equal to 0.5; that is, $(\beta_{31} + \rho' \beta_{32}) = 0.5$, regardless of p' (Christiansen & Frydenberg, 1973). An alternative test is based on those mother-offspring combinations where both parents are heterozygous, C_{26} and C_{27} . The null hypothesis is given by two independently tested components:

$$\epsilon_{1(\gamma_{25})} + \epsilon_{1(\gamma_{27})} = \frac{p'}{2} \left[\Sigma \epsilon_{j(\gamma_{25})} + \Sigma \epsilon_{j(\gamma_{27})} \right]$$
(10*a*)

and

$$\epsilon_{3(\gamma_{26})} + \epsilon_{3(\gamma_{27})} = \frac{1-p'}{2} \left[\Sigma \epsilon_{j(\gamma_{26})} + \Sigma \epsilon_{j(\gamma_{27})} \right]. \tag{10b}$$

The MLE of p'' must be based on a distribution which is truncated by both $N > n_1 \ge 1$ and $N > n_3 \ge 1$, where n_i are the number of progeny of genotype *i* produced by matings between heterozygous parents, N is the total number of progeny produced by these matings, and $N = \sum_i n_i$. Assuming a doubly truncated trinomial distribution, the MLE of p'' for single litters is given by the maximum of

$$L(p'') = \frac{\binom{N}{n_i} \left(\left[p'p'' \right]^{n_1} \left[p'(1-p'') + (1-p') p'' \right]^{n_2} \left[(1-p') (1-p'') \right]^{n_3} \right]}{1 + (p'+p''-2p'p'')^N - (p'+p''-p'p'')^N - (1-p'p'')^N}.$$
 (11)

Assuming that p'' is constant among heterozygous females, the MLE of p'' for more than one litter is given by the maximum of

$$L(p'') = \Pi \frac{\binom{N_{id}}{n_{id}} ([p'p'']^{n_{1d}} [p'(1-p'') + (1-p') p'']^{n_{2d}} [(1-p') (1-p'')]^{n_{3d}}}{1 + (p'+p''-2p'p'')^{N_d} - (p'+p''-p'p'')^{N_d} - (1-p'p'')^{N_d}}, \quad (12)$$

where d refers to the dth of D litters, d = 1, ..., D. While our test requires larger sample sizes to have the same power as the corresponding Christiansen-Frydenberg test (see below), our test is independent of $\beta_{\delta i}$ and should be able to detect a broader range of gametic selection.

The MLE of p'' assumes the p' is known with certainty. In those cases where both p' and p'' must be estimated from the same data, i.e. samples of natural populations, p' and p'' must be jointly estimated by the MLE of the product of equations (5) and (11) for single litters and by the product of equations (6) and (12) for more than one litter.

(iv) Power of the tests. For this purpose, the 50% power point is used and is defined as the degree of distortion from 0.5 at which the probability is 50% of rejecting the null hypothesis when it is false and the alternative hypothesis is correct. This point for the 5% significance level for the male gametic selection component is

$$3 \cdot 84/4 (\Sigma Z_{j(c_{1s})} + \Sigma Z_{j(c_{2s})})^{\frac{1}{2}}.$$
(13)

Approximate power tests can also be derived for the female gametic selection component.

Analysis of the power of the test to detect distortion in males demonstrated that detection of selection is less likely when based on the doubly truncated distribution (T) than when based on the non-truncated distribution (N) (Table 3). The loss of litters composed of only one genotypic class was usually largest when the degree of distortion was large and was primarily responsible for the lack of power. In three lines, lines 2, 3 and 12, the hypothesis of no distortion was not rejected when it was very likely false. Thus while the proposed test permits gametic selection to be tested independently of sexual selection in males, the test requires large sample sizes.

(v) Paternity assignment. If paternal genotype could be assigned to each mother-offspring combination, then the genotypic frequencies among breeding males could be estimated and the reproductive components of selection in males and mating preferences could be studied. Population samples which include mother-offspring combinations can be used to assign paternal genotype, but the assignments preclude testing of sexual selection in males and mating patterns.

While paternal genotype is sometimes exactly defined by each mother-offspring combination, many cases involve probability estimates. For example, an AAfemale that produces only AA progeny could have mated either an AA male or an AB heterozygous male whose B allele was not sampled. The probability estimates depend on the gene frequencies among breeding males (β_{3i}/β_{30}) , litter size (ϵ_{s0}/γ_i) , the degree of segregation distortion in males (ρ') and females (ρ'') , and mate preferences (ω_i) . In monogamous mating systems, the probability that a heterozygous male is the father of n homozygous progeny is given by

$$p(B_2|D) = ([p(D|B_2)]^n p(B_2|B_0)) / ([p(D|B_2)]^n p(B_2/B_0) + p(B_i|B_0)), \quad (14)$$

where, depending on the mating combinations, $p(D|B_2)$ is equal to p' or (1-p')and i = 1 or 3. Non-random mating can be taken into account by weighing the frequencies of breeding males by the factor (ω_i) which reflects mate preferences. The specific probability for each combination is given in Table 4.

The problem with using the assigned paternal genotypes to study selection is that to make the assignments, assumptions must be made concerning the frequencies of breeding males and the randomness of mating. Subsequent tests of sexual selection in males and random mating would be tautologous. Thus unless

1

using the non-truncated	
ble 3. Comparison of the degree of segregation distortion based on estimates v	(N) and doubly truncated (T) binomial distribution
T_{a}	

unpublished observations). Both alleles of the heterozygous parent must have occurred in each litter for the litter to be included in the analysis. P < 0.05.1

	12		200	138		0.42	0-47		0.03	0.05		5.12*	0.46		0.07	0.08
	11		196	191		0.52	0.52		0.03	0.04		0.18	0.13		0.07	0.07
	10		110	78		0.44	0.45		0.04	0.06		1.78	0.82		0.09	0.11
	6		125	95		0.54	0.53		0.04	0-06		0-97	0.26		0-0	0.10
	8		130	109		0.29	0.31		0.04	0.05		24.12*	11.24*		60·0	60.0
	2		135	109		0.53	0.54		0.04	0.05		0.36	0.45		0.08	60.0
	9		181	122		0-44	0.42		0.03	0.05		2.44	2.10		0.07	60.0
	ũ		52	45		0.50	0.57		0-02	0-08		0	0.56		0.14	0.15
	4		62	54		0.48	0.52		0.06	0-07		0-02	0-02		0.12	0.13
	er		86	47		0.23	0.39		0.04	60·0		24.61*	1.04		0.11	0.14
< 0.05.)	67		107	83		0.40	0.48		0.04	0-06		4·12*	0-11		60·0	0.11
Iysis. P	1		34	29		0.59	0.61		0.10	0.01		1.06	0.86		0.17	0.18
ed in the ana	Line	Sample size	z	H	'n,	N	т	s.D.	z	Т	χ²	N	Т	Power	N	Т
ğ																

Selection component analysis

independent evidence can be used to identify breeding males, sexual selection in males and mating patterns can only be studied by the methods of Christiansen & Frydenberg (1973). As emphasized by Christiansen & Frydenberg, these tests detect only certain patterns of selection.

Table 4. Conditional probabilities for paternal genotype associated with each motheroffspring combination, given the frequencies of breeding males and the degrees of segregation distortion in males and females

]	Litte	r								
	composition			Probability that a heterozygous male is the father of n observed							
Maternal				progeny in each mother-offspring combination, or the proportion							
genotype	Z_1	Z_2	Z_3	of mother-offspring combinations associated with heterozygous males							
M_1	+			$(p')^n \beta_2 / [(p')^n \beta_2 + \beta_1]$							
	+	+		1.0							
	•	+	•	$(1-p')^n \beta_2 / [(1-p')^n \beta_2 + \beta_3]$							
M_2	+	•		$(p')^n \beta_2 / [(p')^n \beta_2 + \beta_1]$							
	•	+	•	Uninformative							
			+	$(1-p')^n \beta_2/[(1-p')^n \beta_2 + \beta_3]$							
	+	+		$(p'p'' + p'(1-p'') + (1-p')p'')^n \beta_2 / [(p'p'' + p'(1-p'') + (1-p'))]$							
				$(p'')^n\beta_2+\beta_1]$							
	٠	+	+	$[p'(1-p'')+(1-p')p''+(1-p')(1-p'')]^n\beta_2/[p'(1-p'')+$							
				$(1-p^{r}) p^{r} + (1-p^{r}) (1-p^{r})^{n} p_{2} + p_{3}$							
	+	+	+	1:0							
	+	•	+	1.0							
M_3		+		$(p')^n \beta_2 / [(p')^n \beta_2 + \beta_1]$							
		+	+	1.0							
	•	•	+	$(1-p')^n \beta_2/[(1-p')^n \beta_2 + \beta_3]$							

3. DISCUSSION

With the new tests for gametic selection, several changes are recommended in the SCA. The following is the revised sequence in which the data should be analysed:

(a) H_1 : gametic selection in males can be tested by equation (5). Rejection of the null hypothesis may indicate segregation distortion, multiple insemination and non-random mating, or both (see below).

(b) H_2 : random mating can be tested by comparing the frequency with which paternal allele 1 is transmitted through females of each genotype;

$$(\beta_1 + p'\beta_2)_1 = (\beta_1 + p'\beta_2)_2 = (\beta_1 + p'\beta_2)_3, \tag{15}$$

where the subscript outside the parentheses denotes the genotype of the mother through which paternal allele I was transmitted.

(c) H_3 : sexual selection in males can be tested by comparing the frequency of allele 1 in breeding males and the frequency in progeny:

$$(\beta_1 + p'\beta_2) = (\epsilon_1 + 0.5\epsilon_2). \tag{16}$$

(d) H_4 : gametic selection in females can be tested by equation (10).

(e) H_5 : sexual selection in females can be tested by comparing the genotypic frequencies in breeding and non-pregnant females:

$$\phi_i = \sigma_i = \alpha_{\mathfrak{Q}_i}. \tag{17}$$

(f) H_6 : fecundity selection can be tested by comparing the total number of progeny produced by females of each genotype with that predicted by the product of average litter size and the total number of mothers of each genotype:

$$\sum_{sj} \epsilon_{sj(\gamma_{ik})} = \left(\sum_{sj} \epsilon_{sj} / \sum_{i} \phi_{i} \right) \phi_{i}.$$
(18)

Rejection of the null hypothesis may indicate that litter size is dependent on maternal genotype.

(g) H_7 : zygotic selection should ideally be studied by a temporal study of the survival of zygotes within a cohort. However, if genetic equilibrium is assumed, then zygotic selection can be tested by comparing adult (parental) and zygotic genotypic frequencies:

$$\epsilon_{sj} = \alpha_{sj} \text{ and } \epsilon_j = \alpha_j.$$
 (19)

Inclusion of entire litters enables independent testing of gametic selection in males and females, sexual selection in males, and random mating, and removes much of the ambiguity in the interpretation of the tests for reproductive selection in males and mating patterns. For example, the analog of H_3 in the Christiansen-Frydenberg SCA may be interpreted as either gametic or sexual selection in males. By testing male gametic selection independently of sexual selection and including an estimate of the degree of distortion in the test of sexual selection in males, interpretation of the corresponding tests in the proposed SCA is unambiguous. The principal problem which remains despite inclusion of entire litters is that many regimes of very strong selection may occur and not produce a significant deviation from either H_2 or H_3 . As a result, a complete study of the components of reproductive selection in males seems impossible with an SCA of mother-offspring combinations.

The biological phenomena measured by p' depend not only on segregation distortion but also on the mating system. Consequently, we must consider the conditions under which the determinants of the mating system, the number of male mates per mother-offspring combination and mate preferences, affect the allelic ratio. Assuming no gametic, fecundity, sexual, or zygotic selection, it can be shown that neither multiple insemination nor non-random mating alone significantly alter the ratio of gametes from that expected in monogamous, randommating breeding systems. However, certain combinations of multiple insemination and non-random mating can alter the expected gametic ratio. Thus interpretation of distortion measured by p' depends on the mating system. In monogamous systems and in many random-mating systems, significant deviations reflect segregation distortion. In polygamous and non-random mating systems, significant deviations may reflect segregation distortion, particular combinations of multiple insemination and non-random mating, or both.

An analysis of selection components more completely characterizes selection regimes than is possible by any other method and is the only way to identify polymorphisms which are maintained by selection components which act in equal but opposite directions. Examples of such polymorphisms are provided by an SCA of the alcohol dehydrogenase and haemoglobin (β chain) polymorphisms in *Peromyscus maniculatus* (Nadeau & Baccus, 1980) wherein sexual selection in females balanced gametic selection in males. Distortion in males was detected only by the proposed tests for gametic selection in males and not by the corresponding tests of the Christiansen-Frydenberg SCA. These results clearly demonstrate that selection components must be examined as completely as possible in order to characterize selection in natural populations.

We thank F. B. Christiansen and an anonymous reviewer for their helpful comments on an earlier draft of this paper. We are grateful to D. Schenzle for help with the preparation of the computer program for the MLE of p'. This work was supported in part by NIH grant HD11682 to R.H.T. and by Boston University Sigma Xi and Graduate School grants to J.H.N.

REFERENCES

- BACCUS, R., HILLESTAD, H. O., JOHNS, P. E., MANLOVE, M. N., MARCHINTON, R. L. & SMITH, M. H. (1977). Prenatal selection in white-tailed deer. 31st Proceedings of the Southeast Association of Game and Fish Commissioners, San Antonio, Texas.
- BACCUS, R., JOULE, J. & KIMBERLING, W. J. (1980). Linkage and selection analysis of biochemical variants in the deer mouse, *Peromyscus maniculatus*. Journal of Mammalogy 61, 423-435.
- CHRISTIANSEN, F. B. & FRYDENBERG, O. (1973). Selection component analysis of natural polymorphisms using population samples, including mother-offspring combinations. *Theoretical Population Biology* **4**, 425-445.
- CHRISTIANSEN, F. B., FRYDENBERG, O. & SIMONSEN, V. (1973). Genetics of Zoarces populations. IV. Selection analysis of an esterase polymorphism using population samples including mother-offspring combinations. *Hereditas* 73, 291-304.
- CHRISTIANSEN, F. B., FRYDENBERG, O., GLYDENHOLM, A. O. & SIMONSEN, V. (1974). Genetics of *Zoarces* populations. Further evidence based on age group samples of a heterozygous deficit in the *Est II* polymorphism. *Hereditas* 77, 225-236.
- CHRISTIANSEN, F. B. & FRYDENBERG, O. (1976). Selection component analysis of natural polymorphisms using mother-offspring samples of successive cohorts. In *Population Genetics and Ecology* (ed. S. Karlin and E. Nevo), pp. 277-301. New York: Academic Press.

JOHNSON, N. L. & KOTZ, S. (1969). Discrete Distributions. Boston: Houghton Mifflin.

- LEWONTIN, R. C. (1974). The Genetic Basis of Evolutionary Change. New York: Columbia University Press.
- NADEAU, J. H. & BACCUS, R. (1980). Selection components of four allozymes in natural populations of *Peromyscus maniculatus*. Evolution. (In the Press.)