

Allele frequency estimates when only heterozygotes can be recognized: method of estimation and application in the case of chromosomal inversion polymorphisms in *Drosophila*

L. B. Klaczko,*
P. A. Otto† and
A. A. Peixoto*

* Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Cx. Postal 68011, 21949 RJ, Brazil.

† Departamento de Biologia, Instituto de Biociências, Universidade de São Paulo, Cx. Postal 11461, São Paulo, 05499 SP, Brazil.

A general method to estimate allele frequencies when there are more than two alleles segregating in an autosomal and only the heterozygotes can be identified is presented. The particular three allele case is detailed and the method is applied to chromosomal inversion data of *Drosophila mediopunctata*. The estimates from egg sample data are compared with those from adult male sample data. No significant differences were found between the two methods. However, the combined use of both data unveiled significant variations in the frequencies of inversions at different collecting times, that otherwise would not have been detected. Furthermore, the method is applied to data on inversion frequencies of *D. nebulosa*, *D. willistoni* and *D. flavopilosa*.

INTRODUCTION

Frequently in the analysis of chromosome inversion polymorphism in *Drosophila* a situation arises where it is relatively easy to recognize the heterokaryotypes while very difficult to distinguish different homokaryotypes. Usually under these circumstances only data on the frequency of heterokaryotype (or heterozygosis) are presented since it is impossible to determine which allele has the higher frequency (*i.e.*, a heterozygosity of, say, 32 per cent can be explained either by a frequency of $A=0.8$ or 0.2). See for example for *D. polymorpha*, *D. cardinoides*, *D. guaranunu*, *D. nebulosa*: Da Cunha *et al.*, 1953; for *D. willistoni*: Valente and Araújo, 1986; and for *D. nasuta*: Kumar and Gupta, 1988.

When more than two gene arrangements that overlap occur, they can be treated as “alleles” and it is still possible to identify the different heterozygotes but again it may be difficult to distinguish among homozygotes. In this case, however, one can estimate each one of the allele frequencies and their standard errors. As a matter of fact, such a situation occurs in *D. mediopunctata*.

In the proximal region of this species' second chromosome there are three big overlapping inversions: PA, PB and PC (actually, there are also five inversions included in PC which are all in low frequency, which for the purposes of the present paper can be ignored). The heterozygote PB/PC can be easily distinguished from PA/PB and PA/PC individuals by the proximal breaking point, while the other two have very distinct “subloops” of different sizes: this allows one to rapidly identify the different heterozygotes. On the other hand, homozygotes are difficult to score reliably in routine preparations (Peixoto and Klaczko, 1987; a photographic map of the inversions will be published elsewhere).

In the present paper we present a general method for estimating allele frequencies as well as their standard errors when there are more than two alleles at a locus and only heterozygotes can be identified. We apply this method to *D. mediopunctata* chromosomal data, and compare the results with the conventional technique of genotype counting. Furthermore, we also apply it to data on inversion frequencies of *D. nebulosa*, *D. willistoni* and *D. flavopilosa*.

MATERIAL AND METHODS

The flies used in this work were collected in the Parque Nacional do Itatiaia. This National Park is on the slopes of the Serra da Mantiqueira, States of Rio de Janeiro and Minas Gerais, Brazil (22°25'S, 44°50'W). Its fauna was described by Barth (1957) and its flora by Brade (1956) who characterized it as a sub-tropical hygrosopic forest. We collected *Drosophila* at various points in the Park at different altitudes: area A at 1300 m, area C at 1020 m, area E at 970 m, F at 1200 m, G at 1600 m, H at 700 m and I at 870 m. We used approximately a dozen fermented banana baits per area spaced somewhere between 5 to 15 metres. The flies were collected on 29 and 30 September 1986, 25 and 26 February and 30, 31 May and 1 and 2 June, 1987. Unfortunately, only in areas C and E could we collect on all three occasions.

After collection we separated *D. mediopunctata* from the other flies which were classified to species or group. The *D. mediopunctata* were brought to the laboratory to estimate inversion frequencies. This was done by two methods: (a) "adult male sample" and (b) "egg sample" (cf. Dobzhansky and Levene, 1948; Arnold, 1981).

The adult male sample method consists in crossing individually each collected male with three virgin females of a homokaryotypic strain of known genetic composition. From each cross the salivary gland chromosomes of up to eight F₁ larvae were analyzed to determine the male's genotype. The probability of incorrect assignments is less than 1 per cent (therefore negligible) and the constitution for the second chromosome can be precisely determined given that the heterozygotes can be easily distinguished and only one type of homozygote may appear in the F₁ having the mother's constitution.

The egg sample method consists of analysing one F₁ larva from each collected female. In this case, only the heterozygotes can be individually identified, while the homozygotes are pooled in one single category.

METHOD OF ESTIMATION

The case of two autosomal alleles has an explicit but ambiguous solution, unless the observed number of homozygotes is exactly equal to the number of heterozygotes (then $p = q = 0.5$). The genetic properties of this situation have been studied in detail by Cotterman (1953), who coined for it the

term "unnamed system" (phenogram number 2-2-2 shown in his fig. 1 (p. 201)).

The case of two alleles has theoretical interest only; besides the estimates being ambiguous, the adherence of the expected to the observed frequencies cannot be tested, because there are only two classes of observations and therefore one is left with no degrees of freedom. For this case, letting n_1 and n_2 be the observed numbers of homozygotes ($AA + BB$) and AB heterozygotes respectively, and putting $x = 2p(1-p)$ it follows that the maximum likelihood function is given, in logarithmic form, by

$$L = n_1 \ln(1-x) + n_2 \ln(x) + \text{const.};$$

the estimate is then $x = n_2/N$, with binomial variance $x(1-x)/N$, where $N = n_1 + n_2$. Using the principle of functional invariance the estimate p is given by

$$\begin{aligned} p &= (1/2) \pm (1/2)(1-2x)^{1/2} \\ &= (1/2) \pm (1/2)(1-2n_2/N)^{1/2}; \end{aligned}$$

the variance of the estimate p is obtained straightforwardly from

$$\begin{aligned} \text{var}(p) &= \text{var}(x) \cdot (dp/dx)^2 \\ &= \frac{x(1-x)}{4N(1-2x)} = \frac{2p(1-p)[1-2p(1-p)]}{4N(1-2p)^2} \end{aligned}$$

When the number of alleles segregating at an autosomal locus is greater than two, the estimation of allele frequencies takes place without ambiguity and, furthermore, the adherence of expected frequencies based on the estimates to the observed ones can always be tested, as we show in the lines that follow.

Letting n_1 , n_2 , n_3 and n_4 be respectively the observed numbers of ($AA + BB + CC$), AB , AC and BC individuals, p , q and $r = 1 - p - q$ be the frequencies of the three autosomal alleles (A , B , C) and assuming panmixia, the maximum likelihood function is given by

$$\begin{aligned} L &= n_1 \ln[p^2 + q^2 + (1-p-q)^2] \\ &+ (n_2 + n_3) \ln p + (n_2 + n_4) \ln q \\ &+ (n_3 + n_4) \ln(1-p-q) + \text{const.}; \end{aligned}$$

the system of equations

$$\begin{aligned} \partial L / \partial p &= 2n_1[p - (1-p-q)]/[p^2 + q^2 + (1-p-q)^2] \\ &+ (n_2 + n_3)/p - (n_3 + n_4)/(1-p-q) = 0 \\ \partial L / \partial q &= 2n_1[q - (1-p-q)]/[p^2 + q^2 + (1-p-q)^2] \\ &+ (n_2 + n_4)/q - (n_3 + n_4)/(1-p-q) = 0 \end{aligned}$$

does not have an explicit solution, so the values of p and q that satisfy the equations above are determined using an iterative numerical method, such as the generalized Newton-Raphson procedure:

$$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix} = \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} + \begin{pmatrix} -m_{11} & -m_{12} \\ -m_{21} & -m_{22} \end{pmatrix}^{-1} \begin{pmatrix} \partial L/\partial p \\ \partial L/\partial q \end{pmatrix}$$

$$= \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} + \begin{pmatrix} \text{var}(p) & \text{cov}(p, q) \\ \text{cov}(p, q) & \text{var}(q) \end{pmatrix} \begin{pmatrix} \partial L/\partial p \\ \partial L/\partial q \end{pmatrix},$$

where

$$m_{11} = \partial^2 L/\partial p^2 = 4n_1\{p^2 + q^2 + (1-p-q)^2 - [p-(1-p-q)]^2\} / [p^2 + q^2 + (1-p-q)^2]^2 - (n_2 + n_3) / p^2 - (n_3 + n_4)/(1-p-q)^2$$

$$m_{12} = m_{21} = \partial^2 L/\partial p \partial q = 2n_1\{p^2 + q^2 + (1-p-q)^2 - 2[p-(1-p-q)] \times [q-(1-p-q)]\} / [p^2 + q^2 + (1-p-q)^2]^2 - (n_3 + n_4)/(1-p-q)^2$$

$$m_{22} = \partial^2 L/\partial q^2 = 4n_1\{p^2 + q^2 + (1-p-q)^2 - [q-(1-p-q)]^2\} / [p^2 + q^2 + (1-p-q)^2]^2 - (n_2 + n_4) / q^2 - (n_3 + n_4)/(1-p-q)^2$$

The values $p_0, q_0; p_1, q_1; p_2, q_2;$ etc. converge to estimates p, q provided that p_0 and q_0 , the initial guesses, are not very far from p and q . Generally, since

$$n_4/n_3 \cong 2Nq(1-p-q)/2Np(1-p-q) = q/p \quad q \cong pn_4/n_3$$

and $2Npq \cong 2Np \cdot pn_4/n_3 = n_2$, where $N = n_1 + n_2 + n_3 + n_4$, good guesses for the initial trial values p_0 and q_0 are given by $p_0 = (n_2n_3/2Nn_4)^{1/2}$ and $q_0 = (n_2n_4/2Nn_3)^{1/2}$. The formulae for the case of three autosomal alleles were deliberately left without algebraic simplifications. The mere inspection of these formulae enables us to generalize the situation for the case of k alleles segregating at an autosomal locus: let p_1, p_2, \dots, p_k be the frequencies of the alleles segregating at an autosomal locus. The maximum likelihood function is then given by:

$$L = n_h \ln \left[\sum_{i=1}^{k-1} p_i^2 + \left(1 - \sum_{i=1}^{k-1} p_i\right)^2 \right] + \sum_{i=1}^{k-1} n_i \ln p_i + n_k \ln \left(1 - \sum_{i=1}^{k-1} p_i\right) + \text{const.},$$

where

- n_h = number of homozygotes in the sample;
- n_i = number of individuals that are heterozygotes for the i th allele of the locus; and
- n_k = number of individuals that are heterozygotes for the k th allele, which frequency is estimated by difference since k different alleles define a system of $k-1$ linearly independent equations.

The quantities necessary for the application of the numerical iterative procedure shown in detail for the case of three alleles are the following ones:

$$\begin{aligned} \partial L/\partial p_i &= 2n_h \left[p_i - \left(1 - \sum_{i=1}^{k-1} p_i\right) \right] \\ &\quad / \left[\sum_{i=1}^{k-1} p_i^2 + \left(1 - \sum_{i=1}^{k-1} p_i\right)^2 \right] \\ &\quad + n_i/p_i - n_k / \left(1 - \sum_{i=1}^{k-1} p_i\right) \\ \partial^2 L/\partial p^2 &= 4n_h \left\{ \sum_{i=1}^{k-1} p_i^2 + \left(1 - \sum_{i=1}^{k-1} p_i\right)^2 \right. \\ &\quad \left. - \left[p_i - \left(1 - \sum_{i=1}^{k-1} p_i\right) \right]^2 \right\} \\ &\quad / \left[\sum_{i=1}^{k-1} p_i^2 + \left(1 - \sum_{i=1}^{k-1} p_i\right)^2 \right]^2 \\ &\quad - n_i/p_i^2 - n_k / \left(1 - \sum_{i=1}^{k-1} p_i\right)^2 \\ \frac{\partial^2 L}{\partial p_i \partial p_j} &= 2n_h \left\{ \sum_{i=1}^{k-1} p_i^2 + \left(1 - \sum_{i=1}^{k-1} p_i\right)^2 \right. \\ &\quad \left. - 2 \left[p_i - \left(1 - \sum_{i=1}^{k-1} p_i\right) \right] \left[p_j - \left(1 - \sum_{i=1}^{k-1} p_i\right) \right] \right\} \\ &\quad / \left[\sum_{i=1}^{k-1} p_i^2 + \left(1 - \sum_{i=1}^{k-1} p_i\right)^2 \right]^2 \\ &\quad - n_k / \left(1 - \sum_{i=1}^{k-1} p_i\right)^2 \end{aligned}$$

The initial guesses ($p_{1,0}, p_{2,0}, \dots$, etc.) are determined using algebraic procedures such as the ones shown for the case of three alleles.

For testing the adherence of the expected numbers to the observed ones, we use a chi-square test with $(k-2)(k-1)/2$ degrees of freedom, since there are, in the case of k alleles, $k(k-1)/2 + 1$ different classes and to calculate the expected values we use k parameters estimated from the sample ($k-1$ allele frequencies and the total number of sampled individuals).

Table 1 Observed numbers of homozygotes and heterozygotes per area and collecting date in the egg sample. *N* is the number of individuals analysed.

Area	A+G+F		C			E			H+I
	Feb.	Sep.	Feb.	May	Sep.	Feb.	May	May	
Homozygotes	5	16	8	9	21	12	17	9	
PA/PB	4	7	8	6	15	2	7	7	
PA/PC	1	6	5	9	9	4	14	7	
PB/PC	1	8	5	0	3	6	7	4	
<i>N</i>	11	37	26	24	48	24	45	27	

Table 2 Observed numbers of individuals for the various karyotypes per area and collecting date in the adult male sample. *N* is the number of individuals analysed

Area	A+G+F		C			E			H+I
	Feb.	Sep.	Feb.	May	Sep.	Feb.	May	May	
PA/PA	3	18	8	30	18	1	27	8	
PB/PB	0	7	2	4	2	3	5	3	
PC/PC	0	6	0	995	4	0	2	1	
PA/PB	3	19	9	29	29	6	11	7	
PA/PC	3	16	3	26	16	9	16	18	
PB/PC	2	8	5	18	8	11	8	9	
<i>N</i>	11	74	27	112	77	30	69	46	

RESULTS

Table 1 shows the observed numbers of homozygotes and heterozygotes on the three collecting dates per area in the egg sample. The data from areas A, G and F (above 1200 m of altitude) were pooled due to the small size of the samples in each of them. The same was done for areas H and I (under 900 m). Similarly, the observed numbers of different karyotypes per area and collecting date in the adult male sample are in table 2.

Using these data we estimated the allele frequencies per collecting date. For the adult male sample this was done by a direct count of genotypes; for the egg sample we used the method outlined above. The estimates with their standard errors are shown on table 3. As can be seen, the estimates for each collecting date using the two methods are in good agreement. To test the significance of the differences between male and egg estimates (p_m and p_f , respectively) we used the z approximation to the normal,

$$z = (p_m - p_f) / [p_{m+f}(1 - p_{m+f})(1/n_m + 1/n_f)]^{1/2}$$

where p_{m+f} is a combined estimate obtained by weighing p_m and p_f by the reciprocals of their

variances and n_m and n_f are the total numbers of scored chromosomes in the male and egg samples, respectively. The last column of table 3 shows the calculated z values. The 5 per cent critical value

Table 3 Estimated allele frequencies in each collection. The numbers represent the estimated allele frequencies and standard errors expressed in per cent in the egg, male sample and total sample

Date	Egg	Male	Total	z
September				
PA	51.9 ± 7.9	50.3 ± 2.9	50.5	0.33 (ns)
PB	28.5 ± 6.8	27.2 ± 2.6	27.3	0.31 (ns)
PC	19.6 ± 4.4	22.5 ± 2.4	21.8	0.74 (ns)
2 <i>N</i>	170	302	472	
February				
PA	30.8 ± 9.2	41.9 ± 4.2	40.0	1.82 (ns)
PB	43.3 ± 11.1	33.8 ± 4.1	34.9	1.60 (ns)
PC	25.9 ± 7.0	24.3 ± 3.7	24.6	0.31 (ns)
2 <i>N</i>	122	136	258	
May				
PA	50.1 ± 6.3	52.2 ± 2.3	51.9	0.49 (ns)
PB	20.0 ± 3.9	23.4 ± 2.0	22.7	0.94 (ns)
PC	30.0 ± 5.6	24.5 ± 2.0	25.1	1.49 (ns)
2 <i>N</i>	192	454	646	

is 1.96 and not a single value is above it, demonstrating that there are no differences in the estimates given by the two methods. Thus, we can use the combined estimates for the allele frequencies, which are also shown in table 3.

The heterogeneity among collecting dates was tested pooling the three homozygotes in each of the adult male samples and adding the respective egg sample data. We performed a contingency chi-square test which showed a significant difference among collecting dates ($\chi^2 = 18.2$, 6 df, $P = 0.006$).

We found no difference among areas comparing either the distribution of homo and heterozygotes in the total data ($\chi^2 = 7.9$, 9df, $P > 0.50$) or the observed numbers for the alleles in the adult male sample ($\chi^2 = 3.6$, 6 df, $P > 0.70$). Again, the estimates (not shown) using both samples are in very good agreement.

We tested for deviations from the Hardy-Weinberg equilibrium in every sample (area and collecting date) and in only one case did we find departures from the expected: in the male adult sample of February ($\chi^2 = 9.37$, 3 df, $P < 0.05$). Actually, in all male samples there was a slight but non-significant excess of heterozygotes.

DISCUSSION

Deviations from the Hardy-Weinberg expected distribution with excess of heterozygotes have been found before in the work with chromosome inversions in *Drosophila* using adults. They are usually interpreted as evidence in favour of heterosis (Dobzhansky, 1970). We might, similarly, take as suggestive evidence of heterosis the significant result obtained in the February male sample.

One of the assumptions of the method proposed here is that the genotypes distribution is in Hardy-Weinberg ratios. However, this method is not likely to be used in adult male samples (where the different genotypes can be precisely identified) but only in egg samples. In none of the egg samples analyzed did we find any significant departures from Hardy-Weinberg expectations; nor was there any trend in the data indicating an excess of heterozygotes. This contrasting result with the male sample can be easily understood if one takes into account the nature of the two methods. With the male method one is estimating allele frequencies in adults after selection—or part of it (viability)—has already occurred. With the egg sample method the larvae are grown under optimal conditions

where differences in viability are minimized leaving little room for selection to operate.

Examples of temporal variation in the frequencies of chromosome inversions are also abundant in the *Drosophila* bibliography (for a review see Sperlich and Pfriem, 1986). In some species there is seasonal variation (as in the well known case of *D. pseudoobscura*) while in others the temporal variation may be non-cyclical. Evidently more data are needed to ascertain the nature of the variation observed, as well as for an understanding of *D. mediopunctata*'s chromosome inversions relation to the environment. We are presently working in this direction.

The standard errors associated with the estimates of allele frequencies using the method here proposed are about two to three times larger than the ones that would be found had we been able to distinguish every karyotype. However, had we used only the male sample, where this is possible, we would not detect the difference between collecting dates. The heterogeneity test for the observed allele frequencies in the male sample is non-significant ($\chi^2 = 7.05$, 4 df, $P > 0.10$). Thus, the use of all the available information, increasing the sample size, allowed us to detect differences that otherwise would pass unsuspected.

Finally, a comment on the general applicability of this method seems necessary. Whenever three or more overlapping inversions occur (or if they do not overlap, but the distance among them is small enough to prevent recombination) they can be treated as alleles and this method can be applied to egg sample data. Actually, such a situation seems to exist in a good number of species. For example, the third chromosome of *D. nebulosa* has in its distal region a number of overlapping inversions. In two populations from Rio Grande do Sul (Brazil), in addition to the St gene arrangement there are two inversions segregating: G and H (Regner and Valente, 1988; Regner and Valente, personal communication). Table 4 shows the observed numbers of the different heterozygotes, the estimated frequencies and the test for deviations from the Hardy-Weinberg equilibrium. The application of the method here proposed allows one to identify the difference between the two populations as due mainly to the frequency of the St gene arrangement (44.0 per cent at Ipanema and 73.0 per cent at Estação Experimental).

Da Cunha *et al.* (1950) give for overlapping inversions L and M of *D. willistoni*'s third chromosome the frequencies of heterozygotes L/M, L/- and M/-. For the populations where the frequencies were above 1 per cent we applied the method

Table 4 Observed numbers for the different karyotypes for the third chromosome inversions of *Drosophila nebulosa* and estimated inversion frequencies (in per cent) at two localities in Rio Grande do Sul, Brasil. Data from Regner & Valente, 1988.

	Estação experimental	Ipanema
Observed numbers		
G/St	35	33
H/St	64	40
G/H	5	20
Homozygotes	132	40
N	236	133
Estimated frequencies		
St	73.0 ± 1.5	44.0 ± 4.9
G	9.3 ± 2.3	25.5 ± 3.7
H	17.7 ± 2.5	30.5 ± 4.3
χ^2	1.48 ^{ns}	1.85 ^{ns}

here proposed. Table 5 shows the original data and the inversion estimated frequencies as well as the test for departure from Hardy-Weinberg expectations for two populations from Goiás, Brasil. A third one (Rio Negro, Amazonas) was also analysed but there was not a good fit to the data ($\chi^2 = 9.59$, 1 df, $P < 0.01$).

Brcnic (1966) studied the inversions that occur in chromosome V-R of *D. flavopilosa*. In this chromosome, in addition to the St gene arrangement, there are four overlapping inversions, but only two of them are widely distributed and relatively abundant (A and B). This author in a carefully designed experiment analysed the chromosomes of the larvae present and grown in *Cestrum* flowers brought to the laboratory directly from the field. He collected them on five localities at different altitudes at Maipo Valley, Chile. Table 6 shows the observed and expected numbers for the

Table 5 Observed numbers for the different karyotypes for the third chromosome inversions of *Drosophila willistoni* and estimated inversion frequencies (in per cent) at two localities in Goiás, Brazil. Data from Da Cunha *et al.* (1950)

	Palma	Monjolinho
Observed numbers		
L/-	11	16
M/-	11	12
L/M	21	57
N	72	128
Estimated frequencies		
L	40.2 ± 11.8	46.8 ± 5.7
M	40.2 ± 11.8	41.0 ± 5.7
χ^2	0.58 ^{ns}	2.83 ^{ns}

various genotypes for inversions A and B (other inversions were lumped with the St gene arrangement; data from his paper's table 2). In all localities but one (El Vulcán) we detected deviations from Hardy-Weinberg expectations. This is particularly relevant, since the larvae analysed were grown, not under optimal conditions but rather, at almost natural conditions. For the four samples where there is a significant deviation, there is an excess of heterozygotes. This strongly supports the idea of heterosis under natural conditions, as Brcnic (1966) suspected based on "the high frequency of some heterokaryotypes in certain populations (sometimes close to or above 50 per cent)".

In spite of the observed deviations from Hardy-Weinberg we used the estimates obtained to verify the existence of altitudinal clines. Table 7 shows the estimated frequencies for inversions A and B in each locality. We calculated the correlation between inversion frequencies (after angular transformation) and altitude: it is significant and positive for inversion A ($r = 0.945$; 3 df; $P < 0.02$) and

Table 6 Observed (and expected) numbers for the different genotypes for chromosome V-R inversions A and B of *Drosophila flavopilosa* in various localities at Maipo Valley, Chile. Data from Brcnic, 1966

Locality	Homozygotes	A/-	B/-	A/B	χ^2
El Tabo	584 (596.2)	99 (88.1)	276 (264.0)	10 (20.7)	7.67**
Vizcachas	546 (573.1)	377 (350.3)	117 (95.6)	12 (33.0)	21.42***
San Alfonso	90 (94.4)	63 (58.6)	16 (12.6)	1 (4.4)	4.10*
Queltehues	565 (582.0)	440 (423.1)	70 (58.6)	14 (25.3)	8.49**
El Vulcán	257 (262.0)	195 (190.0)	13 (9.8)	1 (4.2)	3.72 ^{ns}

* $P < 0.05$. ** $P < 0.01$. *** $P < 0.001$. ^{ns} $P > 0.05$.

Table 7 Estimated inversion frequencies of *Drosophila flavopilosa* in various localities (at different altitudes) at Maipo Valley, Chile. Data from Brncic, 1966

Locality	A	B	N
El Tabo (20 ft)	5.97 ± 0.58	17.89 ± 1.23	969
Vizcachas (2700 ft)	23.95 ± 1.38	6.54 ± 0.58	1052
San Alfonso (3600 ft)	24.57 ± 3.51	5.26 ± 1.28	170
Queltehues (4400 ft)	29.00 ± 1.71	4.01 ± 0.44	1089
El Volcán (5500 ft)	29.61 ± 2.73	1.53 ± 0.41	466

highly significant and negative for inversion B ($r = -0.991$; 3 df; $P < 0.001$). This confirms Brncic's interpretation of an altitudinal gradient.

In these species as well as in other ones there are data which have been published and accumulated for decades using heterozygosity in egg samples. Among others, for example, there are the species of the *nasuta* subgroup: *D. sulfurigaster albostrigata*, *D. albomicans*, *D. kohkoa*, *D. nasuta* (Clyde, 1982; Kumar and Gupta, 1988; Mather and Thongmeearkom, 1980). Unfortunately, however, most authors have published only the heterozygosity for each inversion and not the frequencies of the various heterozygotes—which are the critical data for analysis. Nevertheless, these data should be available to these authors in their protocols.

The analysis of heterozygote frequencies may lead in some cases to the discovery of seasonal and geographical patterns. However, one might recall that if the frequency of a given allele varies from 30 to 70 per cent, its heterozygote frequency varies only from 42 to 50 per cent, that is, a difference of 40 per cent in allele frequency translates in a difference of only 8 per cent in heterozygosity. When allele frequencies are intermediate and the amplitude of variation is small, it should be easier to detect patterns using allele frequencies than using heterozygosity. Thus the application of the method here proposed might help to verify the existence of clines or cycles in the inversion frequencies, since it permits the estimation of allele frequencies from heterozygosity data.

Acknowledgements It is a pleasure to acknowledge Ms Blanche C. Bitner-Mathé and Clea K. da Silva for technical assistance. Ms. L. Regner and Dr V. Valente kindly allowed us to use their unpublished data on the inversion frequencies of *D. nebulosa*. This work was supported by grants from CEPEG-Universidade Federal do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa de São Paulo (FAPESP), Fundação Universitária José Bonifácio and Financiadora de Estudos e Projetos (FINEP).

REFERENCES

- ARNOLD, J. 1981. Statistics of natural populations. I. Estimating an allele probability in cryptic fathers with a fixed number of offspring. *Biometrics*, **31**, 145–160.
- BARTH, R. 1957. A fauna do Parque Nacional do Itatiaia. *Bol. Pq. Nac. Itatiaia*, (6).
- BRADÉ, A. C. 1956. A flora do Parque Nacional do Itatiaia. *Bol. Pq. Nac. Itatiaia*, (5).
- BRNCIC, D. 1966. Ecological and cytogenetic studies of *Drosophila flavopilosa*, a neotropical species living in *Cestrum* flowers. *Evolution*, **20**, 16–29.
- CLYDE, M. 1982. Inversion polymorphism in Southeast Asian populations of *Drosophila nasuta* subgroup. *Genetica*, **58**, 161–175.
- COTTERMAN, C. W. 1953. Regular two-allele and three-allele phenotype systems, Part I. *Amer. J. Hum. Genet.*, **5**, 193–235.
- DA CUNHA, A. B., BRNCIC, D. AND SALZANO, F. M. 1953. A comparative study of polymorphism in certain South American species of *Drosophila*. *Heredity*, **7**, 193–202.
- DA CUNHA, A. B., BURLA, H. AND DOBZHANSKY, TH. 1950. Adaptive chromosomal polymorphism in *Drosophila willistoni*. *Evolution*, **4**, 212–235.
- DOBZHANSKY, TH. 1970. *Genetics of the Evolutionary Process*. Columbia University Press, New York.
- DOBZHANSKY, TH. AND LEVENE, H. 1948. Genetics of natural populations. XVII. Proof of operation of natural selection in wild populations of *Drosophila pseudoobscura*. *Genetics*, **33**, 537–547.
- KUMAR, A. AND GUPTA, J. P. 1988. Genetics of natural populations of *Drosophila nasuta*. *J. Hered.*, **79**, 83–88.
- MATHER, W. B. AND THONGMEEARKOM, P. Chromosome map of *D. albomicans*. *Dros. Inf. Serv.*, **55**, 101–102.
- PEIXOTO, A. A. AND KLACZKO, L. B. 1987. Polimorfismos de inversões cromossômicas em *Drosophila mediopunctata*. II: Desequilíbrio de ligação. *Ciência e Cultura*, **39**, (supl.), 767.
- REGNER, L. P. AND VALENTE, V. 1988. Comportamento sexual e polimorfismo cromossômico em *Drosophila nebulosa*. *Ciência e Cultura*, **40**, (supl.), 769.
- SPERLICH, D. AND PFRIEM, P. 1986. Chromosomal polymorphism in natural and experimental populations. In Ashburner, M., Carson, H. L. and Thompson, J. N. (eds) *The Genetics and Biology of Drosophila*, Vol. 3e, Academic Press.
- VALENTE, V. AND ARAUJO, A. 1986. Chromosomal polymorphism, climatic factors and variation in population size of *D. willistoni* in Southern Brazil. *Heredity*, **57**, 149–159.