

GENETIC STRUCTURE AND CHANGE IN NATURAL POPULATIONS OF  
*DROSOPHILA ROBUSTA*: SYSTEMATIC INVERSION AND  
INVERSION ASSOCIATION FREQUENCY SHIFTS  
IN THE GREAT SMOKY MOUNTAINS

WILLIAM J. ETGES

Department of Biology, University of Rochester, Rochester, New York 14627

Received December 14, 1982. Revised October 25, 1983

Few situations in nature suggest the operation of natural selection as strongly as latitudinal and altitudinal clines of chromosomal inversions in natural populations of *Drosophila*. Gradients in inversion frequency have implicated selection with the maintenance of inversion polymorphisms in *Drosophila pseudoobscura*, *D. persimilis* (Dobzhansky, 1948; Speiss, 1950), *D. robusta* (Stalker and Carson, 1948), *D. pachea* (Ward et al., 1974), and *D. melanogaster* (Mettler et al., 1977; Stalker, 1980; Knibb et al., 1981). For *D. robusta*, the effects of altitude on inversion frequencies sometimes recapitulate latitudinal clines in inversion frequency, depending upon where in the species' range the altitudinal cline is located. Levitan (1978) studied X chromosome inversion association clines in populations of *D. robusta* along altitudinal transects throughout the eastern United States. Inversions on opposite arms of the X chromosome in this species are in linkage disequilibrium, forming inversion associations in most populations studied (Levitan, 1958, 1973a). These associations show trends with altitude that vary geographically. That is, while particular inversions on the X chromosome show consistent trends with altitude, different associations of inversions are associated with changes in elevation in different parts of the species range. Thus, the pattern of local adaptation in *D. robusta* may involve a hierarchy of single inversions and linked inversion complexes within and between populations.

Populations of *D. robusta* inhabiting

the Great Smoky Mountains National Park exhibit altitudinal gradients in inversion frequencies over 30 km and 3,800 ft in elevation (Stalker and Carson, 1948; Levitan, 1978; this study). These populations harbor 14 paracentric inversions distributed over all chromosomes with the exception of the left arm of the third chromosome; 11 of these gene arrangements are associated with changes in elevation. All the X linked inversions of this species are in linkage disequilibrium in the Smoky Mountains (Levitan, 1973a) and show nonrandom frequencies of X chromosomes characteristic of southern Appalachian populations (Levitan, 1978).

The first purpose of this paper is to present inversion frequency data from a resampling of the altitudinal transect previously studied by Stalker and Carson (1948, collections made in late July, 1947) and Levitan (1978, collections made in August, 1958 and 1959). A description of genetic change spanning the last 34 years is thus achieved. The second purpose of this study is to present data describing the inversion clines in greater detail than before, through sampling over shorter altitudinal intervals, and through extension of the sampling to higher elevations (>4,000 ft). The current study was planned in an effort to elucidate the genetic continuities or discontinuities actually present in these clines.

#### MATERIALS AND METHODS

Flies were collected along a transect adjacent to the West Prong of the Little Pigeon River and associated feeder streams and along the Little River from

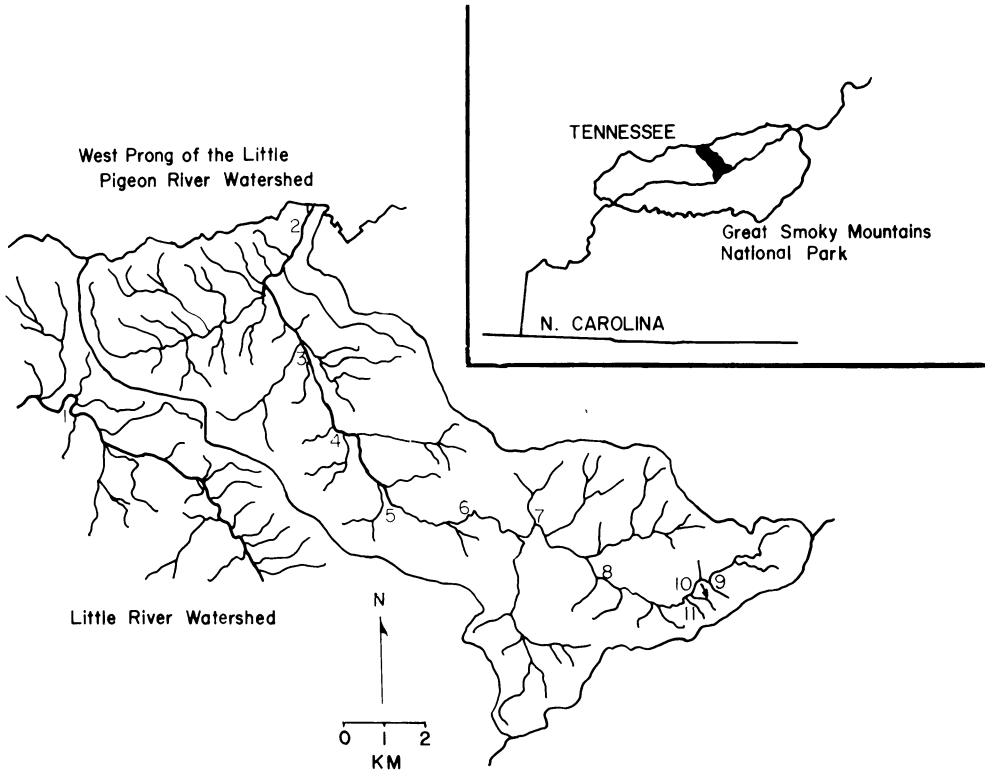


FIG. 1. The geography and locations of the two watersheds in the Smoky Mountains National Park indicated in black (inset) and the numbered collecting sites described in Table 1 where flies were obtained in this study. All streams in the area are indicated and the approximate borders of the watersheds were drawn from topographic maps.

elevations of approximately 1,360 ft to 4,840 ft (Fig. 1). Collections were made from July 23 to August 3, 1981 by sweep-netting flies daily over cans of fermenting bananas. Collecting dates were chosen to minimize seasonal bias in inversion frequencies (Levitan, 1973a). Collecting sites were chosen near streams at elevations reckoned by topographic map information and an altimeter (barometer) so that the populations sampled within the park boundaries were at locations at elevations reported by Stalker and Carson (1948) and Levitan (1978). Elevations of all sites are presented in feet (') for consistency with these earlier studies. Populations at 1,000' and 1,200' were not sampled because of destruction of appropriate habitats outside the park. The pop-

ulation at 1,360' was located approximately 2 km south of the site noted in Stalker and Carson (1948). All traps were tied to the branches or trunks of deciduous trees next to the West Prong of the Little Pigeon River unless otherwise noted in Table 1. The vegetation at these sites is described in Stalker and Carson (1948) and their depiction of the 4,000' elevation site generally applies to all sites above that elevation.

Adult male flies collected at each sampling location were immediately mated to virgin females of a known karyotype, and all wild females were despermed by repeated transfers to fresh food vials until no offspring were observed for at least two transfers. These females were then mated to stock males of the known

TABLE 1. Names, elevations, number of site collecting visits and locations of the collecting sites mentioned in this study.

Site	Elevation	<i>d</i> <sup>1</sup>	Location
1. Elkmont	2,080'	3	about 1 km south along the Little River from U.S. 73
2. Boundary	1,360'	2	about .7 km south of park boundary near Gatlinburg
3. Sugarlands	1,560'	5	near confluence of Sugarland Branch and WPLPR <sup>2</sup>
4. Critter Branch	2,000'	1	about .3 km north of the confluence of Critter Branch and WPLPR
5. Ramp Creek	2,440'	1	near confluence of Ramp Creek and WPLPR
6. Chimneys	3,040'	1	about .7 km north of Chimneys Picnic Area opposite roadside walkway
7. Cole Creek	3,620'	4	along Cole Creek south of loop on U.S. 441
8. Walker Camp I	3,980'	5	about .8 km south of Alum Cave parking lot
9. Walker Camp II	4,520'	4	east along Walker Prong near 4,500' park sign
10. Morton I	4,680'	7	south of parking area on U.S. 441
11. Morton II	4,840'	2	along small drainage southwest of tunnel on U.S. 441

<sup>1</sup> Number of collecting visits to site in days.<sup>2</sup> West Prong of the Little Pigeon River.

karyotype. The salivary glands of at least seven and usually ten  $F_1$  larvae from these crosses were dissected in insect saline and stained in lacto-acetoorcein to obtain inferred karyotypes for each of the wild adults (Levitan, 1958).

Inversion and chromosome designations are the same as those described in Carson and Stalker (1947), Carson (1958) and Levitan (1958, 1978). Single X chromosome and autosome inversion labels designate the chromosome, the arm of the chromosome on which the inversion is located, and an arbitrary number for a particular inversion (or no number designating the assigned standard arrangement). For example, 2L-3 refers to the number three inversion on the left arm of the second chromosome. X chromosome inversion associations are labelled similarly, i.e., a male whose X chromosome karyotype is inferred to be XL-1.XR, with a period denoting the centromere, can be represented in shorthand form as 1S where S refers to the standard arrangement XR (Carson, 1953, 1958; Levitan 1958).

All gene arrangements and arrangement combinations were tested for homogeneity across elevation by contingen-

cy table analyses; correlation and regression analyses were performed to quantify the clines after arcsin transformation of the frequency data. All populations were tested for fit to Hardy-Weinberg expectations for each inversion system. Tests for linkage disequilibria between all possible pairs of inversions on all chromosomes were performed according to the methods of Weir (1979). Between arm X chromosome data were treated separately because males are hemizygous for the X; therefore, the disequilibrium measures of Lewontin and Kojima (1960) were used.

An affinity analysis of the populations along the transect was created using the analysis of Istock and Scheiner (unpubl.). All X chromosome inversion association and autosomal inversion frequency data (Tables 2, 3) were logit transformed (Snedecor and Cochran, 1974) and used to compute a table of all possible pairwise similarity coefficients between populations using the analogue of the Jaccard similarity coefficient (Goodall, 1973). The similarity coefficient of population *i* with each of the other populations,  $S_{ix}$ , is computed and plotted against  $S_{jx}$ , the similarity coefficient of population *j* with each

of the other populations. Here the reference populations  $i$  and  $j$  are at opposite ends of the transect.  $\bar{S}_i$  is the mean similarity coefficient of population  $i$  with all the other populations. From the similarity matrix, a new matrix of Wilcoxon T values (Snedecor and Cochran, 1974) for all pairwise comparisons among sites is computed.  $T_{ij}$  is the Wilcoxon T value calculated for the  $S_{i\cdot}$  and  $S_{j\cdot}$  columns. The sign of the Wilcoxon T is retained. The T matrix is obtained by computing all possible pairs of  $S_{i\cdot}$  and  $S_{j\cdot}$ . The signed Wilcoxon T value is then averaged algebraically for each site giving a mean  $\bar{T}_i$  value against which is plotted  $\bar{S}_i$ . The ranks of the paired similarity values of any two populations with all the other populations in the Wilcoxon T procedure are expected, under the null hypothesis, to show no consistent positive or negative trends. The amount of sharing or matching of inversion frequencies by all populations simultaneously is characterized by  $\bar{T}_i$ . The S-T plot thus portrays an overall pattern of genetic similarity among all populations simultaneously while ordering the populations along  $\bar{T}_i$ , reflecting the tendency for each population to be more or less modal with respect to inversion and inversion association frequencies. Populations which contain increasingly distinctive constellations of inversions distinct from all the other populations have higher positive  $\bar{T}_i$  values.

#### RESULTS

Sites above 3,000' were necessarily sampled more often as fewer *D. robusta* were attracted to the baits each day than at lower elevations (Table 1). Baits were exposed continuously until sufficient numbers of *D. robusta* were obtained; my goal was 50 males and 50 females per site.

X chromosome frequencies and second and third chromosome inversion frequencies are presented in Tables 2 and 3, respectively. All data for the sites at 4,680' and 4,840' were pooled after verifying frequency homogeneity for all in-

versions due to the small sample size at 4,840'. The populations sampled in 1981 showed no statistically significant deviations from Hardy-Weinberg expectations for each inversion system.

Considering the 1981 survey first, inspection of Tables 2 and 4 reveals changes in frequency for different X chromosome combinations with altitude. Frequencies of 1S and 12 combinations increased with elevation, with 1S dominating in frequency above 1,560' (Fig. 2). Inversion combination 1S increased in frequency up to 2,000', then decreased until 4,680'. The combinations S2 and 22 decreased in frequency with altitude, combination S2 showing a significant negative regression with elevation (Table 4). Both S2 and 22 characterized the lower elevation populations. Arrangement 12 accounted for 42% of the X chromosomes at 4,520', shifting from 24% at 3,980', and then dropping to 9% at 4,680'. Overall, XL-containing combinations were associated with lower elevations while XL-1-containing associations predominated above 1,560'.

The left arm of chromosome two was found to contain four gene arrangements in all but the 4,840' population (Table 3). Frequencies of gene arrangements 2L-1 and 2L-3 showed significant negative and positive regressions with elevation respectively, while the negative trend for 2L bordered on significance (Table 4). Arrangement 2L-3 clearly predominated at all elevations, rising to 85% at the 4,680' site; 2L and 2L-1 were more common at lower elevations (Fig. 3). Inversions in the right arms of chromosomes 2 and 3 showed no detectable change with elevation.

Strong X chromosome linkage disequilibria have persisted since 1947 in all populations studied due to an overabundance of S2 and 1S arrangement combinations and a corresponding deficiency of SS and 12 (Levitan, 1978). X chromosome association 22 is seldom broken down by recombination (Nelson, 1951; Levitan, 1958); only four 2S chromosomes have ever been recovered in

TABLE 2. X-chromosome frequencies (percent) in populations of *Drosophila robusta* along an altitudinal transect in the Great Smoky Mountains National Park, in the collections of Stalker and Carson (1948), Levitan (1978), and this study, respectively.

Year	Elevation	N <sup>1</sup>	SS	S2	1S	12	22	χ <sup>2</sup> @
1947	1,000'	53	47.2	37.7	.0	7.5	7.5	
1947	1,200'	78†	33.3	25.6	6.4	7.7	26.9	
1947	1,400'	95†	11.6	34.8	10.5	18.9	24.2	
1958-1959	1,400'	51	7.8	29.4	7.8	15.7	39.2	12.15 NS
1981	1,360'	147	7.0	32.0	16.0	25.0	20.0	
1981	1,560'	123	3.0	29.0	30.0	22.0	16.0	
1947	2,000'	258††	7.0	25.9	13.9	17.1	36.1	
1958-1959	2,000'	798	6.6	21.8	25.4	22.1	24.1	100.42**
1981	2,000'	112	3.0	12.0	60.0	16.0	9.0	
1981	2,080'†††	149	3.0	19.0	43.0	24.0	11.0	
1981	2,440'	127	2.0	12.0	54.0	23.0	9.0	
1947	3,000'	40	.0	15.0	27.5	17.5	40.0	
1958-1959	3,000'	58†	1.7	10.3	51.7	13.8	22.4	15.35 NS
1981	3,040'	33	.0	9.0	55.0	27.0	9.0	
1981	3,620'	87	1.0	10.0	50.0	31.0	8.0	
1947	4,000'	34	.0	14.7	8.8	26.5	50.0	
1981	3,980'	97	4.0	13.0	47.0	24.0	12.0	33.28**
1981	4,520'	99	1.0	14.0	37.0	42.0	3.0	
1981	4,680'	55	4.0	7.0	69.0	9.0	11.0	
1981	4,840'	10	10.0	10.0	60.0	10.0	10.0	

<sup>1</sup> N refers to the number of chromosomes sampled.

† Includes one XL-2.XR not included in the calculations.

†† Includes three XL-2.XR not included in the calculations.

††† The sample from Elkmont, not included in the Chi-square test.

\*\* P < .01.

NS Not significant.

@ Chi-squared values refer to the test for homogeneity between years. The test for homogeneity between elevations for 1981 yielded a χ<sup>2</sup> = 147.19\*\*.

nature (Levitan, 1973a), therefore only XR-2 when not linked to XL-2 chromosomes, i.e., S2 and 12, were included in the analysis. The direction of linkage disequilibrium, i.e., too many 1S and S2 combinations, is typical for the southern Appalachians and differs from the northeastern U.S. where SS and 12 combinations are in excess (Levitan, 1978). Significance of the consistency of linkage disequilibria in all populations can be measured by the χ test (Simpson et al., 1960). Values for  $\chi(\chi = \sqrt{\chi^2})$  are obtained for each population and assigned a negative value if there is an excess of "repulsion" combinations (S2 and 1S) or a positive value if there is an excess of "coupling" combinations (SS and 12). The sum of the χ values divided by the square root of the sum of the degrees of

freedom yields a standard normal deviate. At all elevations, most notably at 1,560'-2440', there were significant excesses of 1S and S2, and deficiencies of SS and 12 (Table 5) which have persisted since 1947 (Levitan, 1978).

No significant disequilibria were detected within the second chromosome or between all pairwise combinations of inversions on different chromosomes. When X linked inversions were considered with autosomal inversions the analysis was performed only on females since Weir's (1979) composite estimator of linkage disequilibrium includes measures of single locus disequilibria (homozygote excess) which are undefined for the hemizygous males. Pooling adjacent populations homogeneous for all inversion frequencies to increase female sam-

TABLE 3. Second and third chromosome gene arrangement frequencies in populations of *D. robusta* in the Great Smoky Mountains National Park from the collections of Stalker and Carson (1948) and this study.<sup>1</sup>

Year	Elevation	N <sup>1</sup>	Second chromosome					Third chromosome				
			2L	2L-1	2L-2	2L-3	$\chi^2$	2R	2R-1	3R	3R-1	$\chi^2$
1947	1,000'	145	26.9	37.3	17.2	18.6		86.9	13.1	33.8	66.2	
1947	1,200'	127	24.4	33.9	15.7	26.0		89.8	10.2	28.3	71.7	
1947	1,400'	278	19.4	19.8	15.5	45.3		91.7	8.3	25.2	74.8	
1981	1,360'	186	17.2	21.5	5.9	55.4	11.48**	91.9	8.1	30.6	69.4	1.67 NS
1981	1,560'	158	12.6	18.5	7.5	61.4		95.6	4.4	31.1	68.9	
1947	2,000'	301	18.6	23.6	7.0	50.8		88.7	11.3	19.9	80.1	
1981	2,000'	168	11.3	17.9	5.3	65.5	9.80*	93.4	6.6	25.6	74.4	2.01 NS
1981	2,080'	200	13.0	12.0	8.0	67.0		92.5	7.5	25.0	75.0	
1981	2,440'	168	7.7	16.0	6.6	69.7		94.6	5.4	32.7	67.7	
1947	3,000'	102	22.5	15.7	7.8	54.0		85.3	14.7	22.5	77.5	
1981	3,040'	52	5.8	11.5	5.8	76.9	9.32*	88.5	11.5	30.8	69.2	1.23 NS
1981	3,620'	122	7.4	13.9	3.3	75.4		94.3	5.7	20.5	79.5	
1947	4,000'	74	20.3	8.1	5.4	66.2		91.9	8.1	21.6	78.4	
1981	3,980'	140	11.4	8.6	4.3	75.7	3.36 NS	89.3	10.7	25.0	75.0	0.31 NS
1981	4,520'	144	11.8	5.6	6.9	75.7		95.1	4.9	26.4	73.6	
1981	4,680'	82	1.2	8.5	4.9	85.4		92.7	7.3	30.5	69.5	
1981	4,840'	14	7.0	21.0	.0	72.0		86.0	14.0	36.0	64.0	

<sup>1</sup> All footnotes as in Table 1, except the test for homogeneity across elevation in 1981 for the left arm of chromosome 2 produced a  $\chi^2 = 63.18^{**}$ .

\*  $P < .05$ .

ple sizes failed to detect additional linkage disequilibria within or between chromosomes.

Comparisons of the frequencies of X chromosomes and of the second and third chromosome inversions from the three collections made over the last 34 years indicate a significant genetic restructuring of the Smokies populations along this altitudinal transect (Figs. 2, 3, 4; Tables 2, 3). Increases in frequency of X chromosome combination 1S and of inversion 2L-3 at all elevations common to the 1947, 1958-1959, and 1981 studies have raised representations of high altitude gene arrangements and combinations of arrangements. The heterogeneity in 3R reported in the 1947 collection (Stalker and Carson, 1948) was not present in 1981 ( $\chi^2 = 9.7$ ,  $P > .4$ , Tables 3 and 4), but different sets of populations were sampled. The cline in arrangement combination 22, which was positively

correlated with altitude in 1947 (Table 4), showed evidence of a decrease in frequency at higher elevations in 1958-1959, and in 1981 showed a negative correlation with elevation. Thus, over a period of 34 years, the cline in combination 22 shows a statistically significant *reversal* with respect to altitude.

At 1,400' the X chromosome frequencies do not show a statistically significant change through the years, yet arrangement combinations 12 and 1S have increased in frequency, while SS and 22 have decreased (Table 2). The 22 combination increased in frequency from 1947 to 1958-1959, then by 1981 had declined in frequency to below its 1947 frequency. The statistically significant changes in the left arm of chromosome 2 are characterized by an increase in 2L-3 and a decrease in 2L-2. Frequencies of arrangements 2L and 2L-1 remained unchanged (Table 3).

TABLE 4. Correlation and linear regression coefficients of X chromosome inversion association and autosomal inversion effects of altitude in 12 populations of *Drosophila robusta* in the vicinity of the Great Smoky Mountains National Park in 1947 and 1981.

Inversion(s) <sup>1</sup>	Correlation (r) with altitude	Slope	F <sup>2</sup>
1. 1947—SS	-.788	-.077	6.54†
1981—SS	-.299	-.003	.79
2. 1947—S2	-.880*	-.043	13.79*
1981—S2	-.726*	-.029	8.91*
3. 1947—1S	.494	.023	1.29
1981—1S	.503	.040	2.71
4. 1947—12	.843	.030	9.86*
1981—12	.172	.007	.24
5. 1947—22	.908*	.069	18.90*
1981—22	-.599*	-.012	4.47†
6. 1947—2L	-.435	-.007	.94
1981—2L	-.623†	-.012	5.08†
7. 1947—2L-1	-.893*	-.049	15.67*
1981—2L-1	-.844*	-.020	19.76***
8. 1947—2L-2	-.885*	-.023	14.40*
1981—2L-2	-.608	-.005	4.69
9. 1947—2L-3	.902*	.088	17.39*
1981—2L-3	.921*	.051	44.79***
10. 1947—3R-1	.695	.026	3.71
1981—3R-1	.164	.003	.22

† The 12 populations are the 6 of Stalker and Carson (1948) and the 10 in this study.

<sup>1</sup> Entries 1-5 are the five X chromosome associations in different years and entries 6-10 are the major autosomal inversions.

<sup>2</sup> F is the variance ratio from the test of the hypothesis that the slope of the regression line equals zero.

† .1 > P > .05.

\* P < .05.

\*\* P < .01.

\*\*\* P < .005.

In 1981 the sample from Elkmont at 2,080' was collected for purposes of comparing that population with the 2,000' sample at Critter Branch (Tables 2 and 3), contrasting geographically isolated populations in different watersheds at similar elevations. Data for all chromosomes for these two populations are in close agreement, except for the predominance of combination 1S at Critter Branch (Table 3). Only the 4,680' population exceeds the Critter Branch population in the frequency of this arrangement combination. Combinations S2 and 22 have declined in frequency since 1947,

while 1S has dramatically increased (Fig. 2). Gene arrangements 2L, 2L-1, and 2L-2 have decreased slightly in frequency while 2L-3 has increased through the years (Fig. 3).

Interpretation of the data from 3,000' is hampered by small sample sizes, yet some change in all but the SS combination seems evident. Combinations 22 and S2 have decreased, 22 markedly so, while 1S and 12 have increased (Table 2, Fig. 2). Gene arrangement 2L-3 has increased at the expense of 2L and 2L-1 (Fig. 2).

At 4,000' X chromosome arrangement combination 1S has increased to 47% by 1981 from the 9% in 1947, while combination 22 has decreased from 50% to 12% (Table 2). In the same period second chromosome inversion frequencies have changed slightly, but not significantly with 2L-3 increasing and 2L decreasing in frequency (Fig. 3).

Closer inspection of the cline using the affinity analysis revealed a higher order genetic structuring of the Smokies populations in both the 1947 and 1981 collections (Fig. 5). Genetic similarity biplots, the genetic similarities of the lowest population with all others versus the highest population with all others,  $S_{ix}$  vs.  $S_{jx}$ , respectively, show an ordering of all populations with elevation (Fig. 5a, b). These reference axes were chosen a priori because they represent a biologically meaningful framework; i.e., endpoints of the transects. The 1947 collection is arrayed exactly in the correct altitudinal order as opposed to the 1981 collection which shows several inconsistencies (Fig. 5b). The 1947 biplot (Fig. 5a) corroborates Stalker and Carson's (1948) conclusions that the lower end of the transect (1,000'-1,400') encompassed more genetic diversification than the upper elevations shown by the separation of the 1,000' and 1,200' populations from the 1,400', 2,000' and 3,000' populations (Fig. 5a). The 1981 data show less overall genetic dissimilarity because the 1,000' and 1,200' sites were not included, but the pattern of similarities shows a less even pattern along the transect than the

TABLE 5. Estimates of X chromosome linkage disequilibrium,  $D$ , the standard deviation of  $D$ ,  $SD$ , and associated  $\chi^2$  and  $\chi$  values in 10 populations of *D. robusta* in the Great Smoky Mountains, 1981. Populations from 4,680' and 4,840' were pooled. The statistical tests are described in the text.

Elevation	$DU^1$	$D$	$SD$	$\chi^2$	$\chi$
1,360'	-.0486	-.0484	.0144	5.39*	-2.32
1,560'	-.1124	-.1119	.0137	22.86***	-4.78
2,000'	-.0871	-.0867	.0147	25.57***	-5.06
2,080'	-.0915	-.0911	.0214	23.33***	-4.82
2,440'	-.0759	-.0755	.0129	20.34***	-4.51
3,040'	-.0610	-.0600	.0216	5.00*	-3.24
3,620'	-.0566	-.0563	.0129	9.64***	-3.11
3,980'	-.0641	-.0637	.0154	9.32***	-3.05
4,520'	-.0537	-.0534	.0109	9.19***	-2.86
4,680'					
4,840'	-.0606	-.0600	.0190	11.42***	-3.38

$\chi = -36.11$   
 $t = -8.07***$

<sup>1</sup>  $DU$  refers to the unbiased estimate of  $D$ , given by Weir (1979) for the case of known linkage relationships with random mating.  
\*  $P < .05$ .  
\*\*\*  $P < .005$ .

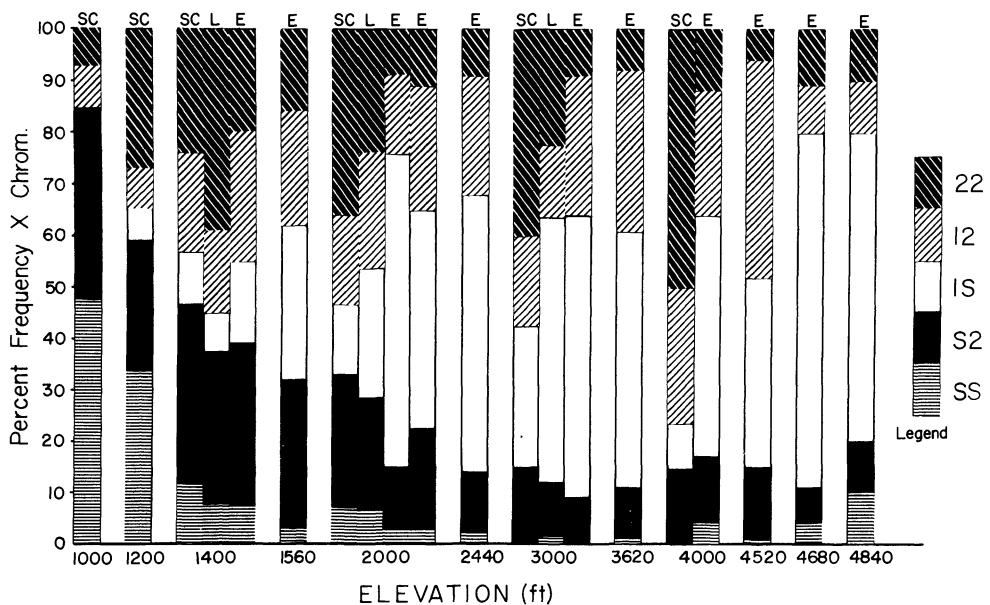


FIG. 2. X chromosome frequencies along the Smoky Mountains transect for 12 populations of *D. robusta* compared from samples obtained in 1947, 1958-1959, and 1981 by Stalker and Carson (1948), Levitan (1978), and the author. Bars in the diagram are separated when differences in elevation are compared; otherwise, adjacent bars imply the same site was sampled at that elevation in different years, except the rightmost bar at the 2,000' elevation which refers to the Elkmont site in the Little River watershed. SC, L, and E refer to the first letters of the authors' names of the three studies.



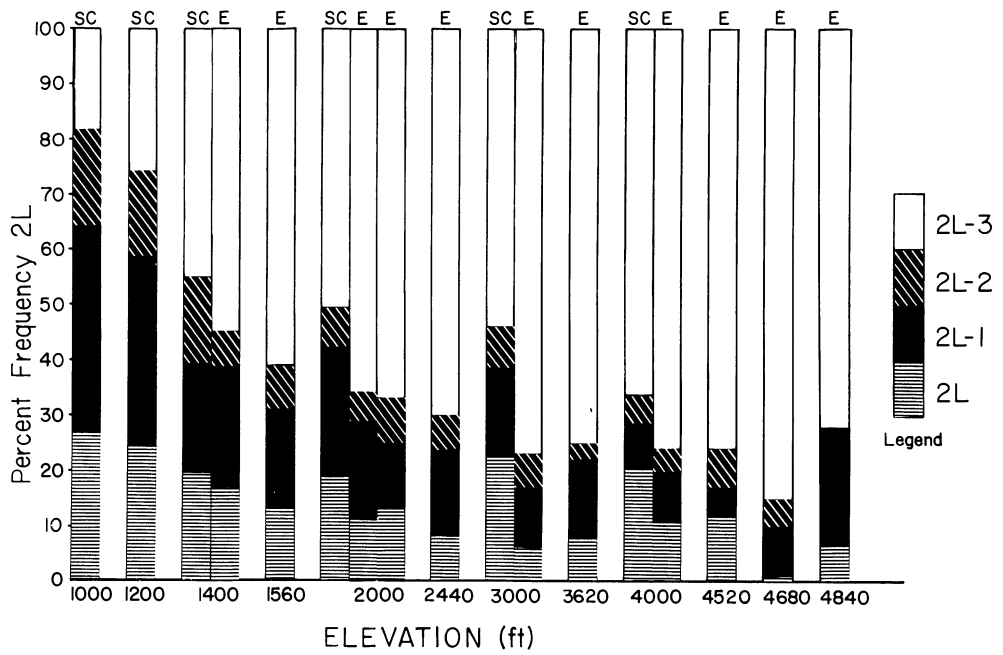


FIG. 3. Second chromosome inversion frequencies along the Smoky Mountains transect for 12 populations of *D. robusta*. Figure descriptions are the same as in Figure 2.

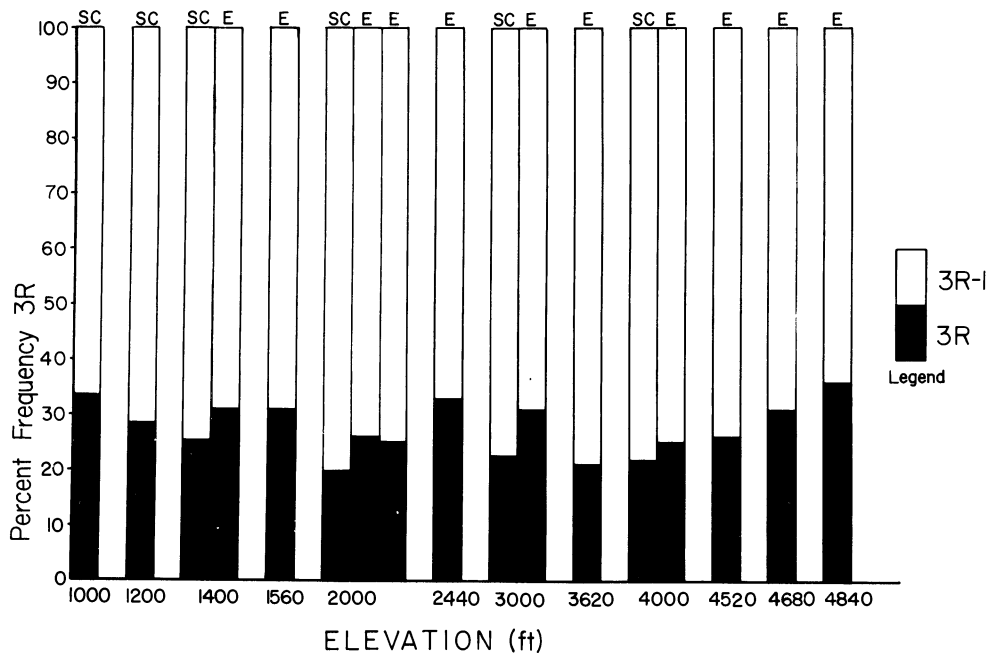


FIG. 4. Third chromosome inversion frequencies in 12 populations of *D. robusta* along the Smoky Mountains transect. Descriptions of the figure are the same as in Figure 2.

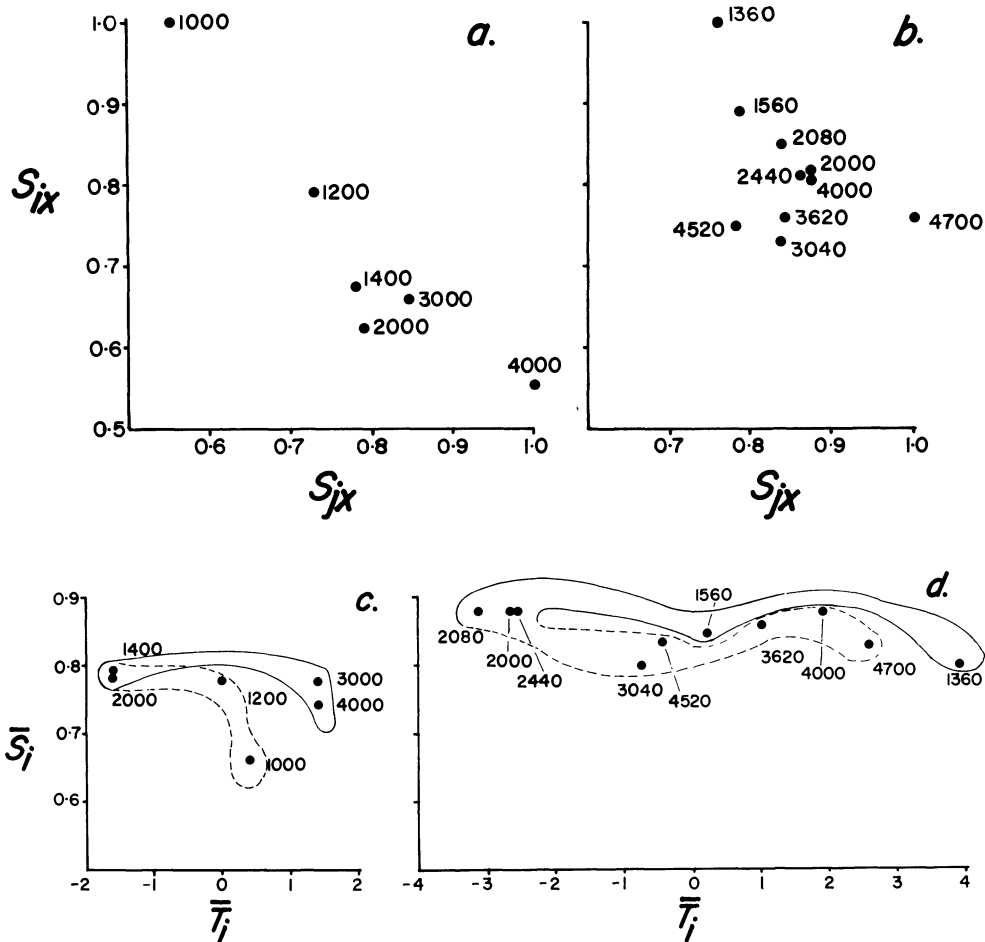


FIG. 5. a) Similarity biplot of the 1947 data of Stalker and Carson (1948) showing the ordering of the populations along the altitudinal transect. b) IBID for the 1981 data. c) S-T plot of the 1947 data showing the higher order genetic structure of the Smokies populations. d) IBID for the 1981 data.  $S_{ix}$  and  $S_{jx}$  are the Jaccard similarity coefficients of the highest and lowest transect populations, respectively with all of the other populations.  $\bar{S}_i$  is the mean genetic similarity coefficient of population  $i$  with all of the other populations.  $\bar{T}_i$  is the mean Wilcoxon T value, calculated from the genetic similarity matrix, derived from pairwise comparisons of the similarity values of any two sites with all of the others. The "arms" in c) and d) were inferred directly from a) and b), and were drawn to show the apparent genetic uniqueness at each end of the transect in terms of  $\bar{T}_i$ . Details of the techniques and their interpretation are given in the text. Data for the 4,680' and 4,840' were pooled for the affinity analysis.

1947 data, evidenced by the placements of the 4,520', 3,620' and 3,040' sites toward the origin of the biplot (Fig. 5b).

The S-T plots (Fig. 5c, d) reveal a genetical aspect of the Smokies populations not readily seen in any of the frequency data. At a glance, these plots show similar

patterns in both the 1947 and 1981 data sets where the middle elevation populations, the 1,400' and 2,000' sites in 1947 and the 2,000', 2,080', and 2,440' sites in 1981 are on average modal in frequency for all inversions and inversion associations. That is, these populations

better match all the other populations simultaneously in terms of inversion and inversion association frequencies. Populations with higher associated  $\bar{T}_i$  values contain more distinct combinations of frequencies due to increasing genetic uniqueness. In both S-T plots (Fig. 5c, d), higher and lower elevation populations are characterized by decreased sharing of similarity in frequency, diagrammed by the two "arms" centered on the middle elevation populations. In spite of the temporal changes in inversion and inversion association frequencies, the middle elevation populations have retained an overall genetic modality and evenness of inversion representation with respect to all the other populations. The upper elevation populations (>2,500') show more heterogeneity amongst themselves than would be expected from a smooth cline with increasing genetic distinctness with elevation.

#### DISCUSSION

Long term genetic changes in natural populations of *D. robusta* have been hitherto undocumented. Results from years of intensive study of several populations suggested a high degree of inversion frequency homeostasis for this species (Carson, 1958). Data from 10 years of sampling a population at Olivette Woods, Missouri showed no evidence of any directional changes (Carson, 1958). It is unclear why the Smoky Mountains populations should exhibit temporal genetic changes over many years, as well as seasonal changes in inversion frequencies as shown in previous studies (Levitan, 1973a), while the Missouri populations exhibit few, if any, temporal or seasonal changes (Carson, 1958).

The long term temporal changes in the genetic structuring of populations of *D. robusta* along the Smokies transect have implications which may extend to other species, especially *D. pseudoobscura*. In the most recent of eight publications devoted to this problem, Anderson et al. (1975) stated: "The search for an under-

standing of the long term changes in inversion frequency of the inversion polymorphs of *D. pseudoobscura* serves as a paradigm for evolutionary studies. After 3 decades of study the problem remains unsolved." Similarly, the specific causes for the temporal shifts in inversion and inversion combination frequencies in populations of *D. robusta* along the Smokies transect are presently unknown. The 34-year increase in frequency of 2L-3, arrangement combination XL-1.XR and decrease in XL-2.XR-2 in every one of the transect populations implies that all of the populations have assumed more of a high elevation genotype. Arrangement combination 22 has been reported in high frequency in the central-southern United States (Carson and Stalker, 1947) yet was found to increase in frequency with elevation in 1947 (Stalker and Carson, 1948), showing a parallel cline "reversal" with 3R-1, a southern gene arrangement, which also increased with elevation (although not significantly, see Table 4). Furthermore, in altitudinal transects where XL-1.XR increases with elevation, XL-2.XR-2 usually decreases (Levitan, 1978).

Such pervasive and systematic temporal genetic changes in adjacent populations demands explanation but interpretations are made difficult by the lack of information on dispersal ability and knowledge of the spatial and temporal availability of food resources (cf. Carson and Stalker, 1951). However, considering the affinity analysis of the transect populations (Fig. 5) the relative altitudinal order of many of the populations has remained stable for 34 years, in spite of the systematic frequency changes of individual genetic elements. The causes for these changes may be due to any one or combination of factors including habitat alteration, genetic changes within the inversions themselves, migration, between year seasonal variations, genetic drift, or sampling error. Sampling error cannot be ruled out, but is unlikely because the sample sizes at most sites are considerable in all three studies (Tables

2, 3), the temporal variations are in some cases quite high, e.g., combination 1S at 2,000', and the direction of change is consistent in all the populations. Mass migration from the upper elevations could account for the overall increases in 2L-3 and combination 1S, but not the cline reversal in combination 22. Because the population sizes increase throughout the summer (Carson, 1958) and the growing season shortens with elevation (Shanks, 1954), population sizes can be expected to increase during the summer from the lower to the upper elevations. Therefore, mass migration should proceed from the bottom to the top of the mountain, if at all. Year to year seasonal variation could also produce frequency shifts, since the inversions in *D. robusta* are known to cycle seasonally (Levitan, 1973a), but this does not explain the systematic directional changes in almost all the populations. Genetic changes within the inversions themselves, altering the fitness relationships of alternate gene arrangements and combinations cannot be excluded as a possibility for the long term frequency changes (cf. Anderson et al., 1975). Historically, the most important environmental alterations in the vicinity of the transect populations were logging activity and farming which continued until the early 1930's (Stupka, 1964; S. P. Bratton, pers. comm.). Land-clearing for farming progressed up to approximately 3,000', above which is predominantly virgin forest. Further yearly sampling of the transect populations is clearly warranted to determine the rate of change of inversion frequencies. Results from a survey in late July, 1982 of these populations were consistent with the 1981 data (Etges, unpubl.).

The degree of genetic differentiation between adjacent populations depends upon a balance between gene flow, drift and selection, a subject of many theoretical (Wright, 1943, 1946; Haldane, 1948; Kimura and Weiss, 1964; Rohlf and Schnell, 1971; Slatkin, 1973; May et al., 1975; Endler, 1977) and empirical studies (Carson and Stalker, 1947; Dob-

zhansky, 1948; Stalker and Carson, 1948; Cain and Currey, 1963; Kettlewell and Berry, 1969; Antonovics and Bradshaw, 1970; Endler, 1973; Levitan, 1978; Knibb et al., 1981). Most theoretical treatments have emphasized the interactions between limited gene flow in subdivided populations and genetic drift in the establishment of genetic differentiation over short distances. Rohlf and Schnell (1971) and Endler (1973) have shown in simulations that limited gene flow between subdivided populations could cause genetic differentiation that once established would persist for hundreds of generations over small distances. Endler (1977) argued that steps in continuous clines may be produced by limited gene flow without selection or steep environmental gradients.

Empirical studies of genetic clines, particularly in *Drosophila* species, have usually assumed natural selection as the primary cause for genetic differentiation (however, see Endler, 1973). The observation by Stalker and Carson (1948) that the altitudinal gradient of inversions in *Drosophila robusta* strongly recapitulates the north-south latitudinal clines (Carson and Stalker, 1947) provides evidence that particular inversions are under selection, especially XL-1 and 2L-3, which are common to northern populations (Carson, 1958). Furthermore, Levitan (1978) showed that two types of XL-1 bearing chromosome associations found throughout the eastern part of the species range, i.e., 1S and 12, show local variation associated with elevation. In northern areas, combination 12 is associated with higher elevations where 1S is not. In the southern part of the species range, including the Smoky Mountains transect, combination 1S shows increases in frequency with elevation. Combination 12 was evenly distributed in all populations in 1981, except at 4,520'. This observation that XL-1, albeit in a complex way, is associated with altitude in widely separated populations is prima facie evidence for natural selection. Crossing-over between inversions on opposite arms of

the X chromosome is infrequent—approximately  $10^{-3}$  per gamete per generation (Levitan, 1958); therefore the dynamics of individual gene arrangements on the X chromosome, due to the degree of linkage, cannot be properly evaluated in natural populations without first determining their linkage relationships. The observations that one of the linkage associations, XL-2.XR, is exceedingly rare, and that linkage disequilibrium is strong and consistent in direction suggests interaction between linked inversions on the X chromosome (Levitan, 1973*b*, 1978).

The ongoing search for causation underlying the long term genetic changes in natural populations of *Drosophila* has provided little understanding of how inversions interact under *natural* conditions to produce fitness differences contributing to the observed temporal changes in inversion frequencies. For *D. pseudoobscura* the search for causation has taken 40 years with few tangible results. The Smokies populations of *D. robusta* provide another model system for long term studies of inversion polymorphism which may be more amenable to conclusive analyses due to the proximity of populations which have all undergone systematic genetic changes with time and because the populations are located in an area with a known history of environmental alteration. The present study indicates the need for discovering the actual forces of selection here and the estimation of dispersal abilities and population sizes at different elevations.

#### SUMMARY

Populations of *Drosophila robusta* inhabiting the Great Smoky Mountains exhibit clines in inversion and inversion association frequencies over an altitudinal transect spanning 3,800' in elevation. Collections made in the summer of 1981 showed systematic directional shifts in inversion and inversion association frequencies in all populations, when compared to the earlier collections in 1947 (Stalker and Carson, 1948) and in 1958–

1959 (Levitan, 1978), toward higher elevation genotypes in the intervening 34 years. A cline in X chromosome inversion association XL-2.XR-2 shifted from a statistically significant *positive* relationship with altitude to a statistically significant *negative* relationship over that time period.

Linkage disequilibria between inversions on opposite arms of the X chromosome is strong and shows patterns characteristic of the Southern Appalachians. In all populations studied, there is an excess of "repulsion" combinations 1S and S2 with a corresponding deficiency of combinations SS and 12. Linkage disequilibrium of inversions on opposite arms of the second chromosome or between arms on different chromosomes was not detected.

A genetic ordination of the transect populations showed for both the 1947 and 1981 data sets that the middle elevation populations contained modal frequencies of all genetic elements surveyed. The populations at lower and higher elevations were more genetically distinct from each other and the middle elevation populations, and were more extreme in combined inversion and inversion association frequencies than the middle elevation populations.

#### ACKNOWLEDGMENTS

I thank C. A. Istock for the use of the affinity analysis computer program and who with T. Caraco, U. Nur, E. W. Caspari, R. J. MacIntyre, and an anonymous reviewer made useful criticisms of earlier drafts. Peter White at Uplands Field Research Lab provided work space in the park. Computing costs were provided by the Laboratory for Laser Energetics and the University of Rochester. D. Foltz provided the linkage program and help in its implementation. J. Ricciardi and E. S. Etges provided technical assistance, M. Grassi typed the manuscripts, and Dr. Max Levitan provided stocks, karyotyping, and encouragement through all phases of this study. The author was sup-

ported by NIH Predoctoral Genetics Training Grant 5-27667-169.

## LITERATURE CITED

- ANDERSON, W. W., TH. DOBZHANSKY, O. PAVLOVSKY, J. R. POWELL, AND D. YARDLEY. 1975. Genetics of Natural Populations. XLII. Three decades of genetic change in *Drosophila pseudoobscura*. *Evolution* 29:24-36.
- ANTONOVICS, J., AND A. D. BRADSHAW. 1970. Evolution in closely adjacent populations. VIII. Clinal patterns at a mine boundary. *Heredity* 25:349-362.
- CAIN, A. J., AND J. D. CURREY. 1963. Area effects in *Cepea*. *Phil. Tran. Roy. Soc. London B*:1-81.
- CARSON, H. L. 1953. The effects of inversions on crossing over in *Drosophila robusta*. *Genetics* 38:168-186.
- . 1958. The population genetics of *Drosophila robusta*. *Adv. Genet.* 9:1-40.
- CARSON, H. L., AND H. D. STALKER. 1947. Gene arrangements in natural populations of *Drosophila robusta*. *Evolution* 1:113-133.
- . 1951. Natural breeding sites for some wild species of *Drosophila* in the eastern United States. *Ecology* 32:317-330.
- DOBZHANSKY, TH. 1948. Genetics of natural populations. XVI. Altitudinal and seasonal changes produced by natural selection in natural populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. *Genetics* 33:158-176.
- ENDLER, J. A. 1973. Gene flow and population differentiation. *Science* 179:243-250.
- . 1977. Geographic Variation, Speciation and Clines. Princeton Univ. Press, Princeton.
- GOODALL, D. W. 1973. Sample similarity and species correlation, p. 106-156. *In* R. H. Whittaker (ed.), *Ordination and Classification of Communities*. W. Junk, The Hague.
- HALDANE, J. B. S. 1948. The theory of a cline. *J. Genet.* 48:277-284.
- KETTLEWELL, H. B. D., AND R. J. BERRY. 1969. Gene flow in a cline. *Heredity* 24:1-14.
- KIMURA, M., AND G. H. WEISS. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49:561-576.
- KNIBB, W. R., J. G. OAKESHOTT, AND J. B. GIBSON. 1981. Chromosome inversion polymorphism in *Drosophila melanogaster*. I. Latitudinal clines and associations between inversions in Australasian populations. *Genetics* 98:833-847.
- LEVITAN, M. 1958. Nonrandom association of inversions. *Cold Spr. Harb. Symp. Quant. Biol.* 23:251-268.
- . 1973a. Studies of linkage in populations. VI. Periodic selection for X-chromosome arrangement combinations. *Evolution* 27:215-225.
- . 1973b. Studies of linkage in populations. VII. Temporal variation and X-chromosome linkage disequilibria. *Evolution* 27:476-485.
- . 1978. Studies of linkage in populations. IX. The effects of altitude on X-chromosome arrangement combinations in *Drosophila robusta*. *Genetics* 89:751-763.
- LEWONTIN, R. C., AND K. KOJIMA. 1960. The evolutionary dynamics of complex polymorphisms. *Evolution* 14:458-472.
- MAY, R. M., J. A. ENDLER, AND R. E. MCMURTRIE. 1975. Gene frequency clines in the presence of selection opposed by gene flow. *Amer. Natur.* 109:659-676.
- METTLER, L. E., R. A. VOELKER, AND T. MUKAI. 1977. Inversion clines in populations of *Drosophila melanogaster*. *Genetics* 87:169-176.
- NELSON, S. E. 1951. The influence of crossing over in the X chromosome of *Drosophila robusta*. Master's thesis. Washington Univ., St. Louis.
- ROHLF, F. J., AND G. D. SCHNELL. 1971. An investigation of the isolation by distance model. *Amer. Natur.* 105:295-324.
- SHANKS, R. E. 1954. Climates of the Great Smoky Mountains. *Ecology* 35:354-361.
- SIMPSON, G. G., A. ROE, AND R. C. LEWONTIN. 1960. *Quantitative Zoology*, Rev. ed. Harcourt, Brace and World, N.Y.
- SLATKIN, M. 1973. Gene flow and selection in a cline. *Genetics* 75:733-756.
- SNEDECOR, G. W., AND W. G. COCHRAN. 1974. *Statistical Methods*. Iowa State Univ. Press, Ames.
- SPIESS, E. B. 1950. Experimental populations of *Drosophila persimilis* from an altitudinal transect in the Sierra Nevada. *Evolution* 14:14-33.
- STALKER, H. D. 1980. Chromosome studies in natural populations of *Drosophila melanogaster*. II. Relationship of inversion frequencies to latitude, season, wing loading, and flight activity. *Genetics* 95:221-233.
- STALKER, H. D., AND H. L. CARSON. 1948. An altitudinal transect of *Drosophila robusta* Sturtevant. *Evolution* 2:295-305.
- STUPKA, A. 1964. *Trees, Shrubs, and Woody Vines of Great Smoky Mountains National Park*. Univ. Tenn. Press, Knoxville.
- WARD, B. L., W. T. STARMER, J. S. RUSSELL, AND W. B. HEED. 1974. The correlation of climate and host plant morphology with a geographic gradient of an inversion polymorphism in *Drosophila pachea*. *Evolution* 28:565-579.
- WEIR, B. 1979. Inferences about linkage disequilibrium. *Biometrics* 35:235-254.
- WRIGHT, S. 1943. Isolation by distance. *Genetics* 28:114-138.
- . 1946. Isolation by distance under diverse systems of mating. *Genetics* 31:39-59.

Corresponding Editor: R. J. MacIntyre