F Statistics in *Drosophila buzzatii*: Selection, Population Size and Inbreeding

Timothy Prout* and J. S. F. Barker†

**Department of Genetics, University of California, Davis, California 95616, and †Department of Animal Science, University of New England, Armidale, New South Wales 2351, Australia*

Manuscript received December 10, 1991
Accepted for publication January 26, 1993

ABSTRACT

Drosophila buzzatii is confined to reproducing in a well defined patchy environment consisting of rotting cactus cladodes which are ephemeral, permitting at most three generations. Flies emerging from such rots were used to estimate the additive genetic variance within rots and the genetic variance between rots for body size and also were electrophoresed to determine their genotypes at six polymorphic loci. *F* statistics were estimated from body size and allozyme data. The F_{ST} derived from body size was significantly larger than the allozyme F_{ST}. It is proposed this is due to selective differentiation of body size. The allozyme F_{ST} is used to estimate effective population size: $10 < N < 50$. It is suggested that the regularly observed positive F_{IS}'s could be due to partial sib mating, S. If so, the estimated lower bound is $S = 0.258$. Experiments are identified which could support or contradict these interpretations.

THE cactophilic *Drosophila* species provide a model system for studies in evolutionary biology (BARKER and STARMER 1982). One of these species, *Drosophila buzzatii*, is specific to the cactus niche, feeding and breeding in rotting cladodes and fruits of *Opuntia* cactus species. Extensive studies of electrophoretic variation in Australian populations have suggested that selection influences allozyme frequencies on different spatial scales for different loci (SOKAL, ODEN and BARKER 1987). If selection is involved in this geographic variation, these analyses cannot distinguish between direct effects on allozymes, or hitchhiking. However, if any selection is operating, the relevant variation determining it is most likely that among individual rots. BARKER, EAST and WEIR (1986) found significant allele frequency heterogeneity among rots within a locality, which results could not distinguish between direct selection, hitchhiking or simply genetic drift among parental flies ovipositing in different rots. The results reported in this note are an extension of these studies.

There are two different topics. One involves data on allozyme frequencies that are supplementary to our study of body size in *D. buzzatii* and the F_{ST} derived from body size (PROUT and BARKER 1989). We compare F_{ST} estimates from body size and from allozyme frequencies of the same flies to make inferences about selection in natural populations. The second topic concerns the interpretation of F_{ST} and F_{IS} derived from allozyme studies in this species in terms of breeding structure. *F* statistics were estimated using the methods of WEIR and COCKERHAM (1984), and we use their notation, *viz.* $F = F_{IT}$, $\theta = F_{ST}$, $f = F_{IS}$.

COMPARISON OF θ ESTIMATED FROM BODY SIZE AND FROM ALLOZYMES

In PROUT and BARKER (1989), the progeny of flies emerging from 19 rotting cactus cladodes were used to estimate the additive genetic variance for body size within each rot from the covariances of full siblings. We also observed significant differences between the means of these progeny, when reared under uniform laboratory conditions, and assumed the differences in rot means were genetic. From these data, it was possible to calculate θ as a measure of the genetic differentiation of the rot populations. However, there was no way to determine whether the cause of this differentiation was genetic drift or selection.

Here we report some additional results which suggest a solution to this problem. The new data come from an electrophoretic assay of six enzyme loci in 526 flies from 13 of the rots used for the body size study, which provides an independent estimate of θ.

The calculation of θ from body size was as follows: The mean additive genetic variance within rots was $V_W = 0.3720$ and the genetic variance between rots was $V_B = 0.1295$. Following FALCONER (1981), the ratio of V_B to V_W is related to θ as follows:

$$\frac{V_B}{V_W} = \frac{2\theta}{1 - \theta}.$$

So that,

$$\theta = \frac{V_B}{V_B + 2V_W}.$$

This assumes that *D. buzzatii* is like *Drosophila melanogaster*, in that there is essentially no dominance

TABLE 1

Estimates of θ and f from allozyme data on flies emerging from 13 rots collected at Trinkey, N.S.W.

Enzyme	f	θ
Pgm	−0.0718	0.0455
Aldox	0.0914	0.0234
Hex	0.1231	0.0473
Adh-1	−0.0010	0.0314
Est-1	0.1836	0.0225
Est-2	0.2106	0.0277
Mean	0.1233	0.0317
Bootstrap confidence intervals	0.0334–0.1802	0.0273–0.0400

variance for body size. ROBERTSON (1987), in his study of body size in *D. buzzatii*, also made this assumption, and in fact provided some direct evidence for the lack of directional dominance, as F_1 progeny from crosses among geographically separate populations were close to the midparent. The above estimate of θ also assumes that any genotype by environment interaction between laboratory and field operates the same way on the genetic variance among populations, V_B, as it does among individuals, V_W, and also that the genotype by environment interaction does not magnify the variance in the laboratory as compared to the field. This calculation resulted in $\theta = 0.1483$. A bootstrap procedure, resampling individual rot contributions 1000 times, resulted in a 95% confidence interval on this θ of 0.0417–0.2165.

Table 1 shows the results of analysis of the electrophoretic data for the six enzyme loci. F statistics were estimated from allele and genotype frequency data using the program DIPLOID (WEIR 1990). For θ, the jackknife mean = 0.0317 with a 95% confidence interval of 0.0273–0.0400.

The confidence intervals for θ estimated from allozyme data and from body size data do not overlap, indicating a significantly larger θ for body size. If differentiation among rots were due only to drift affecting all loci, θ estimates from allozymes and body size would not be significantly different. Thus, the larger θ for body size is most likely due to some form of selective differentiation for this trait. However, there is evidence that some of the allozymes may be subject to selective differentiation (BARKER, EAST and WEIR 1986), in which case this result indicates stronger selection for body size differentiation than allozyme differentiation.

In PROUT and BARKER (1989), we suggested a plausible mode of selection which could cause rot differentiation for body size. We observed considerable variation in the number of flies emerging from different rots, and the reduced size of these flies in the high yield rots compared to the low yield rots indicated substantial differences in larval crowding. ROFF (1981) conjectured, and WILKINSON (1987) has indirect evidence that the selection favoring large adult flies, which applies also to *D. buzzatii* (SANTOS *et al.* 1992), is countered by the advantage of faster developing larvae which produce genetically smaller adults (L. NUNNEY, personal communication). We propose that the more severe the larval crowding, the more intense will be this selection; so that variation in larval crowding results in the genetic differentiation of body size between rots. The positive, but nonsignificant, regression of the presumed genetic differences between offspring from different rots on the mainly phenotypic differences between their parents (PROUT and BARKER 1989) is suggestive. A larger survey of rots might confirm this relationship.

ZENG and COCKERHAM (1991) have recently derived the theory for the genetic variances within and between populations for a quantitative character, assuming selective neutrality. This is for Wright's island model where the local populations diverge due to drift and are held together by migration. In this model, the local populations are permanent and not ephemeral as in our case of rotting cactus cladodes. We suggest that there is much opportunity for a joint analysis of allozymes and quantitative characters, such as in our study, the only limitation being that it must be possible to perform "common garden" experiments in order to measure the genetic variance between populations. This would permit a test of ZENG and COCKERHAM's neutrality assumption. We surmise such studies would confirm that there is more selection on quantitative characters than there is on allozymes.

POPULATION SIZE AND INBREEDING

Interpreting θ: Before the electrophoretic data were available, there was no way to distinguish the neutral *vs.* selective cause of body size differentiation. Proceeding under the neutral hypothesis, we used the body size θ to estimate the effective population size within rots, but the smaller θ given by the allozyme data from the same rots suggests not only selection on body size, but also that the estimate of population size requires revision upwards.

For this purpose, we combined the means from Table 1 with those from eight other Australian studies in Table 2, seven of which are summarized from THOMAS and BARKER (1990), and one from Table 4 (see below). For θ, the mean, SD and approximate confidence limits are at the bottom of column 2.

The rotting cladodes constitute ephemeral patches allowing two generations (SANTOS, RUIZ and FONTDEVILA 1989), or at most three generations (THOMAS and BARKER 1990), before drying out. Assuming the allozymes are nearly neutral, the θ_t resulting from t generations of sizes N_i ($i = 1, 2, \ldots t$) drawn from a

TABLE 2

Estimates of θ and f derived from allozyme frequencies in flies emerging from nine sets of field collected rots

Source	θ	f	S_L
"Thorax" allozymes, Table 1	0.0317	0.1233	0.439
"Late" from Table 4	0.0483	0.0708	0.264
Trinkey, N.S.W., June 1986	0.059**	0.090*	0.330
Trinkey, N.S.W., Dec 1986	0.025**	0.115	0.413
O'Hara, N.S.W.	0.002*	0.143	0.500
Grandchester, Queensland	0.038***	0.070	0.262
Grandchester Hill, Queensland	0.056**	0.035	0.135
Borallon, Queensland	0.048***	0.153***	0.531
Hemmant, Queensland	0.037***	0.065	0.244
Mean	0.0383		0.3464
SD	0.0176		0.1323
SE	0.0059		0.0441
2 SE	0.0117		0.0882
CI	(0.0266–0.0500)		(0.2582–0.4346)

* Significant at $P < 0.05$, ** at $P < 0.01$, *** at $P < 0.001$.

TABLE 3

Estimates of population size derived using the upper confidence limits on $\theta(\theta_U, \theta_L)$, which provide N_L and N_U, respectively

t	$\tilde{N}_L$	$\tilde{N}_U$
1	10	18.8
2	20	37.6
3	30	56.4

large population is as follows:

$$\theta_t = 1 - \prod_{i=1}^{t} \left(1 - \frac{1}{2N_i}\right).$$

Neglecting higher order terms results in the following approximation:

$$\theta_t = \sum_{i=1}^{t} \left(\frac{1}{2N_i}\right).$$

From this relation, the harmonic mean size, $\tilde{N}$, over generations is:

$$\tilde{N} \cong \frac{t}{2\theta_t}$$

Table 3 gives estimates of $\tilde{N}$ from the upper and lower 95% confidence limits of θ for $t = 1, 2, 3$. It should be noted these estimates permit some variation in interpretation. If the flies collected from the rot are a mix of cohorts from different generations (see below), then the mean N for these cohorts would fall somewhere in the interval between those for the earliest and the latest generation. Also, if only one generation is small, say the founder generation, and the rest are large, $1/2N \cong 0$, then even if there were three generations, the estimated size would be that for the one small generation, which means that $N = 10$ is the lower bound when there are actually two or three generations. The values of $\tilde{N}$ in Table 3 are larger than those reported in PROUT and BARKER (1989) because the value of $\theta = 0.1483$ used there. Using the mean of the nine studies, $\theta = 0.0383$ (which is very close to the $\theta = 0.0317$ for the body size study), $\tilde{N} = 13.1$ for one generation, and $\tilde{N} = 26.1$ for two generations, which are in reasonable agreement with the estimates of $N = 10$ made by SANTOS, RUIZ and FONTDEVILA (1989) using inversion data, and by THOMAS and BARKER (1990) using allozyme data. The main conclusion is that average population sizes in the rots are probably less than 50 individuals. However, they could be larger if there were some selective differentiation at one or two of the loci contributing to θ.

Interpreting *f*: The estimates of f, within rots, were significantly positive which is a typical finding in these studies as shown in Table 2, where all f estimates are positive. Random mating within the rots should, if anything, result in a negative f amounting to $-1/(2N - 1)$ (KIMURA and CROW 1963).

There are several possible reasons for a positive f. Positive assortative mating is a formal possibility. However, this seems unlikely because with the exception of Adh the remaining five loci of Table 1 regularly have positive f's [Pgm in Table 1 is exceptional; BARKER, EAST and WEIR (1986)]. Null alleles are another possibility, but for these six loci, null alleles have been detected only at esterase-1 and esterase-2. Even for these loci, the estimated frequencies of null alleles are only 0.034 and 0.017, respectively (KNIBB *et al.* 1987), which would make an insignificant contribution to the positive f values. Another possibility is inbreeding resulting from sib mating, and finally there could be a temporal Wahlund effect. The last could result from pooling of flies from different cohorts over time. With the data of Table 1, there is no way to resolve these two most likely explanations of

TABLE 4

***F* statistics (SE) estimated from allele and genotype frequencies at six enzyme loci (same loci as in Table 1) in flies emerging from 43 rots collected at two localities in the Hunter Valley, N.S.W. (localities 5 and 60 in SOKAL, ODEN and BARKER, 1987)**

Emergencies	F	θ	f
Early	0.1058 (0.0643)	0.0476 (0.0035)	0.0610 (0.0650)
Late	0.1157 (0.0409)	0.0483 (0.0036)	0.0708 (0.0427)
All	0.1107 (0.0523)	0.0358 (0.0022)	0.0776 (0.0524)

sib mating and a temporal Wahlund effect. However, we have data from another study where the flies emerging from 43 rots were divided approximately into an early emerging half and a late emerging half and assayed separately. If there were a temporal Wahlund effect, then for the f within the early collection and within the late collection, each should be lower than that resulting from pooling the two collections. These data (Table 4) show no such tendency.

If there were sib mating, a positive f requires at least two generations. It is assumed that the founder females of a rot have mated with random males; so that the first generation should be in Hardy-Weinberg frequencies. But if some of these individuals remained and mated in the rot, then there is the possibility of sib mating providing that some sibships were sequestered in different parts of the rot. (Random mating between individuals drawn from a cluster of sibships results in Hardy-Weinberg or excess heterozygotes in a small population.) The early-late f estimates appear to contradict this theory, since $f > 0$ should appear only in the late collections. However, the early collections in this case could have included some second generation progeny from the older rots, and the two do vary in the right direction, the late f being larger than the early f, but not significantly so.

A partial sibmating model is derived in the APPENDIX. The object is simply to determine whether the amount of sib mating, S, required to explain the f estimates is large or small.

Assuming most rots dry up after two generations, the model assumes two generations where the female founders have mated with random males and a proportion S of their offspring engage in sib mating. The model also takes into account the possibility that the founders will colonize a rot over a period of time, so that flies collected from a rot are a mixture composed of a fraction v of second generation progeny and $(1 - v)$ first generation progeny (from the later colonists). The model also assumes that the founders, coming from other rots, are a mixture with a fraction u from second generation progeny and $(1 - u)$ first generation progeny. If the flies of these two generations emerging from a rot are equally successful at founding new rots, then $u = v$. Finally, the model assumes that the number of parents of the first and second generation are small, N_1 and N_2, respectively. The result is as follows:

$$f = \frac{S/4(v - 2v\theta_1 + u\theta_2 - u(1 - v)\theta_1) + \theta_1 - \theta_2}{(1 - \theta_2)(1 - Su/4)} \quad (1)$$

Solving for S, which is a mean over rots:

$$S = \frac{4[f(1 - \theta_2) + \theta_2 - \theta_1]}{v(1 - 2\theta_2) + u(f(1 - \theta_2) + \theta_2) - u(1 - v)\theta_1} \quad (2)$$

where the subscripts on θ refer to generation number. The only parameters on the right side of Equation 2 for which there are estimates are $\hat{\theta}_2$ and $\hat{f}$ ("^" = estimates), and even the observed $\hat{\theta}$ could be a mix of θ_1 and θ_2. The parameters u and v are unknown. It is shown in the appendix that with no information on u and v, S could be as large as its maximum, $S = 1$. However, it is also shown that it is possible to obtain a lower bound on S, or S_L. Fortunately this is perhaps the most interesting information. The lower bound occurs when $u = v = 1$ and $\theta_1 = \theta_2$ and is simply,

$$S_L = \frac{4f}{1 + f}$$

which is independent of $\hat{\theta}_2$. We apply this estimate of S_L to the data from the nine independent studies of Table 2. This table gives not only the estimates of θ used earlier, but also the mean f over the same loci as in Table 1 for each study. The first two listed are the mean f from Table 1 and the late f from Table 4. The remaining seven are taken from the large study of THOMAS and BARKER (1990). The table gives S_L for each case, and the mean S_L, its empirical standard error (SE), and the approximate confidence interval, *i.e.* ±2 SE. The lower confidence limit of the lower bound is $S_L = 0.2582$. The conclusion is that this is not small, and if the positive f is in fact due to partial sib mating, then the true value of S must be larger because of the extremely unlikely assumptions that $\theta_1 = \theta_2$ and $u = v = 1$.

A qualitative test of this seemingly large amount of partial sib mating can be obtained by sampling flies from young rots where not more than one generation could have possibly elapsed, where S must be zero under this sib mating model, and also samples of flies from very old rots where at least two generations could have elapsed, where f should be maximal. If the flies from the very young rots still give a positive f, this would falsify the above model and would suggest what at this point seem very unlikely possibilities such as sibmated females as rot founders, or that allozyme alleles are hitchhiking on assortative mating genes with different rot preferences (BARKER 1992), which, if so, would require a different interpretation of $\hat{\theta}$.

Finally, we would like to point out that in the great many allozyme studies the focus of interest is on θ, the measure of population differentiation, rather than on

the interpretation of the f's which always accompany such data [reviewed in CROUAU-ROY (1988)]. However, an attempt to interpret f can generate questions and perhaps answers concerning the organism's local life history, especially for species such as *D. buzzatii* that breed in ephemeral patches.

LITERATURE CITED

BARKER, J. S. F., 1992 Genetic variation in cactophilic Drosophila for oviposition on natural yeast substrates. Evolution **46:** 1070–1083.

BARKER, J. S. F., P. D. EAST and B. S. WEIR, 1986 Temporal and microgeographic variation in allozyme frequencies in a natural population of *Drosophila buzzatii*. Genetics **112:** 577–611.

BARKER, J. S. F., and W. T. STARMER (Editors), 1982 *Ecological Genetics and Evolution. The Cactus-Yeast-Drosophila Model System*. Academic Press, Sydney, Australia.

CROUAU-ROY, B., 1988 Genetic structure of cave dwelling beetles populations: significant deficiencies of heterozygotes. Heredity **60:** 321–327.

FALCONER, D. S., 1981 *Introduction to Quantitative Genetics*. Longman Group Limited, London.

KIMURA, M., and J. F. CROW, 1963 The measurement of effective population number. Evolution **17:** 279–288.

PROUT, T., and J. S. F. BARKER, 1989 Ecological aspects of the heritability of body size in *Drosophila buzzatii*. Genetics **123:** 803–813.

ROBERTSON, F. W., 1987 Variation of body size within and between wild populations of *Drosophila buzzatii*. Genetica **72:** 111–125.

ROFF, D., 1981 On being the right size. Am. Nat. **118:** 405–422.

SANTOS, M., A. RUIZ and A. FONTDEVILA, 1989 The evolutionary history of *Drosophila buzzatii*. XIII. Random differentiation as a partial explanation of chromosomal variation in a structured natural population. Am. Nat. **133:** 183–197.

SANTOS, M., A. RUIZ, J. E. QUEZADA, A. BARBADILLA and A. FONTDEVILA, 1992 The evolutionary history of *Drosophila buzzatii*. XX. Positive phenotypic covariance between field adult fitness components and body size. J. Evol. Biol. **5:** 403–422.

SOKAL, R. R., N. L. ODEN and J. S. F. BARKER, 1987 Spatial structure in *Drosophila buzzatii* populations: simple and direct spatial autocorrelation. Am. Nat. **129:** 122–142.

THOMAS, R. H., and J. S. F. BARKER, 1990 Breeding structure of natural populations of *Drosophila buzzatii*: effects of the distribution of larval substrates. Heredity **64:** 355–365.

WEIR, B. S., 1990 *Genetic Data Analysis*. Sinauer Assoc., Sunderland, Mass.

WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating *F*-statistics for the analysis of population structure. Evolution **38:** 1358–1370.

WILKINSON, G. S., 1987 Equilibrium analysis of sexual selection in *Drosophila melanogaster*. Evolution **41:** 11–21.

ZENG, Z.-B., and C. COCKERHAM, 1991 Variance of natural genetic variances within and between populations for a quantitative character. Genetics **129:** 535–553.

Communicating editor: B. S. WEIR

APPENDIX

Derivation of cyclic random mating and partial sib mating: An elementary model will be derived first with some restrictive assumptions. The behavior of the model will then be examined when these assumptions are relaxed.

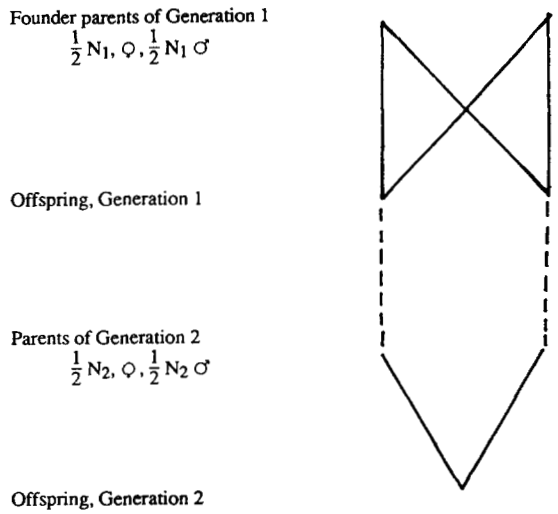

FIGURE 1.—Sib mating diagram. Symbols defined in the text.

TABLE 5

Basic relationships

Gene sources	Probability
Both descended from one gene in one of the two grandparents (*i.e.*, identical by descent)	1/4
Each descended from two different genes in one of the two grandparents	1/4
Each descended from two different grandparents	1/2

The model is diagrammed in Figure 1. A finite number of females, $N_1/2$, are founders of the rot after having mating with $N_1/2$ males. The subscript refers to the parents of generation one. These female parents lay a large number of eggs which constitute generation one. Then there is high egg to adult mortality leaving a small number of females, $N_2/2$ mated to $N_2/2$ males, which produce the second generation on which the observations are made.

The derivation will proceed by first calculating the frequency of homozygotes, P, for one allele at a diallelic locus. From this the value of F is obtained. Then f is derived using F and θ.

The parameter S denotes the fraction of generation 1 offspring which engage in sibmating; so that $(1 - S)$ mate at random. Table 5 shows the basic relationships between the sources of the two genes in the generation 2 offspring of sibmating, derived from their random mated grandparents who are parents of generation 1.

We now consider a diallelic locus with allele frequency p in the population from which founders of the rots are drawn. The finite numbers of parent founders of generation 1 results in a distribution of allele frequencies across rots with mean p. Consider a

subset of the rots drawn from this distribution, all with allele frequency x.

First consider the case of random mating in generation 1, with frequency $1 - S$. These parents of generation 2, drawn from the eggs of generation 1, consist of two independent samples of size ½ (N_2), namely, females and males. Therefore the frequency of homozygotes, $P_R(x)$, in their generation 2 offspring will be x^2, with no associated covariance (for the monoecious case, the homozygotes would be $x^2 + [x(1 - x)/2(N_2)]$). So,

$$P_R(x) = x^2.$$

Now consider the progeny of full-sib mating, with frequency S. Using Table 5, the frequency of homozygotes, $P_S(x)$, is

$$P_S(x) = \frac{1}{4}x + \frac{1}{4}\tilde{P}(x) + \frac{1}{2}x^2$$

where,

$\tilde{P}(x)$ = the frequency of homozygotes among the parents of generation 1.

This is based on the fact that among all populations with allele frequency x, the expected value of the allele frequency in their parents is x. The x^2 in the last term again has no covariance because the alleles transmitted by the two independent parents of generation 2 were derived from two independent grandparents. (See last item in Table 5)

Combining sibmating and random mating, the homozygotes, $P(x)$ in generation 2 in populations with allele frequency x, is:

$$P(x) = (1 - S)x^2 + S\left(\frac{1}{4}x + \frac{1}{4}\tilde{P}(x) + \frac{1}{2}x^2\right).$$

The frequency of homozygotes, P, over all rots is $E(P(x))$, with $E(x) = p$ is:

$$P = (1 - S)(P^2 + V) + S\left(\frac{1}{4}P + \frac{1}{4}\tilde{P} + \frac{1}{2}(P^2 + V)\right) \quad \text{(A1)}$$

where,

V = the variance of allele frequencies over rots in the first generation.

$\tilde{P}$ = the frequency of homozygotes over rots in the parents of the first generation.

At equilibrium $P = \tilde{P} = \hat{P}$: Solving (1) for $\hat{P}$,

$$\hat{P} = p^2 + \frac{V(1 - S/2) + (S/4)pq}{1 - S/4}$$

where

$$q = 1 - p. \quad \text{(A2)}$$

In general, θ is defined as V/pq. In this case, the variance (V) refers to the first generation, so $V = pq\theta_1$.

Substituting for V in (2),

$$\hat{P} = p^2 + \frac{pq(\theta_1(1 - S/2) + S/4)}{1 - S/4}.$$

F is defined by $P = p^2 + Fpq$, and in this case, the observed F is in the second generation, or F_2.

$$F_2 = \frac{\theta_1(1 - S/2) + S/4}{1 - S/4}. \quad \text{(A3)}$$

From the relation,

$$(1 - F_2) = (1 - \theta_2)(1 - f_2)$$

$$f_2 = \frac{F_2 - \theta_2}{1 - \theta_2}$$

substituting (3) for F_2 in the above,

$$f_2 = \frac{S/4\ (1 - 2\theta_1 + \theta_2) + \theta_1 - \theta_2}{(1 - S/4)(1 - \theta_2)}. \quad \text{(A4)}$$

It should be noted that when $S = 0$,

$$f_2 = \frac{\theta_1 - \theta_2}{1 - \theta_2}$$

and since,

$$\theta_2 = \frac{1}{2N_2} + \left(1 - \frac{1}{2N_2}\right)\theta_1$$

and,

$$\theta_1 = \frac{1}{2N_1}$$

then

$$f_2 = -\frac{1}{2N_2 - 1}$$

showing the negative f, or excess heterozygotes expected in small bisexual populations (KIMURA and CROW 1963).

Solving (A4) for S,

$$S = 4\left(\frac{\theta_2 - \theta_1 + f_2(1 - \theta_2)}{(1 - \theta_2)f_2 + 1 - 2\theta_1 + \theta_2}\right) \quad \text{(A5)}$$

It is useful at this point to summarize the behavior of this system by examining the deterministic case when $\theta_1 = \theta_2 = 0$, in which case (A4) becomes,

$$f_2 = \frac{S/4}{1 - S/4}.$$

This is the result of a cycle where a population alternately random mates and partial sib mates and the $f = F > 0$ occurs in the progeny of the partial sib mating phase, generation 2. In generation 1, the progeny of the random mating phase, of course, $f = F = 0$. If the grandparents of these progeny were in

Hardy-Weinberg proportions then, $f = S/4$ which is the standard result for partial sib mating. In the cyclic system, $f_2 > S/4$ because of the excess homozygotes (Equation A2) in the grandparents of the generation 2 progeny whose parents (generation 1) partially sib mated, but themselves were in Hardy-Weinberg proportions because the grandparents mated at random. This deterministic result can be easily derived using the method of identity by descent.

We now relax the assumption of high egg to adult mortality of the first generation and permit large numbers of the first generation to survive, but only a small number of these, N_2, remain to produce the next generation. We also relax the assumption of strictly discrete generations, so that the founders can colonize the rot over a period of time. These new assumptions lead to two modifications of the basic model. The first modification is that some mix of first and second generation progeny from other rots can colonize the rot over a period of time. The second is that a mix of first and second generation progeny are collected from the rot and are used to estimate f.

Consider the consequences of the first modification of a mix of first and second generation founders. In Equation A1, $\tilde{P}$, is the frequency of homozygotes among the founders. When first generation progeny are included $\tilde{P}$ becomes

$$\tilde{P} = (1 - u)p^2 + uP_2 \qquad \text{(A6)}$$

u = frequency of second generation founders and P_2 is the homozygotes among them. Substituting (A6) for P in Equation A1 and replacing P_2 for P on the left side of (A1), then solving for the equilibrium of $\hat{P}_2$, results in,

$$\hat{P}_2 = p^2 + \frac{pq(S/4 + \theta_1(1 - S/2))}{1 - \frac{Su}{4}}. \qquad \text{(A7)}$$

We now consider the second modification of including first generation progeny among the flies collected from the rot, which might not necessarily be the same numbers as those included in founders of new rots. Let v be the proportion of flies from the second generation, $(1 - v)$ from the first, and $\hat{P}_0$ the frequency of observed homozygotes, which is determined by the equilibrium, $\hat{P}_2$, and v.

$$\hat{P}_0 = (1 - v)(p^2 + \theta_1 pq) + V\hat{P}_2,$$

$\theta_1 pq = V$, the variance of first generation progeny.

Substituting the right side of Equation A7 for $\hat{P}_2$ in the above, the remaining calculations of F_2 and f_2 proceed as before with the following results,

$$f_2 = \frac{S/4(v - 2v\theta_1 + u\theta_2 - u(1 - v)\theta_1) + \theta_1 - \theta_2}{(1 - \theta_2)(1 - Su/4)}. \qquad \text{(A8)}$$

Solving for S,

$$S = \frac{4[f_2(1 - \theta_2) + \theta_2 - \theta_1]}{v(1 - 2\theta_1) + u(f_2(1 - \theta_2) + \theta_2) - u(1 - v)\theta_1}. \qquad \text{(A9)}$$

With no *a priori* information about u and v it is not possible to obtain a point estimate of S. However, it is possible to determine the bounds on S. The frequency of second generation flies, v, must be greater than zero in order for $f_2 > 0$. A small enough v requires S to be at its upper bound, $S = 1$, *i.e.*, no information of interest can be obtained concerning the upper bound. However, a lower bound can be determined which is of more interest, in any case. The lower bound with respect to u and v occurs when $u = v = 1$, which gives the maximum value of the denominator of (A9) with respect to these parameters, which do not appear in the numerator. The remaining unknown parameter in (A5) is θ_1; so it is now necessary to seek a lower bound on S with respect to this parameter.

With the restriction $0 \leq S \leq 1$, the right side of (A5) is a decreasing function of θ_1; so the lower bound on S is given by the maximum value of θ_1. It can be shown that this one generation of partial sibmating has no effect on θ_2; so that

$$\theta_2 = \frac{1}{2N_2} + \left(1 - \frac{1}{2N_2}\right)\theta_1$$

θ_2 here is larger than θ_1 except when $1/2N_2 = 0$, in which case $\theta_2 = \theta_1$. This provides the maximum value of θ_1 and therefore the lower bound on S. In this case (A5) gives

$$S_L = \frac{4f_2}{1 + f_2}$$

which is the result for the fully deterministic case discussed above. This shows that the small number of founding grandparents would play no role in their deterministic progeny (*i.e.*, when $1/2N_2 = 0$). This simple result, independent of θ_2, is the most information which can be obtained from data of this sort.