Biological Journal of the Linnean Society, 2004, **81**, 395–411. With 6 figures

Palaeoclimatic variation, adaptation and biogeography of inversion polymorphisms in natural populations of *Drosophila robusta*

WILLIAM J. ETGES[1]* and MAX LEVITAN[2]

[1]*Department of Biological Sciences, SCEN 632, University of Arkansas, Fayetteville, AR 72701, USA*
[2]*Center for Anatomy and Functional Morphology and Department of Human Genetics, Mount Sinai School of Medicine, Box 1007, NY 10029, USA*

Received 20 March 2003; accepted for publication 11 August 2003

Studies of natural and sexual selection in wild populations of *Drosophila* have historically provided strong inference for the maintenance of inversion polymorphism. Analysis of geographical variation in the *Drosophila robusta* chromosomal data collected over more than 50 years from 133 natural populations across eastern North America has confirmed several north–south and east–west clines in the frequencies of some gene arrangements and linked arrangement combinations. Patterns of geographical variation, including several north–south clines, revealed by regression and spatial autocorrelation analyses are concordant with palaeoclimatic shifts, Pleistocene glaciations and historical changes in the composition of North American forest communities. Because *D. robusta* is a sap-breeder, using the microbe-infested sap exudates of a number of deciduous tree species in which they carry out their life cycle, shifts in climate and palaeovegetation types since the formation of the eastern deciduous forests in the Miocene are hypothesized to be major factors influencing patterns of inversion polymorphisms across the range of this drosophilid species. In areas where sharp deviations in frequencies have been observed, particularly in the mid-western and western portions of the range, these divisions parallel historical geographical disjunctions in the species range that have yet to promote divergence and species formation despite the long history of *D. robusta* in North America.

ADDITIONAL KEYWORDS: clines – deciduous forests – Diptera – evolution – linked inversions – Miocene – natural selection – North America – palaeoecology.

INTRODUCTION

A major challenge to studies of geographical variation in natural populations is untangling historical forces from local adaptation. Across a wide variety of terrestrial and aquatic organisms, historical events have shaped patterns of geographical variation that have been inferred from current distributions of genetic polymorphisms (Avise, 1994; Templeton, 1998; Hewitt, 2000) or documented in long-term studies (Bradshaw & Holzapfel, 2001; Levitan, 2001; Solé, Balanyà and Serra, 2002; Levitan, 2003). While studies of *Drosophila* chromosomal polymorphisms have provided much of the foundation for understanding evolutionary processes in the 20th century (Lewontin *et al*., 1981; Krimbas & Powell, 1992), understanding of long-term historical causes shaping their distributions, including shifting climatic and vegetation patterns, remains unclear (e.g. Anderson *et al*., 1991). There exists abundant evidence confirming the roles of natural and sexual selection in maintaining the abundant chromosomal polymorphisms documented over the last 50 years in populations of *D. robusta* Sturtevant (Stalker & Carson, 1948; Levitan, 1961; Prakash, 1968; Levitan, 1992; Etges, 1996; Levitan, 2001), yet there have been few systematic attempts to relate geographical patterns of genetic polymorphisms to the evolutionary and biogeographical history of this species.

A major goal of this paper is to investigate the roles of current and historical forces of causation in shaping

*Corresponding author: E-mail: wetges@uark.edu

the geographical distribution of genetic polymorphisms in *D. robusta*. Because much of the data implicate genetic responses to climatic variability across the species range, we have reanalysed the geographical patterns in the polymorphism and related its current distributions to historical, climatic and vegetation community changes in North America. In addition, because many clinal patterns in the data are so distinctive that gaps in, and reverses from, them merit documentation and discussion of their possible import for the future evolution of the species. Some of these results were described by Etges & Levitan (2000).

HISTORY AND ECOLOGY OF THE *D. ROBUSTA* GROUP

Molecular and biogeographical evidence suggest that *D. robusta* and its closest relatives arose *c*. 36 Mya along with the closely allied *melanica* group (Throckmorton, 1975; Powell & DeSalle, 1995). The *D. robusta* species group is thought to have originated in south-eastern Asia for there are found a majority of all members of the group: *D. sordidula* Kikkawa & Peng, *D. pseudosordidula* Kaneko, Tokumitsu & Takada, *D. lacertosa* Okada, *D. okadai* Takada, and *D. neokadai* Kaneko & Takada from Japan, Korea, Taiwan, Nepal, India and Burma (Toda, 1988); and *D. cheda* Tan, Hsu & Sheng, *D. pullata* Tan, Hsu & Sheng, *D. gani* Watabe, Liang, & Zhang, *D. yunnanensis* Watabe, Liang & Zhang, *D. bai* Watabe, Liang & Zhang, *D. fluvialis* Beppu, Peng & Xie, and *D. medioconstricta* Watabe, Liang & Zhang from China (Narayanan, 1973; Watabe & Peng, 1991). Two closely related species traditionally assigned to the *robusta* group, *D. moriwakki* Okada & Kurokawa from Asia and *D. colorata* Walker from North America have recently been reassigned to the closely aligned *melanica* group, and a former member of the *virilis* group, *D. unimaculata* Strobl, from eastern Europe has been reassigned to the *robusta* group (Beppu, 1988). Because *D. robusta* and *D. colorata* are found only in North America, their ancestors must have crossed the Bering land bridge at a time when ecological conditions permitted transcontinental dispersal of these small, sap-feeding insects.

Many temperate species of *Drosophila* use the sap exudates of a variety of tree species, woody plants and vines as oviposition sites (Carson & Stalker, 1951; Carson, 1971; Shorrocks, 1982). In the eastern deciduous forests of North America, the principal breeding sites of *D. robusta* are sap fluxes of trees including the genera *Ulmus, Quercus, Morus, Prunus, Salix* and *Robinia* (Carson & Stalker, 1951). In northern Japan, *D. moriwakki, D. lacertosa*, *D. sordidula, D. pseudosordidula*, *D. okadai*, and *D. neokadai* have been reared from sap fluxes and rotting bark from trees in the genera *Ulmus, Morus, Betula* and *Fraxinus*, as well as timber yard logs and decaying vegetation (Kimura *et al*., 1977; Ichijo & Beppu, 1990). Families of deciduous trees, e.g. Ulmaceae, Betulacae and Salicaceae, are characterized by physiologically distinct groups of sap flux yeasts (Lachance & Starmer, 1982; Lachance, Metcalf & Starmer, 1982). Therefore, oviposition site preference for the sap fluxes of these forest tree species is a likely mechanism for ecological specificity in *robusta* group species that may have existed since the early Miocene.

From the beginning of the Quaternary 2.4 Mya until 0.9 Mya, glacial advances and retreats occurred on a *c*. 41 000 year cycle and more recently on about a 100 000 year cycle due to periodic fluctuations in the earth's orbit (reviewed in Delcourt & Delcourt, 1987). Analysis of long-term climatic variation in the temperate zone from deep ice cores (Stauffer, 1999) and palynological reconstructions of the more recent Holocene have provided a detailed record of climate and vegetation change over the last 20 000 years (Delcourt & Delcourt, 1987; Hewitt, 2001). Because much of the current range of *D. robusta* was covered by ice during glacial maxima, like other species, it was periodically driven into southern, mixed deciduous forest refugia. Thus, it is likely that repeated range expansions north during glacial minima from these southern refugia could have allowed the accumulation of genetic polymorphisms associated with these changing environments. Certainly, the magnitude of chromosomal polymorphism in *D. robusta* is likely to be relatively ancient given the time it has existed in North America.

Therefore, the goals of this study are (i) to reanalyse the geographical patterns of inversion polymorphism in *D. robusta* based on all published and unpublished data, and (ii) to relate current distributions of chromosomal polymorphisms in *D. robusta* to historical, climatic and vegetation community changes in North America.

MATERIAL AND METHODS

Drosophila robusta is one of the more common species of the genus in the deciduous forest of North America east of the Rocky Mountains and north of 28°N latitude in Florida. The haploid chromosome number of *D. robusta* is 4. The X-chromosome and the largest autosome, chromosome 2, are nearly equal in size, and both are nearly metacentric. A smaller autosome, chromosome 3, is somewhat less metacentric, and chromosome 4 is an acrocentric dot.

In the first detailed description of gene arrangement variation caused by inversions in this species (Carson & Stalker, 1947), certain arrangements were designated as 'Standard', and named for the respective left

or right arm: XL, XR, 2L, etc. (the dot was named 4). Most of the polymorphic arrangements in natural populations have proven to result from one-step inversions from these standards of both arms of the two largest chromosomes and the short arm of chromosome 3. Other arrangements were named and numbered by each arm in the order of their discovery (e.g. XL-1, XL-2, XR-1, 2L-1). While lacking paracentric inversions, the left arm of the third chromosome is involved in a polymorphic pericentric inversion spanning approximately two thirds of both arms.

The short-hand notation used by Carson (1953) will be used in reference to linked combinations of gene arrangements. The Standard arrangement of each arm is labelled 'S', and the other arrangements are referred to by the Arabic numerals in their names. So, a fly with gene arrangements XL and XR-2, for example, would be S2 in this notation. The arrangements discussed in this report have also been described in detail by Carson (1958) and Levitan (1982, 1992).

Adult *D. robusta* were baited over fermenting bananas and returned to the lab. Much of the data came from salivary gland smears from larvae derived from matings in the wild, so-called 'egg samples'. These often indicated full X-chromosomal karyotypes of the collected females but not their autosomes. To remedy this (Levitan, 1955) wild-caught males were mated to homokaryotypic females and their chromosomal constitution inferred from the salivary gland smears of at least six test cross larvae. Wild females were serially transferred to fresh medium until no F1 larvae were seen. These 'despermed' females were mated to homokaryotypic males and their chromosomal constitution inferred as in the case of the wild males. Some females that did not survive the desperming cross contributed data via the 'egg sample' route.

We analysed the gene arrangement frequency data for all populations sampled since 1946: these comprised the data summarized in Levitan (1992), Levitan & Etges (1995), and previously unpublished collections described in Table 1. The sampled localities are shown in Figure 1. Populations sampled at greater than 1500′ (457 m) were excluded unless there were no lower elevation sites nearby, such as the sites in western Nebraska and some on the Ozark Plateau. Arrangements such as XL-1 and 2L-3 have been found to increase in frequency with elevation throughout the species range, especially in the various divisions of the Appalachian Mountains, even when rare or absent in the adjacent lowlands (cf., e.g., the GA, NC and TN data of table 25 and those of tables 19 and 20 in Levitan, 1992). We averaged the frequencies of populations sampled over several years when the variation appeared to be relatively uniform. Samples of fewer than 20 were pooled with those of nearby samples whenever possible.

A total of 133 populations including 29 459 X chromosomes, 37 599 second chromosomes and 37 155 third chromosomes were sampled. These data are available upon request from the first author (WJE). Numbers of observed karyotypes were converted to percent frequencies and tested for normality using PROC UNIVARIATE (SAS Institute, 1989). In no instances did transforming the data have any effect on the results of any statistical tests, so all analyses were

Table 1. Recent, unpublished X chromosome inversion association and autosome inversion frequency data (as percentages) for populations of *D. robusta*. All sites are indicated in Fig. 1. N_x refers to the number of X chromosomes sampled, N_2 refers to the numbers of second chromosomes samples, and N_3 refers to the numbers of third chromosomes sampled

Locality	N_X	SS	S1	S2	1S	11	12	22
Robert Allerton Park, IL	264	12.5	36.4	23.9	0.8	10.2	11.0	5.3
Norfolk, NB	123	2.4	21.1	0.0	38.2	38.0	0.0	0.0
Fort Robinson, NB	117	0.9	0.9	0.0	95.7	2.6	0.0	0.0
Chadron, NB 1997[1]	190	0.0	0.5	0.0	96.8	2.6	0.0	0.0
Valentine, NB	26	0.0	4.0	0.0	72.0	24.0	0.0	0.0
Murdo, SD	77	0.0	7.8	0.0	62.3	30.0	0.0	0.0

	N_2	2L	2L-1	2L-2	2L-3	2R	2R-1	N_3	3R	3R-1
Robert Allerton Park, IL	354	64.7	21.4	2.0	11.9	82.5	17.5	354	65.5	34.5
Norfolk, NB	159	15.1	10.7	0.0	74.2	98.7	1.3	155	100.0	0.0
Fort Robinson, NB	166	0.6	0.6	0.0	98.8	100.0	0.0	165	100.0	0.0
Chadron, NB 1997	262	0.0	0.8	0.0	99.2	100.0	0.0	259	100.0	0.0
Valentine, NB	38	18.4	0.0	0.0	81.6	100.0	0.0	38	100.0	0.0
Murdo, SD	99	2.0	1.0	0.0	97.0	100.0	0.0	98	100.0	0.0

[1]Chadron was also sampled in 1955 (Carson, 1956) and both data sets were included for analysis in this study.

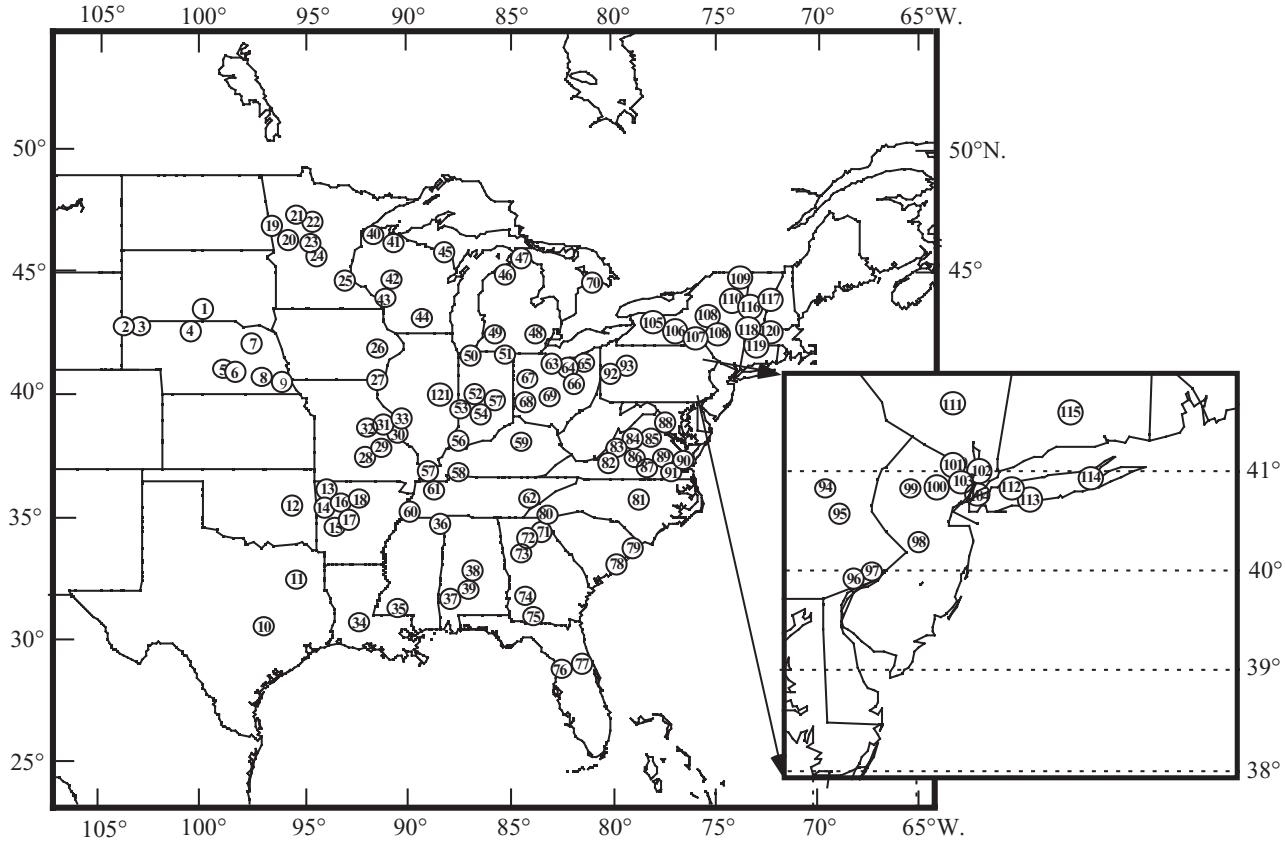

Figure 1. Site localities in the United States and Canada where *D. robusta* were collected as described in this study and Levitan (1992). Not all 133 populations mentioned in the text are indicated due to map scale. The following numbered sites are listed by state in a roughly west–east direction. SOUTH DAKOTA: (1) Murdo; NEBRASKA: (2) Fort Robinson, (3) Chadron, (4) Valentine, (5) Ravenna, (6) Rockville, (7) Norfolk, (8) Lincoln and Crete Reller, (9) Humboldt; TEXAS: (10) Navasota, (11) Tyler; OKLAHOMA: (12) Checotah; ARKANSAS: (13) Fayetteville, (14) Shores Lake, (15) Mill Creek, (16) Fane Creek, (17) Mt. Magazine, (18) Haw Creek Falls; MINNESOTA: (19) Moorhead, (20) Lake Ida, (21) Itasca State Park, (22) Park Rapids, (23) Lake Shamineau, (24) Little Falls, (25) Woodbury; IOWA: (26) Mt. Vernon, (27) Keokuk; MISSOURI: (28) Montauk State Park, (29) Steeleville, (30) University City, (31) Olivette, (32) Creve Coeur, (33) Pauline Hills and Eureka; LOUISIANA: (34) Opelousas; MISSISSIPPI: (35) Percy Quinn State Park and Eastabouchie, (36) Burnsville; ALABAMA: (37) Wagarville, (38) Demopolis, University, Pelham, Oak Park and Verbena, (39) Abbeville, Orion, Ozark and Troy; WISCONSIN: (40) Iron River, (41) Copper Falls State Park, (42) Augusta, (43) La Crosse, (44) Madison; MICHIGAN: (45) Iron Mountain, (46) Petoskey, (47) Cheboygan, (48) Ann Arbor, (49) Plainwell; INDIANA: (50) Michigan City, (51) Pokagon State Park, (52) Crawfordsville, (53) Terre Haute, (54) Bloomington, (55) Unionville, (56) Evansville, New Harmony and Vincennes; KENTUCKY: (57) Clinton, (58) Hopkinsville, (59) Lexington; TENNESSEE: (60) Shelby Forest State Park, (61) Milan and Greenfield, (62) Great Smoky Mountains National Park; OHIO: (63) Castalia, Bowling Green, Grand Rapids, Portage, and Van Buren, (64) Oberlin, (65) Brecksville, (66) Wooster, (67) Lima, (68) Dayton, (69) Columbus; ONTARIO, CANADA: (70) Owen Sound; GEORGIA: (71) Lake Russell, (72) Alto, (73) Emory, (74) Albany, Edison, Leary and Thomasville; FLORIDA: (75) Wellborn, Monticello and Perry, (76) Homosassa Springs, Lake Tsala Apopka and Inverness, (77) Crows Bluff and Astor Park; South Carolina, (78) Bull Island, (79) Myrtle Beach; NORTH CAROLINA: (80) Transylvania County, (81) Raleigh; Vg: (82) Shawsville, (83) Douthat State Park, (84), Afton, (85) Charlottesville, (86) Willis Mountain, (87) Sprouse's Corner, (88) Mt. Vernon, (89) Chester, (90) James City County, (91) Seashore State Park and Ingleside; PENNSYLVANIA: (92) Mt. Lebanon, (93) Clarion, (94) Jim Thorpe, (95) Allentown, (96) Swarthmore, (97) Philadelphia; NEW JERSEY: (98) Princeton, (99) Ledgewood, (100) Parsippany, (101) Paramus, (102) Englewood, (103) Englewood Cliffs; NEW YORK: (104) New York City Central Park, (105) Mendon Ponds County Park, (106) McDougall, (107) Deerfield, (108) Oneonta, (109) Plattsburgh, (110) Warren County, (111) Catskills Region, (112) Huntington, (113) Dix Hills, (114) Riverhead; CONNECTICUT: (115) Meridan; VERMONT: (116) Rutland, (117) Bridgewater; MASSACHUSETTS: (118) Williamstown, (119) Pittsfield, (120) Amherst; and the latest collection from ILLINOIS: (121) Robert Allerton Park.

performed with raw percentage frequency data. Although hierarchical analysis of the original observed diploid karyotype data would have allowed greater insight into patterns of population structure, these data have not been completely archived and thus were not available for all populations. Because of strong linkage (Levitan, 1958), X chromosome gene arrangement combinations (SS, S1, S2, 1S, 11, 12, 13, 22) were included for analysis, not single arm X chromosome arrangements. For the second chromosome, however, we focused on gene arrangements (2L, 2L-1, 2L-2, 2L-3, 2R) rather than arrangement combinations because of the relatively low frequency of 2R-1 in most of the species range. Arrangements 3R, 3R-1, and 3L-R were also included. Rare arrangements and arrangement combinations (Levitan, 1992) were not included for analysis.

Multiple regressions were performed for each X chromosome and gene arrangement with latitude (degrees N), longitude (degrees W), elevation (ft) and all interactions among these independent variables (PROC GLM; SAS Institute, 1989). Polynomial regression analyses were performed with latitude as the independent variable because of the known and often matching latitudinal and elevational clines in gene arrangement frequencies. We evaluated the Studentized residuals with the effects of elevation removed to identify outlier populations from the frequency trends.

Spatial autocorrelations were also calculated to assess the degree of spatial structure in populations of *D. robusta* (Oden, 1984). Spherical distances were calculated between all pairs of point locations based on degrees latitude and longitude. These distances were equally allocated into ten distance classes (Wartenberg, 1989). Both Moran's *I* and Geary's *C* coefficients (Sokal & Oden, 1978) were calculated for each X chromosome gene arrangement combination and the major autosomal gene arrangements and graphed with distance class.

In order to characterize patterns of covariation in chromosome and gene arrangement frequencies, we estimated the partial correlations between them with the effects of latitude, longitude and elevation removed. We compared this partial correlation matrix with one in which only the effects of elevation were partialled out, in order to identify covariation in gene arrangement and arrangement combination frequencies resulting from latitude and longitude. A principal components analysis (PROC PRINCOMP; SAS Institute, 1989) was also performed to graphically display the covariation in chromosomal frequencies.

RESULTS

Recent collections from western Nebraska and South Dakota confirmed that these populations are 'marginal' with respect to inversion polymorphism (Table 1), although the Chadron, Nebraska population is apparently not completely homokaryotypic as it was during an earlier sampling (Carson, 1956).

Perhaps the most remarkable feature across the species range of *D. robusta* is the north–south variation in chromosomal polymorphism as the polynomial regression equations with latitude were significant for all gene arrangements and arrangement combinations (Tables 2 and 3). For SS, S1, 11, 12, however, the

Table 2. Polynomial regression equations of gene arrangement and arrangement combination frequency on latitude (lat, °N). Coefficients in italics are significant at $P < 0.05$ ($F_{1,120} > 3.92$)

R^2		Regression equation
0.194	Freq(SS)****	= 19752 – *2107.56*(lat) + *83.42*(lat)^2 – 1.45(lat)^3 + *0.009*(lat)^4
0.120	Freq(S1)**	= 14330 – 1568.35(lat) + *62.21*(lat)^2 – 1.08(lat)^3 + 0.007(lat)^4
0.437	Freq(S2)****	= 12676 – *1365.91*(lat) + 55.23(lat)^2 – 0.99(lat)^3 + *0.007*(lat)^4
0.481	Freq(1S)****	= –37294 + *4126.98*(lat) – 169.74(lat)^2 + *3.07*(lat)^3 – *0.021*(lat)^4
0.322	Freq(11)****	= –12379 + *1361.80*(lat) – *55.59*(lat)^2 + 0.99(lat)^3 – *0.007*(lat)^4
0.161	Freq(12)***	= 4415.37 – 514.35(lat) + *21.97*(lat)^2 – 0.41(lat)^3 + 0.003(lat)^4
0.805	Freq(13)****	= 14117 – *1579.11*(lat) + *65.79*(lat)^2 – *1.21*(lat)^3 + *0.008*(lat)^4
0.618	Freq(22)****	= –15748 + *1631.97*(lat) – 62.53(lat)^2 + *1.05*(lat)^3 – *0.007*(lat)^4
0.543	Freq(2L)****	= 20349 – 2188.79(lat) + *86.87*(lat)^2 – *1.51*(lat)^3 + *0.010*(lat)^4
0.717	Freq(2L-1)****	= 3700.96 – *329.41*(lat) + *11.25*(lat)^2 – 0.17(lat)^3 + 0.001(lat)^4
0.585	Freq(2L-2)****	= –474.05 + *12.01*(lat) + *1.10*(lat)^2 – *0.05*(lat)^3 + 0.001(lat)^4
0.696	Freq(2L-3)****	= –23701 + *2531.70*(lat) – *100.31*(lat)^2 + *1.75*(lat)^3 – *0.011*(lat)^4
0.531	Freq(2R)****	= –11801 + *1246.52*(lat) – *49.10*(lat)^2 + 0.86(lat)^3 – *0.006*(lat)^4
0.914	Freq(3R)****	= –2352.65 + *396.14*(lat) – *21.20*(lat)^2 + *0.52*(lat)^3 – *0.004*(lat)^4
0.850	Freq(3L-R)****	= 7382.65 – 844.41(lat) ± 36.01(lat)^2 – 0.68(lat)^3 ± 0.005(lat)^4

Significance of the regression model; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

Table 3. Slopes of the multiple regressions of (a) X chromosome arrangement combination and (b) autosomal gene arrangement frequencies regressed on degrees north latitude, west longitude, elevation and the interactions between these effects for 133 populations of *D. robusta*. Per cent presence was calculated as the proportion of populations in which a chromosome or gene arrangement was present with a frequency greater then five per cent. R^2 is a measure of the percentage of variance explained by latitude (lat), longitude (long), elevation (elev) and the interactions between them

(a)

Variable	d.f.	SS	S1	S2	1S	11	12	13	22
Per cent presence		0.597	0.336	0.403	0.545	0.299	0.418	0.067	0.388
R^2		0.393	0.401	0.798	0.725	0.442	0.331	0.524	0.507
Lat	1	30.293***	0.019	–61.646****	43.196****	–14.295	0.480	–3.517	5.542
Long	1	12.872***	0.723	–27.261****	17.511****	–6.679	–0.136	–0.943	3.842
Elev	1	1.417**	–1.053*	–1.149****	0.118	–0.575	0.030	0.641**	0.526
Lat*long	1	–0.338***	0.026	0.656****	–0.504****	0.205*	–0.005	0.046	–0.084
Lat*elev	1	–0.037***	0.025*	0.029****	–0.003	0.015	0.000	–0.014*	–0.012
Long*elev	1	–0.016**	0.013*	0.013***	–0.003	0.007	0.000	–0.008**	–0.006
Lat*Long*elev	1	0.0004**	–0.0003*	–0.0003****	0.0000	–0.0002	–0.000	0.0002**	0.000

(b)

Variable	d.f.	2L	2L-1	2L-2	2L-3	2R	3R	3L-R
Per cent presence		0.910	0.806	0.448	0.276	0.985	0.896	0.119
R^2		0.343	0.835	0.500	0.767	0.761	0.750	0.615
Lat	1	2.153	–9.461*	–0.679	8.588	39.814****	24.384**	–4.068
Long	1	–0.871	–2.001	0.123	3.019	16.059****	6.316	–1.340
Elev	1	–0.964*	–0.424	0.281*	1.109**	1.747****	–0.727	0.540
Lat*long	1	0.017	0.042	0.001	–0.067	–0.406****	–0.200*	0.058
Lat*elev	1	0.022	0.009	–0.007*	–0.025*	–0.043****	0.015	–0.012*
Long*elev	1	0.013*	0.004	–0.003	–0.014*	–0.019****	0.009	–0.007**
Lat*long*elev	1	–0.0003*	–0.000	0.00007	0.0003*	0.0005****	–0.0002	0.0002**

Significance: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

amount of variance explained by the model, R^2, was low even with higher order effects included. This suggests that factors other than latitude, or those correlated with it, are responsible for the observed trends. In the case of S1, 12 and 2L, the first-order regressions on latitude were not statistically significant, but the higher order terms were, implicating non-linear frequency changes with latitude. In several cases, notably 2L and 3R (Table 2; Fig. 2) such non-linear trends were apparent when multiple gene arrangements replaced one another in a latitudinal fashion; high frequencies of 2L in central populations (Missouri, Illinois, Indiana and Ohio) are replaced by 2L-1 in the lowest southern latitudes and by 2L-3 in the highest northern ones, and 3R, after increasing clinally in frequency with latitude is replaced in many northern populations by the pericentric inversion 3L-R.

Spatial correlograms were statistically significant for all chromosome arrangements (Fig. 3, Table 4). Moran's *I* statistics were positive in the first three to four distance classes indicating strong, positive spatial autocorrelations for populations nearest to each other and up to 844 km apart. The upper limits for the ten distance classes were 339, 533, 693, 844, 993, 1143, 1325, 1525, 1808 and 2589 km. Frequencies of arrangement combinations S1, 13 and gene arrangement 2R were positively autocorrelated in the first distance class and then became negative with increasing distance (>340 km) as shown by the highly non-monotonic correlograms for both Moran's *I* and Geary's *C* (Fig. 3). These trends are consistent with the locally high frequencies of these genetic elements in different parts of the species range (Figs 2, 5). In some cases, chromosome frequencies were positively autocorrelated in several of the larger distance classes, suggesting frequency similarity among populations in disparate parts of the species range, such as S1, 12, 22, 2R and 3L-R.

Multiple regression analyses of frequency on latitude, longitude, elevation and all interactions between these variables revealed the extent of genetic variation across the range of *D. robusta* and helped to assess the patterns of spatial autocorrelations. Although we eliminated all rare gene arrangements

Table 4. Spatial autocorrelation results for each chromosome or gene arrangement in this study. All pairwise distances between localities were divided into equal numbered groups in ten distance classes

	Distance boundary (km)										
	339	533	693	844	993	1143	1325	1525	1808	2589	
	Distance class										Correlogram
	1	2	3	4	5	6	7	8	9	10	probability§
(a) Moran's *I*											
SS	0.51**	0.28**	0.09**	−0.01	−0.15**	−0.19**	−0.19**	−0.22**	−0.16**	−0.06*	0.000
S1	0.28**	−0.12**	−0.15**	−0.03	0.05*	−0.01	−0.10**	−0.05	0.10**	−0.04	0.000
S2	0.48**	0.18**	0.05*	0.00	−0.20**	−0.10**	−0.17**	−0.25**	−0.13**	0.06**	0.000
1S	0.41**	0.31**	0.12**	−0.14**	−0.16**	−0.15**	−0.08**	−0.11**	−0.15**	−0.11**	0.000
11	0.32**	0.31**	0.06**	0.01	−0.15**	−0.06*	−0.11**	−0.17**	−0.16**	−0.14**	0.000
12	0.35**	0.04	−0.08**	−0.11**	−0.01	−0.15**	0.19**	−0.02	−0.10**	−0.20**	0.000
13	0.33**	−0.13**	−0.10**	−0.02	−0.08**	−0.03	0.01	−0.02	−0.05	0.02	0.000
22	0.41**	0.10**	−0.05	−0.04	−0.19**	−0.15**	−0.11**	−0.05	−0.04	0.05*	0.000
2L	0.27**	0.21**	0.00	−0.11**	−0.09**	−0.14**	−0.14**	−0.06*	0.04	−0.05	0.000
2L-1	0.43**	0.36**	0.33**	0.09**	0.06*	−0.06	−0.20**	−0.26**	−0.43**	−0.38**	0.000
2L-2	0.45**	0.35**	0.28**	0.15**	−0.01	−0.14**	−0.24**	−0.22**	−0.32**	−0.37**	0.000
2L-3	0.58**	0.48**	0.34**	0.09**	0.04	−0.15**	−0.32**	−0.34**	−0.43**	−0.36**	0.000
2R	0.44**	−0.16**	−0.15**	−0.15**	0.13**	0.14**	0.04	−0.18**	−0.25**	0.06**	0.000
3R	0.43**	0.08**	−0.15**	−0.20**	−0.13**	−0.10**	0.17**	0.01	−0.05	−0.14**	0.000
3L-R	0.28**	0.06*	−0.02	−0.09**	0.01	−0.17**	0.00	0.05*	−0.09**	−0.10**	0.000
Average	0.40	0.16	0.16	−0.04	−0.06	0.10	−0.08	−0.13	−0.15	−0.12	
(b) Geary's *C*											
SS	0.31**	0.67**	1.03	1.15*	1.14*	1.14*	1.21**	1.28**	1.12	0.95	0.000
S1	0.45**	1.39**	1.40**	0.93	0.79*	0.98	1.08	0.97	0.86	1.15	0.001
S2	0.47**	0.62**	0.70**	0.82**	1.07	0.97	1.25**	1.28**	1.29**	1.56**	0.000
1S	0.31**	0.58**	0.88	1.35**	1.48**	1.57**	1.07	1.04	0.98	0.74	0.000
11	0.89	1.30*	1.06	0.90	1.18	0.78*	0.89	1.07	1.05	0.87	0.388
12	0.32**	0.54**	0.78	0.88	0.92	1.06	1.08	1.14	1.83**	1.49*	0.000
13	0.57**	1.72**	1.62**	1.16	1.05	0.96	0.77*	0.72*	0.84	0.56	0.000
22	0.44**	0.82**	1.02	1.15**	1.17**	1.20**	1.17**	1.05	1.04	0.95	0.000
2L	0.81*	0.65**	0.88	0.92	0.91	1.15*	1.13*	1.19**	1.05	1.31*	0.000
2L-1	0.09**	0.44**	0.48**	0.69**	0.96	1.00	1.25**	1.57**	1.82**	1.70**	0.000
2L-2	0.24**	0.33**	0.45**	0.70**	0.87*	1.10	1.37**	1.36**	1.68**	1.90**	0.000
2L-3	0.15**	0.45**	0.55**	0.76**	1.00	1.12*	1.38**	1.50**	1.58**	1.50**	0.000
2R	0.30**	1.12	1.13	0.96	0.79**	0.88	1.00	1.21**	1.42**	1.19	0.000
3R	0.31**	0.94	1.00	1.10	1.09	1.04	0.98	1.06	1.38**	1.11	0.000
3L-R	1.01	0.95	1.05	1.03	0.99	1.23**	0.84*	0.89	0.96	1.06	0.042
Average	0.44	0.84	0.94	0.97	1.03	1.08	1.10	1.16	1.26	1.20	

§Bonferroni approximation.
Significance: *$P < 0.05$, **$P < 0.01$.

from this analysis, some X chromosome gene arrangements and arrangement combinations were less geographically widespread than others, and therefore not expected to show significant latitudinal or longitudinal trends as compared to widespread arrangements. In order to test the null hypothesis that there was no relationship between geographical variability and range size for each X chromosome and gene arrangement, we calculated 'per cent presence', the proportion of all populations sampled in which each gene arrangement and arrangement combination was observed at frequencies greater than 5%. This estimate was a direct measure of the geographical commonness of each X chromosome or autosomal gene arrangement across all populations sampled. There was no correlation ($r = 0.223$, d.f. = 13) between per cent presence and R^2 for the 15 X chromosome arrangement combinations and gene arrangements considered. So, because a *D. robusta* X chromosome arrangement combination or autosomal gene is geographically widespread does not imply that its frequency varies significantly across its range.

The two least widespread chromosomes, chromosome arrangement combination 13 and pericentric inversion 3L-R, showed positive associations only with elevation or interactions with elevation (Table 3). However, some widespread forms such as SS, 2L-1, 2R and 3R, exhibited significant frequency changes with latitude or longitude, yet 2L, which exhibited the highest percent presence, showed only a negative association with elevation and frequencies influenced by the interaction between longitude and elevation. Its high percent presence may suggest widespread positive interactions with other second chromosome arrangements across the species range, i.e. heterozygote advantage, or linkage associations with right arm gene arrangements, although these effects are localized mostly in several south-eastern populations (Levitan & Etges, 1998).

Frequencies of arrangement combinations SS, S2, 1S, and gene arrangements 2R and 3R exhibited significant correlations with latitude as well as longitude, and those of 2L-1 with latitude alone (Figs 2, 4). In the case of 2R and 3R, the relation to latitude involves significant north–south clines (Table 3) despite the aforementioned reduction in frequency of 3R by 3L-R in northern Minnesota and Wisconsin (Figs 2, 5).

Although 2L-3 clearly increases in frequency at higher latitudes, it appears to do so mostly in the western part of the range. Many populations in which it reaches its highest frequencies are in the east: these were excluded to avoid confounding the effects of elevation with latitude. Thus, even though 2L-3 is common in these eastern populations, their removal from our analyses likely reduced latitude as a significant predictor of the frequency of 2L-3 (Table 3). Even so, elevation and interactions with latitude and longitude were significantly associated with variation in 2L-3. Over all collecting sites, elevation was correlated with latitude (Pearson $r^2 = 0.287$, $P = 0.0008$) and longitude (Pearson $r^2 = 0.555$, $P < 0.0001$) in this study. Therefore, the northerly distribution of 2L-3, its repeated increases in frequency with elevation in geographically isolated populations, and replacement of 2L and 2L-1 are responses to cooler climates associated with higher elevations and latitudes, particularly in the north and north-west parts of the species range in Minnesota, Wisconsin and western Nebraska (Fig. 2).

Several other chromosomal arrangements show remarkable patterns of local abundance, genetic discontinuities or disjunct distributions. Many of these cases were the consequence either of the expected decreases in frequency of certain gene arrangements in 'marginal' populations (Carson, 1955; Carson, 1956; Carson, 1959), variation in cline steepness or replacement of a clinally varying arrangement with an alternate one near the limits of the species range. Because of chance fluctuation, seasonal variation (Levitan, 1973a; Levitan, 1973b) and historical changes in several localities (Levitan, 2001; Levitan, 2003), the data were expected to exhibit some heterogeneity, for most of the samples were taken in different years and at different times of the year. Significant deviations from geographical trends were assessed by examination of the Studentized residuals from the multiple regression analyses. Localities with residual values greater than |2| were considered to be statistically significant outliers, i.e. were greater or lesser in frequency than that predicted from the regression model. In Florida and southern Georgia, for example, X chromosome combination S2 is near fixation and is significantly higher in frequency than would be predicted by the data along the South Carolina and Virginia coasts, confirming these as 'marginal' populations (Fig. 4). Further west in southern Alabama, Mississippi and Louisiana, S2 declines rapidly and is replaced by 22, which reaches its highest frequencies here and in the Ouachita mountains in Arkansas (Fig. 5). Another X chromosome combination, S1, exhibits significantly higher than predicted frequencies in Oklahoma, eastern Nebraska, Iowa and Missouri but drops off quite abruptly in the Arkansas Ozarks and in eastern Ohio (Fig. 5). Similarly, X chromosome combination SS shows a disjunct distribution, reaching high frequencies (45–65%) from Texas to central Louisiana, and north into Tennessee, eastern Missouri and Kentucky. Further to the north-east in New Jersey and Long Island, New York, populations reach the highest frequencies (80–90%) of X chromosome combination SS anywhere in the species range (Fig. 4), yet no positive spatial autocorrelations were observed in these larger distance classes (Table 4).

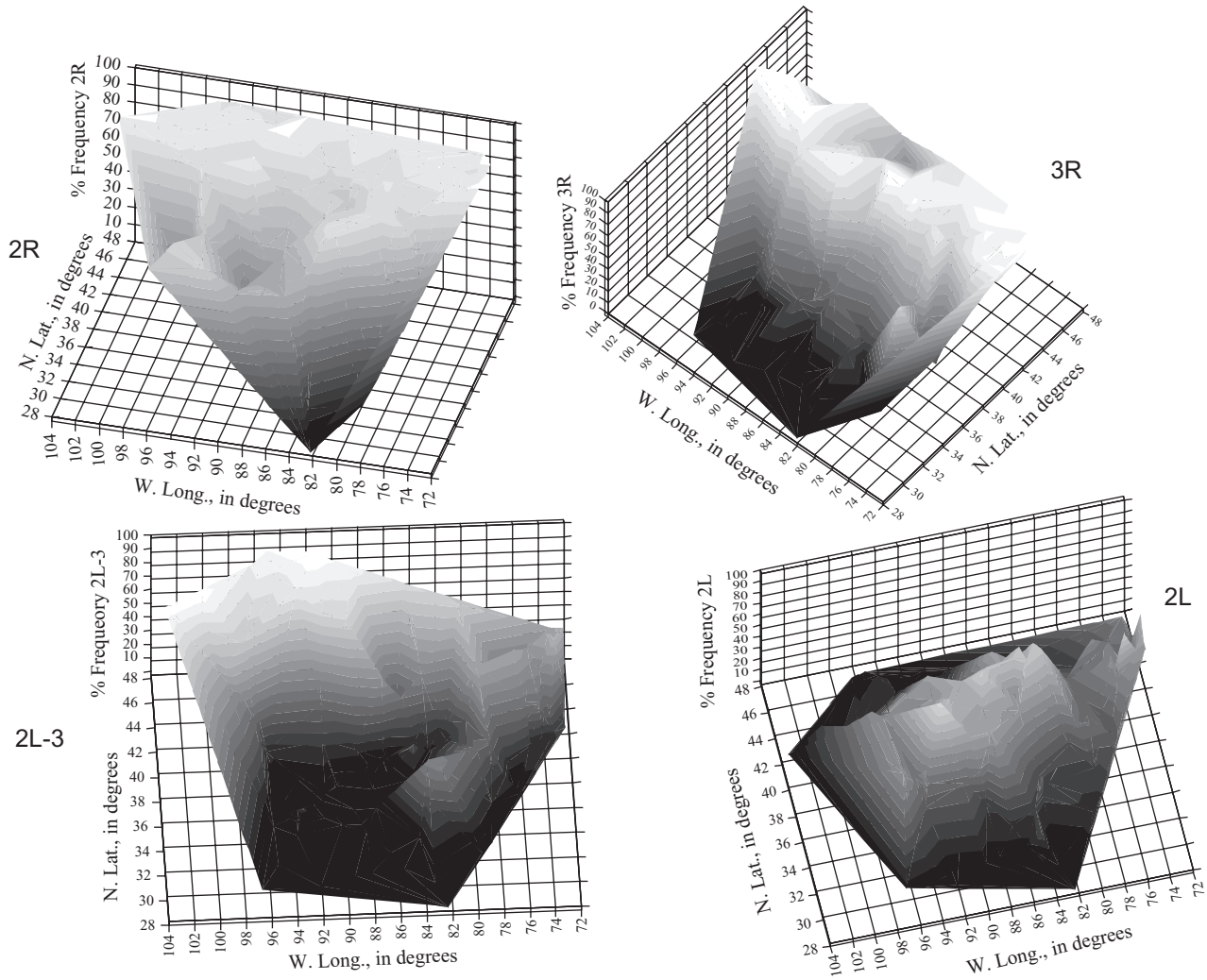

Figure 2. Geographical variation in the frequencies of second chromosomal inversions 2R, 3R, 2L-3, and 2L plotted by latitude (degrees N) and longitude (degrees W). Each graph was plotted to maximize the observed degree of variation for that chromosome. Names of chromosomes are defined in the text.

X chromosome combination 1S almost reaches fixation in the extreme north-east (Connecticut, Vermont, northern New York) and north-west (western Nebraska) parts of its range (Fig. 4). Combination 11, although more widely distributed than 3L-R (Fig. 5) in the northern part of the range, exhibits a similar, but more patchy pattern of abundance, and it reaches higher frequencies than would be expected on a clinal basis in Michigan, northern Indiana and southern Wisconsin. In north-western Minnesota and Wisconsin, 11 is replaced in high frequencies by 13, a relatively narrowly distributed northern X chromosome combination.

X chromosome combinations 12, 22 and gene arrangement 2L-2 exhibit disjunct distributions correlated with the major division between the Appalachian forest chain and the southern Ozarks and Ouachitas to the west, but incomplete sampling cannot be ruled out as there are no data from central Mississippi and northern Alabama. In particular, frequencies of combination 12 are significantly higher than predicted (based on the Studentized residuals) in the southern Arkansas Ozarks and across the Arkansas River south to Mt. Magazine (50.7%), and in southeastern Pennsylvania and northern New Jersey (Fig. 5) consistent with the positive spatial autocorrelation in the interval from 1143 to 1325 km (Table 4). The high frequency of combination 12 in the southern Ozarks represents a major break in frequency from other populations in the same region. Elevation cannot account for the Ozarks data inasmuch as samples from the Ouachita mountains (Mill Creek, AR) and the north central Ozarks (Steelville, MO) are at similar elevations and contain 12 frequencies that are more

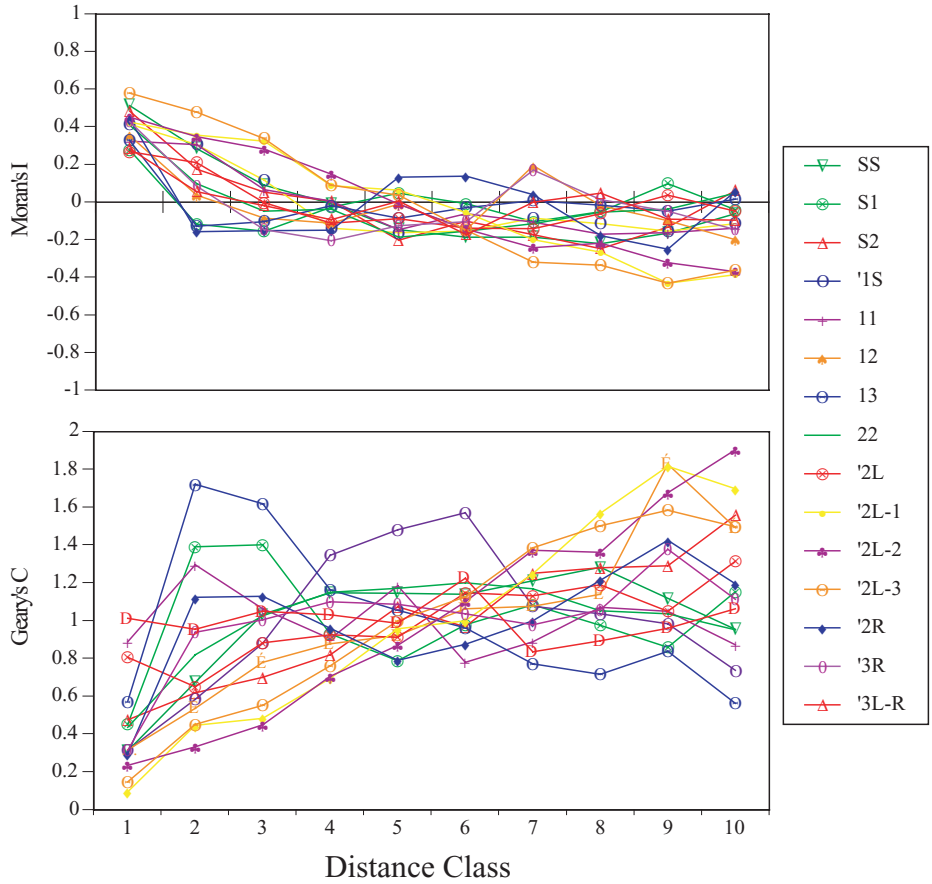

Figure 3. Spatial autocorrelation correlograms showing both Moran's *I* and Geary's *C* estimates plotted along the ten distance classes separating populations for each of the X chromosome associations and autosomal gene arrangements in this study. Each correlogram was statistically significant ($P < 0.0001$).

typical (<10%) for these latitudes. Fayetteville, at a much higher elevation (1540′) in the north-western Ozarks, contains an average of only 19.3% of all XL-1 bearing chromosomes, i.e. 1S + 11 + 12 ($N = 1716$ in seven samples from 1988 to 2001; Levitan & Etges, 1995; M. Levitan, unpubl. data), and even in the Appalachian Mountains XL-1 frequencies in the 50% range are not seen below 1500′ elevations (Levitan, 1992, table 25). X chromosome combination 22 shows a similar frequency 'peak' in the Ouachitas and on Mt. Magazine, and another from central Alabama to Emory, Georgia (Fig. 5). The discontinuity in frequency of 2L-2 is reinforced by the high Studentized residuals from the populations in Virginia to central Alabama, then dropping below expectations in northern Mississippi and western Tennessee, and then rising again to higher than predicted frequencies in the Ouachitas of south-western Arkansas.

Results of the partial correlation and Principal Components analyses revealed patterns of covariation among gene arrangements and arrangement combinations for individual chromosome arms and X chromosomes as well as between chromosomes (Table 5, Fig. 6). With the effects of latitude, longitude and elevation partialled out, the partial correlations showed the geographical range-wide degree of co-occurrence of all chromosomal elements. These correlations revealed significant geographical karyotypic structure in *D. robusta*, much of which was a reflection of clinal patterns, e.g. negative correlations among frequencies of SS and S2, and of 13 and 3R. That the first three principal components (Fig. 6) accounted for 65.3% of the variance in the data was surprising given that there were only three dependent variables. Partialling out elevation had little effect on the results (not shown), so only the PC plot for the raw frequency data is presented. PC 1 clearly grouped these chromosomal elements along a latitudinal gradient. PC 2 arrayed the chromosomes along an axis of broad distribution (2L, 3R, SS, 2R) to smaller, restricted distributions, concordant with the differences in percent presence. S2, 22, 2L-1, and 2L-2 are clearly grouped as 'south-

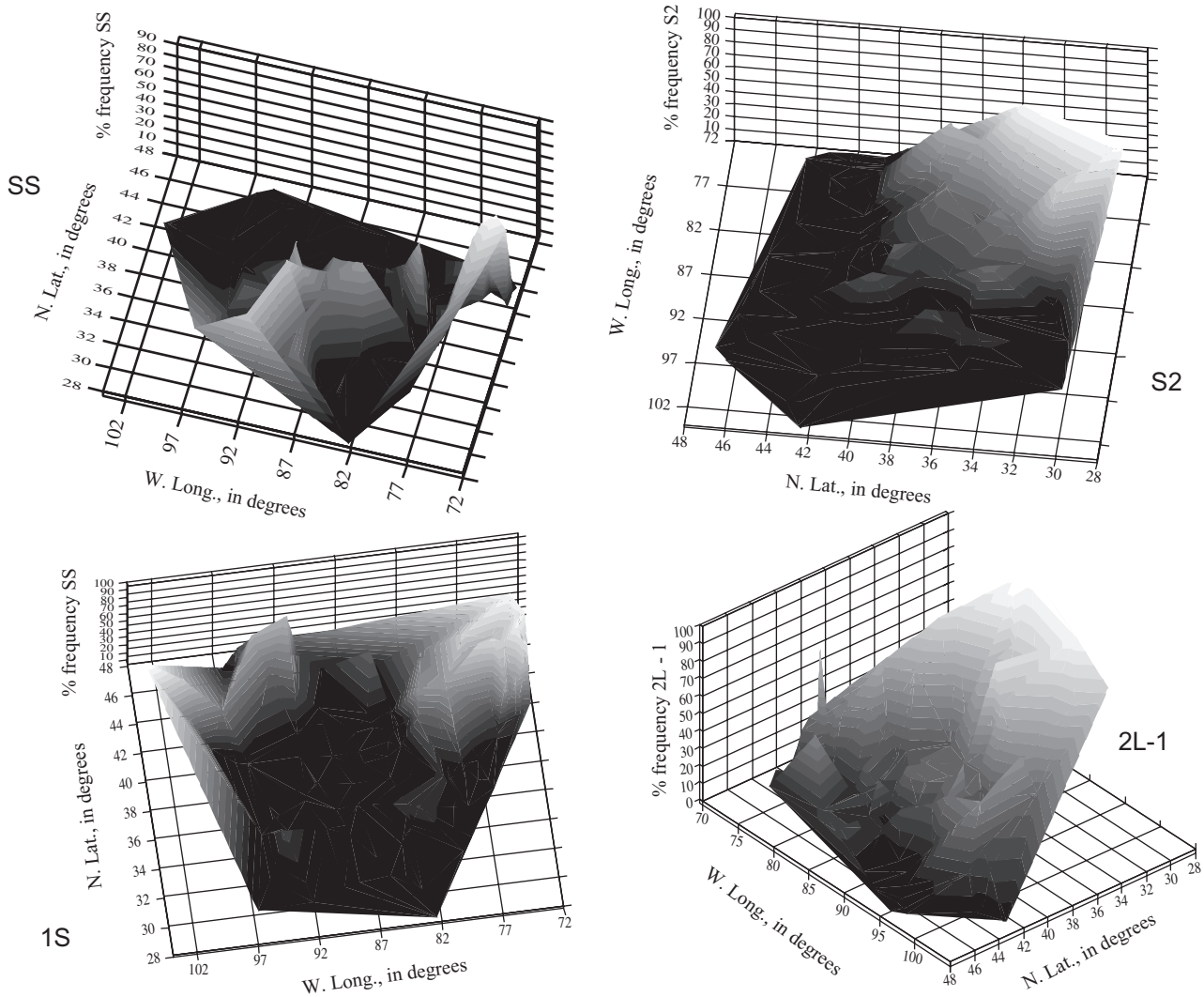

Figure 4. Geographical variation in the frequencies of X chromosome inversion combinations SS, S2, and 1S and second chromosome inversion 2L-1. See Fig. 2 for details.

ern' chromosomal forms, while 1S, 11, 2L-3, 2R, 3R and 3L-R are 'northern', with 12, S1, SS, and 2L as intermediate.

DISCUSSION

The exuberant variation in chromosomal polymorphism in *D. robusta* has retained a great deal of historical information about the great range shifts in climate in North America, as well as current patterns of natural selection. Despite huge anthropogenic alterations in the structure of the once contiguous eastern deciduous forests, a large number of the chromosomal arrangements in *D. robusta* display clinal distributions that reveal adaptation to temperate zone differences in eastern North America, as well as along multiple elevational transects. In particular, the 'warm' vs. 'cold' adapted X chromosome arrangement combinations S2 vs. 1S and gene arrangements 2L-1 vs. 2L-3, respectively, show similar latitudinal and elevational clines (Stalker & Carson, 1948; Carson, 1958; Levitan, 1978; Etges, 1984; Levitan & Scheffer, 1993).

The structure of some of these clines is also likely to be a consequence of historical forces shaping frequency variation along each transect. For example, detailed inspection of the Smoky Mountains cline revealed abrupt breaks in gene arrangement frequency and adult body size, particularly at elevations between 1400 and 2000′. Stalker & Carson (1948) interpreted this break as evidence for possible secondary contact between formerly disjunct lowland and mountain populations. Finding similar patterns 34

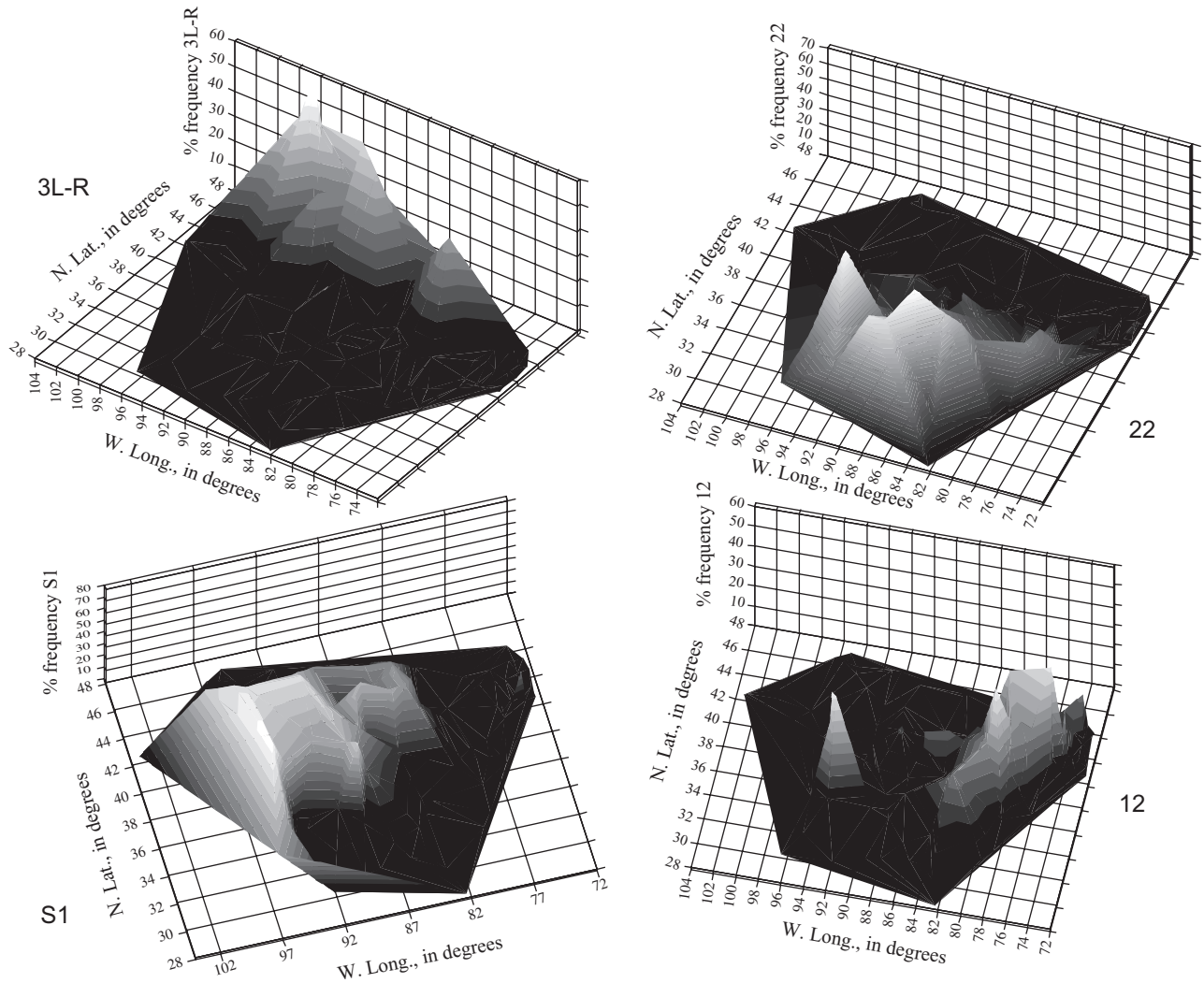

Figure 5. Geographical variation in the frequencies of pericentric inversion 3L-R and X chromosome combinations 22, S1, and 12. See Fig. 2 for details.

years later, Etges (1984) noted that these mid-elevation populations contained the highest modal frequencies of gene arrangements along the transect, and populations towards lower and higher elevations were characterized by increasing frequency differences. Gene arrangement-associated differences in components of fitness in populations along this transect demonstrated, however, that several of these clines are maintained by natural selection (Etges, 1989).

Such stepped clines are not surprising given the climatic history of high-elevation Appalachian forests over the last 18 000 years (Delcourt & Delcourt, 1987) and the altitudinal movements of deciduous tree species during glacial maxima (Ware, 1999). However, the Smokies clines have persisted through time despite short-term temporal shifts in gene arrangement frequencies. Frequencies of 'northern' gene arrangements and arrangement combinations such as 2L-3 and 1S have increased in frequency in all populations sampled from 1947 to 1981 with corresponding decreases in 'southern' arrangements like 2L-1, S2, and 22. Such systematic, directional shifts in frequencies were ascribed to forest canopy regrowth since the Great Smokies National Park was established in 1934, when most logging ceased (Etges, 1984). Hence, it is likely that this cline and the eight others like it in the Appalachian mountains have been influenced by Holocene climate-driven changes in forest composition, and over longer time intervals, perhaps since the most recent cooling period starting *c*. 5000 years ago (Whittaker, 1956). Until other supporting data are available, it

Table 5. Partial correlations between frequencies of the various X chromosomes and autosomal gene arrangements for the 133 populations of *D. robusta* in this study. The effects of latitude, longitude and elevation were partialled out

	SS	S1	S2	1S	11	12	13	22
SS	–	0.121	–0.489****	–0.051	–0.376****	–0.174	–0.123	–0.202*
S1		–	–0.305***	–0.234***	–0.042	–0.106	–0.394****	–0.416****
S2			–	–0.300***	0.152	–0.087	0.301***	–0.152
1S				–	–0.321***	–0.264**	–0.201*	–0.020
11					–	–0.074	–0.263**	–0.011
12						–	0.020	0.170
13							–	0.172

	2L	2L-1	2L-2	2L-3	2R	3R	3L-R
SS	0.273**	0.163	–0.146	–0.394****	0.363****	0.318***	–0.220*
S1	0.663****	–0.304***	0.066	–0.604****	0.052	0.589****	–0.506****
S2	–0.339****	0.136	0.093	0.277***	–0.577****	–0.469****	0.359****
1S	–0.239**	0.205*	–0.332****	0.256**	–0.038	0.216*	–0.203*
11	–0.129	–0.131	–0.015	0.253**	–0.045	0.008	–0.011
12	0.168	–0.301***	0.173	–0.047	0.195	–0.133	0.018
13	–0.399****	0.160	0.098	0.327***	–0.207*	–0.705****	0.848****
22	–0.188*	–0.031	0.413****	0.118	0.337****	–0.410****	0.212*

	2L	2L-1	2L-2	2L-3	2R	3R	3L-R
2L	–	–0.555****	0.093	–0.849****	0.105	0.515****	–0.533****
2L-1		–	–0.391****	0.105	–0.159	–0.114	0.149
2L-2			–	–0.160	0.211*	–0.252***	0.099
2L-3				–	–0.076	–0.454****	0.497****
2R					–	0.204*	–0.222*
3R						–	–0.827****

Significance: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

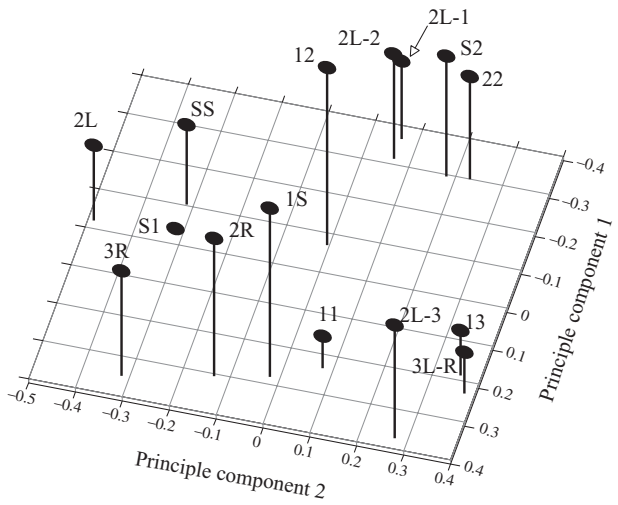

Figure 6. Results of the Principal Components Analysis showing variation in the frequencies of the 15 chromosomal elements in this study. See text for details.

is not currently possible to infer the age of many of these clines, but they could be very recent.

Biogeography and North American climate change

The similarities in the floras of temperate eastern North America and eastern Asia, noted by Linnaeus in 1750 and described in the late 1800s by Asa Gray (Graham, 1972), originated 20–25 Mya when eastern Asia and North America were closely connected. The temperate Arcto-Tertiary flora common to these two areas includes over 120 plant genera with 45 genera in 28 families of trees and shrubs (Little, 1983; Wu, 1983). By the mid-Miocene, *c*. 15 Mya, climates began to cool and the western and eastern North American forest biotas became isolated by western mountain building and the origin of dry interior grasslands in the late Miocene and early Pliocene 6–8 Mya. This time was marked by worldwide climate change associated with global changes in vegetation and lowering

of CO_2 abundance (Cerling *et al*., 1997). Thus, *D. robusta* and its distant relative, *D. colorata*, were isolated in North America from their Asian relatives at least 20–25 Mya.

Carson (1959) pointed out that in addition to the observed clinal variation in frequencies, several 'common' gene arrangements exhibited 'radiate' distributions, i.e. distributions that seemed centred in various parts of the species range, likely to be relicts of earlier population subdivision. These are 13 and 3L-R in the north, S1 and 2L in central populations, and 22 and 2L-2 in the south. Although these arrangements do not show consistent associations with environmental variables, we assume that *D. robusta* could never have persisted any further north or west than the extent of hardwood forests. Current northern and high elevation limits to the range of *D. robusta* are the mixed conifer hardwood forest–boreal forest ecotone described in Delcourt & Delcourt (1987) and prairie grasslands to the west. Carson (1959) suggested that the radiate distributions may have resulted from previous geographical subdivision of the species during phases of glacial maxima when isolated populations may have existed at the northern edge of the range, as well as in southern refugia. The many glacial advances and retreats would have permitted many opportunities for range expansion into northern latitudes from southern refugia in the Quaternary and before. Once established in northern refugial populations, frequencies could have increased by genetic drift through interglacial periods. However, in order to avoid extinction, these 'northern' gene arrangements must have dispersed back into southern refugia with each glacial maximum, or originated quite recently, perhaps in the last 20 000 years. The latter hypothesis is supported by the large number of 'rare' gene arrangements (2L-6, 2L-9, 2R-2, 2R-4, 2R-5 and 2R-6) that, along with 13 and 3L-R, have restricted northern distributions (Levitan, 1992). However, allozyme frequency clines associated with different gene arrangements and high allozyme heterozygosities in 'marginal' populations support repeated cycles of interglacial dispersal and constitute evidence for some form of balancing selection, even in small populations at the edge of the species range (Prakash, 1973).

Changes in north-eastern American forest tree species composition in the last 20 000 years revealed in pollen cores provide a detailed record of the dynamics of shifting climates in the late Quaternary (Delcourt & Delcourt, 1987). Because of the ancient ecological association of *D. robusta* and its host plants, the changing distributions of tree species in the genera *Celtis, Fraxinus, Ulmus, Populus, Salix* and other common genera such as *Acer, Carya* and *Quercus* over this time period suggest another possible mechanism influencing population subdivision and the current distribution of inversion polymorphisms. Many of these tree species show irregular and sometimes radiate distributions across eastern North America since the last glacial retreat. These distributions have changed markedly in addition to changes in overall forest composition (e.g. deciduous to conifer), such as the dominance of *Salix* in the Lower Mississippi alluvial valley beginning 10 000 years ago (chapter 5 in Delcourt & Delcourt, 1987). Others such as *Fraxinus, Populus* and *Ulmus* show concordant changes in dominance and dynamic east–west changes in abundance, as well as increases in frequency northwards. Assuming *D. robusta* populations were then using these resources, these historical shifts in host plant distributions may be partly responsible for local inversion frequency variation and the 'radiate' distributions of some elements (e.g. S1) first observed by Carson. Unfortunately, there exists no data implicating particular host species and the maintenance of individual gene arrangement frequencies. We can only speculate that overall climatic changes driving the geographical mosaic of tree community composition may have influenced the dynamics and genetic structure of *D. robusta* populations over the last 20 000 years, or perhaps more recently.

Distributions of southerly arrangements such as 22, S2 and 2L-2 seem easiest to interpret because of the long-term persistence of warm temperate forests below 33–34°N latitude. Sea-surface temperatures in the Gulf of Mexico have not changed more than 2 °C since the previous interglacial period 125 000 years ago, suggesting long-term persistence of these refugia until *c*. 14 000 years ago when cool temperate forests began to migrate northwards again (Delcourt & Delcourt, 1987). Increased numbers of ancient glaciations reaching further south may have caused the east–west differentiation seen now in the distributions of 22 and S2, although the high frequencies of 22 in the Ouachitas and 12 in southern Ozarks must be rather recent because the Ozarks were part of a boreal forest as recently as 18 000 years ago. The genetic discontinuities seen between the Appalachians and the Ouachitas and Ozarks may reflect the formation of mesic forest isolates caused by the Hypsithermal postglacial warming period 4000–8000 years ago. The distributions of these 'southern' arrangements is now north of the extent of these southern refugia, suggesting that they have not drifted to higher frequencies further north as other gene arrangements, e.g. S1, 1S, 2L-3, migrated to their current northern distributions. The speed with which *D. robusta* could have recurrently expanded northwards is suggested by the rates of Holocene tree species range extensions, 200–350 m/year (Davis, 1981).

Similar responses to temperate zone climatic variation have been documented in a few other *Drosophila*

species in eastern North American: latitudinal clines for chromosome polymorphisms in *D. melanogaster* Meigen (Mettler, Voelker & Mukai, 1977; Stalker, 1980) and *D. americana* Spencer (Spencer, 1938; McAllister, 2002). In contrast to *D. robusta* are the nearctic *D. melanica* Sturtevant and *D. funebris* Sturtevant groups that have speciated rather than maintaining a widespread distribution and abundant inversion polymorphism. Based on patterns of inversion sharing, reproductive isolation, and geographical distributions, Stalker (1966) proposed a phylogeny of the *melanica* group showing a step-wise ancestral-derived pattern of species relationships, where 'southern' species give rise to 'northern' species and these 'northern' species then give rise to 'southern' species, all over a rather recent time frame. Since these species are also sap-breeders closely allied with the *robusta* group (Throckmorton, 1975), their geographical pattern of diversification suggests a very different response to recent climatic cycling than that of *D. robusta*. The three members of the North American *funebris group, D. macrospina macrospina* Stalker & Spencer*, D. macrospina limpiensis* Mainland, and *D. macrospina ohioensis* Spencer show a south-west–northcentral, allopatric chain of ancestral-derived subspecies distributions also thought to have arisen in the Holocene (Mainland, 1942). Though as ancient as either of these other species groups, and ecologically similar to the *melanica* group, *D. robusta* has not speciated since its Miocene introduction into North America. Understanding how it has maintained this genetic cohesiveness would provide a great deal of illumination into the process of speciation.

Although extensive north–south clinal variation (Carson, 1959), concordant elevational clines in different parts of the species range (Levitan, 1978), and seasonal variation (Levitan, 1973a; Levitan, 1973b) suggest otherwise, the ecological sensitivity of *D. robusta* inversion polymorphism was thought to be relatively static over shorter time intervals in some populations. The patterns of spatial autocorrelations are certainly concordant with many of the observed latitudinal clines and several 'centres' of high chromosome frequencies (Table 4, Figs 3, 5), but these results certainly include uncontrolled temporal frequency changes given the long time intervals over which the data were collected. We assume that these temporal changes are small relative to the geographical scale of this study, and without more historical information, further sampling is necessary (cf. Sokal, Oden & Barker, 1987). For example, Carson (1958) sampled a population from Olivette, Missouri from 1946 through 1957, and found no significant shifts in frequencies even though the forest was demolished by commercial development in the last few years of the study. However, over longer intervals, inversion frequency shifts have become apparent. Etges (1984) demonstrated parallel increases in frequency for several 'northern' gene arrangements from 1947 to 1981 in multiple populations along an elevational cline in the Great Smoky Mountains National Park, Tennessee. More recently, Levitan (2001, 2003) has documented significant parallel increases in 'southern' gene arrangement frequencies in Central Park, New York City, Englewood, New Jersey, Allentown and Philadelphia, Pennsylvania, and the aforementioned Olivette, Missouri site over 38–56-year intervals. Thus, chromosome polymorphism in *D. robusta* is flexible over relatively short time intervals, the latter examples suggesting a possible response to recent larger scale climatic shifts (Levitan, 2001; Levitan, 2003; H. L. Carson, pers. comm.). Since invading North America, *D. robusta*'s ecology and inferred history have provided a detailed understanding of the current patterns of inversion polymorphisms that continue to respond to environmental change. Analyses aimed at detailing the rates at which these X chromosome arrangements and inversion polymorphisms have assumed their present frequencies are currently underway.

ACKNOWLEDGEMENTS

We thank H. L. Carson for continuous insight, moral support, and along with two anonymous reviewers, comments on the manuscript. L. Duncan and J. Tubbs provided valuable statistical advice, and C. L. Sagers suggested using the SAAP software. W. B. Heed and S. Ware provided encouragement and comments on the manuscript. J. Whitfield facilitated, D. B. Bowman, Park Superintendent, gave permission, and P. O'Grady helped in obtaining the collection in Robert Allerton Park. The Myron Kaplan Charitable Foundation and the Department of Biological Sciences, University of Arkansas provided financial and logistical support.

REFERENCES

Anderson WW, Arnold J, Baldwin DG, Beckenbach AT, Brown CJ, Bryant SH, Coyne JA, Harshman LG, Heed WB, Jeffery DE, Klaczko LB, Moore BC, Porter JM, Powell JR, Prout T, Schaeffer SW, Stephens JC, Taylor CE, Turner ME, Williams GO, Moore JA. 1991. Four decades of inversion polymorphism in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences, USA* **88:** 10 367–10 371.

Avise JC. 1994. *Molecular markers, natural history, and evolution*. New York: Chapman & Hall.

Beppu K. 1988. Systematic positions of three *Drosophila* species (Diptera: Drosophilidae) in the *Virilis-Repleta* radiation. *Proceedings of the Japanese Society of Systematic Zoology* **37:** 55–58.

Bradshaw WE, Holzapfel CM. 2001. Genetic shift in photoperiodic response correlated with global warming. *Proceedings of the National Academy of Sciences, USA* **98:** 14 509–14 511.

Carson HL. 1953. The effects of inversions on crossing over in *Drosophila robusta*. *Genetics* **38:** 168–186.

Carson HL. 1955. The genetic characteristics of marginal populations of *Drosophila*. *Cold Spring Harbor Symposium on Quantitative Biology* **20:** 276–287.

Carson HL. 1956. Marginal homozygosity for gene arrangement in *Drosophila robusta*. *Science* **123:** 630–631.

Carson HL. 1958. The population genetics of *Drosophila robusta*. *Advances in Genetics* **9:** 1–40.

Carson HL. 1959. Genetic conditions which promote or retard the formation of species. *Cold Spring Harbor Symposium on Quantitative Biology* **24:** 87–105.

Carson HL. 1971. The ecology of *Drosophila* breeding sites. University of Hawaii, Honolulu. *Harold L. Lyon Arboretum Lecture Number Two*: 1–27.

Carson HL, Stalker HD. 1947. Gene arrangements in natural populations of *Drosophila robusta*. *Evolution* **1:** 113–133.

Carson HL, Stalker HD. 1951. Natural breeding sites for some wild species of *Drosophila* in the eastern United States. *Ecology* **32:** 317–330.

Cerling TE, Harris JM, MacFadden BJ, Leakey MG, Quade J, Eisenmann V, Ehleringer JH. 1997. Global vegetation change through the Miocene/Pliocene boundary. *Nature* **389:** 153–158.

Davis MB. 1981. Quaternary history and the stability of forest communities. In: West DC, Shugart HH, Botkin DB, eds. *Forest succession: concepts and application*. New York: Springer-Verlag, 132–153.

Delcourt PA, Delcourt HR. 1987. *Long-term forest dynamics of the temperate zone*. New York: Springer-Verlag.

Etges WJ. 1984. Genetic structure and change in natural populations of *Drosophila robusta*: Systematic inversion and inversion association frequency shifts in the Great Smoky Mountains. *Evolution* **38:** 675–688.

Etges WJ. 1989. Chromosomal influences on life history variation along an altitudinal transect in *Drosophila robusta*. *American Naturalist* **133:** 83–110.

Etges WJ. 1996. Sexual selection operating in a wild population of *Drosophila robusta*. *Evolution* **50:** 2095–2100.

Etges WJ, Levitan M. 2000. *Clines, leaps, gaps and reverses in inversion polymorphism. XXI*. International Congress of Entomology, Iguacu Falls, Brazil, 555.

Graham A, ed. 1972. *Floristics and paleofloristics of Asia and Eastern North America*. Amsterdam: Elsevier.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907–913.

Hewitt GM. 2001. Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology* **10:** 537–549.

Ichijo N, Beppu K. 1990. Bionomics of Drosophilidae (Diptera) in Hokkaido XI. Five species of the *Drosophila robusta* species group. *Japanese Journal of Entomology* **58:** 625–636.

Kimura MT, Toda MJ, Beppu K, Watabe H. 1977. Breeding sites of drosophilid flies in and near Sapporo, Northern Japan, with supplementary notes on adult feeding habits. *Kontyû* **45:** 571–582.

Krimbas CB, Powell JR. 1992. *Drosophila inversion polymorphism*. Boca Raton, FL: CRC Press, Inc.

Lachance MA, Metcalf BJ, Starmer WT. 1982. Yeasts from exudates of *Quercus*, *Ulmus*, *Populus*, and *Pseudotsuga*: New isolations and elucidation of some factors affecting ecological specificity. *Microbial Ecology* **8:** 191–198.

Lachance MA, Starmer WT. 1982. Evolutionary significance of physiological relationships among yeast communities associated with trees. *Canadian Journal of Botany* **60:** 285–293.

Levitan M. 1955. Studies of linkage in populations. I. Associations of second chromosome inversions in *Drosophila robusta*. *Evolution* **9:** 62–75.

Levitan M. 1958. Non-random association of inversions. *Cold Spring Harbor Symposium on Quantitative Biology* **23:** 251–268.

Levitan M. 1961. Proof of an adaptive linkage association. *Science* **134:** 1617–1619.

Levitan M. 1973a. Studies of linkage in populations. VI. Periodic selection for X-chromosome gene arrangement combinations. *Evolution* **27:** 215–225.

Levitan M. 1973b. Studies of linkage in populations. VII. Temporal variation and X-chromosomal linkage disequilibriums. *Evolution* **27:** 476–485.

Levitan M. 1978. Studies of linkage in populations. IX. The effects of altitude on X-chromosome arrangement combinations in *Drosophila robusta*. *Genetics* **89:** 751–763.

Levitan M. 1982. The *robusta* and *melanica* groups. In: Ashburner M, Carson HL, Thompson JN Jr, eds. *The genetics and biology of Drosophila*. New York: Academic Press, 141–192.

Levitan M. 1992. Chromosomal variation in *Drosophila robusta* Sturtevant. In: Krimbas CB, Powell JR, eds. *Drosophila inversion polymorphism*. Boca Raton, FL: CRC Press, Inc., 221–338.

Levitan M. 2001. Studies of linkage in populations. XIV. Historical changes in frequencies of gene arrangements and arrangement combinations in natural populations of *Drosophila robusta*. *Evolution* **55:** 2359–2362.

Levitan M. 2003. Climatic factors and increased frequencies of 'southern' chromosome forms in natural populations of *Drosophila robusta*. *Evolutionary Ecology Research* **5:** 597–604.

Levitan M, Etges WJ. 1995. Convergent linkage disequilibrium in disparate populations of *Drosophila robusta*. In: Levine L, ed. *Genetics of natural populations: the continuing importance of Theodosius Dobzhansky*. New York: Columbia University Press, 105–119.

Levitan M, Etges WJ. 1998. Studies of linkage in populations. XIII. A unique cause of linkage disequilibrium in natural populations of *Drosophila robusta*. *Heredity* **80:** 660–667.

Levitan M, Scheffer SJ. 1993. Studies of linkage in populations. X. Altitude and autosomal gene arrangements in *Drosophila robusta*. *Genetical Research* **61:** 9–20.

Lewontin RC, Moore JA, Provine WB, Wallace B, eds. 1981. *Dobzhansky's genetics of natural populations, I-XLIII*. New York: Columbia University Press.

Little EL. 1983. North American trees with relationships in eastern Asia. *Annals of the Missouri Botanical Garden* **70:** 605–615.

Mainland GB. 1942. Genetic relationships in the *Drosophila funebris* group. *University of Texas Publication* **4228:** 74–112.

McAllister BF. 2002. Chromosomal and allelic variation in *Drosophila americana*: selective maintenance of a chromosomal cline. *Genome* **45:** 13–21.

Mettler LE, Voelker RA, Mukai T. 1977. Inversion clines in populations of *Drosophila melanogaster*. *Genetics* **87:** 169–176.

Narayanan Y. 1973. The phylogenetic relationships of the members of the *Drosophila robusta* group. *Genetics* **73:** 319–350.

Oden NL. 1984. Assessing the significance of a spatial correlogram. *Geographical Analysis* **16:** 1–16.

Powell JR, DeSalle R. 1995. *Drosophila* molecular phylogenies and their uses. *Evolutionary Biology* **28:** 87–138.

Prakash S. 1968. Chromosome interactions affecting mating speed in *Drosophila robusta*. *Genetics* **60:** 589–600.

Prakash S. 1973. Patterns of gene variation on central and marginal populations of *Drosophila robusta*. *Genetics* **75:** 347–369.

SAS Institute I. 1989. *SAS/STAT user's guide*, *Version* 6. SAS Institute Inc., Cary, NC.

Shorrocks B. 1982. The breeding sites of temperate woodland *Drosophila*. In: Ashburner M, Carson HL, Thompson JN Jr. eds. *The genetics and biology of Drosophila*. New York: Academic Press, 385–428.

Sokal RR, Oden NL. 1978. Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnaean Society* **10:** 199–228.

Sokal RR, Oden NL, Barker JSF. 1987. Spatial structure in *Drosophila buzzatii* populations: simple and two-dimensional spatial autocorrelation. *American Naturalist* **129:** 122–142.

Solé E, Balanyà J, Serra L. 2002. Long-term changes in the chromosomal inversion polymorphism of *Drosophila subobscura*. I. Mediterranean populations from southwestern Europe. *Evolution* **56:** 830–835.

Spencer WP. 1938. *Drosophila virilis americana,* a new subspecies. *Genetics* **23:** 169–170.

Stalker HD. 1966. The phylogenetic relationships of the species in the *D. melanica* group. *Genetics* **53:** 327–342.

Stalker HD. 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 HD, Carson HL. 1948. An altitudinal transect of *Drosophila robusta* Sturtevant. *Evolution* **2:** 295–305.

Stauffer B. 1999. Cornucopia of ice core results. *Nature* **399:** 412–413.

Templeton AR. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7:** 381–397.

Throckmorton L. 1975. The phylogeny, ecology, and geography of *Drosophila*. In: King RC, ed. *Invertebrates of genetic interest*. New York: Plenum Press, 421–469.

Toda MJ. 1988. Drosophilidae (Diptera) in Burma. III. The subgenus *Drosophila*, excepting the *D. immigrans* species-group. *Kontyû* **56:** 625–640.

Ware S. 1999. Paleoclimatic causes of altitudinal and geographical distribution of high elevation woody species in Virginia. In: Eckerlin RP, ed. *Proceedings of the Appalachian Biogeography Symposium: Virginia Museum of Natural History.* Martinsville, VA: Virginia Museum of Natural History, 45–57.

Wartenberg D. 1989. *SAAP – A spatial autocorrelation analysis program. 4.3 ed.* Piscataway, NJ: Exeter Software.

Watabe H, Peng TX. 1991. The *Drosophila virilis* section (Diptera: Drosophilidae) from Guangdong Province, southern China. *Zoological Science* **8:** 147–156.

Whittaker RH. 1956. Vegetation of the Great Smoky Mountains. *Ecological Monographs* **26:** 1–80.

Wu Z. 1983. On the significance of Pacific intercontinental discontinuity. *Annals of the Missouri Botanical Garden* **70:** 577–590.