Molecular Analysis of an Allozyme Cline: Alcohol Dehydrogenase in *Drosophila melanogaster* on the East Coast of North America

A. Berry[1] and M. Kreitman[2]

Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544-1003

Manuscript received July 17, 1992
Accepted for publication March 31, 1993

ABSTRACT

Clines may either be selectively maintained or be the by-product of nonadaptive processes related to population structure and history. *Drosophila melanogaster* populations on the east coast of North America show a latitudinal cline in the frequencies of two common electrophoretically distinguishable alleles at the alcohol dehydrogenase locus (*Adh*), designated *Adh-S* and *Adh-F*. This cline may either be adaptive or an artifact of a possible recent dual founding of North American *D. melanogaster* populations in which frequencies of *Adh* alleles differed between founder populations. By means of a high resolution restriction-mapping technique, we studied the distribution of 113 haplotypes derived from 44 polymorphic DNA markers within the *Adh* region in 1533 individuals from 25 populations throughout the cline. We found significant clinal differentiation at the polymorphism determining the mobility-difference causing amino acid replacement between *Adh-F* and *Adh-S* alleles. Hitchhiking was limited, despite extensive linkage disequilibrium, and other sites did not vary clinally. Such a pattern of differentiation implies that selection is responsible for the cline. To investigate whether selection acts only on the *Adh-F/S* site, we performed a "selective equivalence" test under the assumption that all variability within the specified allelic class is selectively neutral. This revealed selective equivalence among *Adh-S*-bearing haplotypes, whose frequencies showed no differentiation throughout the cline, implying high levels of frequency-homogenizing gene flow. Geographical heterogeneity among *Adh-F*-bearing haplotypes implied the action of selection on one or more additional variants in linkage disequilibrium with *Adh-F*. In a further study of a subset of the data (n = 1076 from 18 populations), we found a combined insertion/deletion polymorphism, designated $\nabla 1$, located in the 5′ adult intron and in linkage disequilibrium with *Adh-F*, to show more marked clinal variation than *Adh-F/S*. Although the unequivocal identification of the precise target(s) of selection requires further study, we suggest that clinal selection may be acting epistatically on the *Adh-F/S* and $\nabla 1$ polymorphisms.

NO enzyme polymorphism has been more intensively studied than the *Fast/Slow* (*Adh-F/Adh-S*) polymorphism at the alcohol dehydrogenase locus (*Adh*) of *Drosophila melanogaster*; CHAMBER's (1988) review cites 364 papers. Among these are attempts to resolve whether the latitudinal clines in *Adh* allozyme frequencies result from selection or drift (*e.g.*, OAKESHOTT *et al.* 1982). Experimental approaches to identify the selective factor(s), either by manipulation of artificial populations or by direct *in vitro* or *in vivo* studies of enzyme activity, have yielded equivocal results. The allozymes differ in their kinetic properties (CHAMBERS 1988; HEINSTRA, SCHARLOO and THÖRIG 1987) and their thermostabilities (VIGUE and JOHNSON 1973), and on average *Adh-F* alleles yield twice as much active protein as *Adh-S* alleles (LAURIE-AHLBERG 1985). But these differences do not correlate with known environmental variables (VAN DELDEN 1982; HEINSTRA, SCHARLOO and THÖRIG 1987). This inconsistency may result from a failure to identify or mimic adequately natural selection pressures, or from an inability to detect small fitness differences. Selection coefficients only need to be larger than the reciprocal of population size, N_e, to influence gene frequencies (EWENS 1979). Given *D. melanogaster*'s estimated N_e of 10^6 (KREITMAN 1983), and that selection coefficients of less than a few percent are experimentally undetectable in *Drosophila* (ÁRNASON 1991), there remains a large window of selection coefficients causing effective but undetectable selection.

The most compelling evidence in favor of selection comes not from direct fitness measurements but from surveys of genotype frequencies in natural populations. PIPKIN *et al.* (1976) identified a discontinuity in allele frequencies between Mexican highland and lowland populations. OAKESHOTT *et al.* (1982) observed parallel clines on a number of continents, including one on the east coast of Australia in which the north-south cline of the Northern Hemisphere has become a complementary south-north cline. Nevertheless, not

[1] Current address: MCZ Laboratories, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138.
[2] Current address: Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637.

all comparable ranges of latitude show clines: SMITH *et al.* (1984) failed to find a cline on the west coast of the United States.

We studied the *Adh* cline on the east coast of the United States where frequencies of *Adh-S* range between 90% in Florida and 55% in Maine (VIGUE and JOHNSON 1973). This, like the other clines identified by OAKESHOTT *et al.* (1982), may either be the product of selection or of species history and population structure. Genetic and historical information suggest that *D. melanogaster* has recently expanded its range (JOHNSON 1913; DAVID and CAPY 1988): it originated in tropical Africa, possibly migrated into Eurasia early in its history, and arrived in the New World and Australia with the human movements of historical times. *D. melanogaster* has therefore possibly been on the east coast of the United States for only the past 200–300 years. Such a recent introduction might have facilitated the formation of a cline, especially in view of the likely colonization of the north by flies coming from northern Europe (where populations are largely *Adh-F*) and of the south by flies from Africa (largely *Adh-S*) (DAVID and CAPY 1988). Under this "historical" hypothesis, limited subsequent gene flow in the short time since colonization has established the observed cline between the two points of introduction.

How can history and selection be distinguished? Because selection affecting allozyme frequencies may be undetectable in the laboratory, a test based on the distribution of genotypes in nature is desirable. CAVALLI-SFORZA (1966) pointed out that selection determining genotype frequencies at a specified locus will have no effect on allele frequencies at unlinked polymorphic loci. Drift or historical processes, on the other hand, should affect all loci equally. Thus selection predicts clinal variation at the loci under selection superimposed on a background of drift/migration-determined variation at other loci, whereas history predicts similar clines, or gene frequency differentiation, at all sites whose frequencies differed between the founder populations. A formal statistical test of this idea was developed by LEWONTIN and KRAKAUER (1973) but was subsequently shown to be of limited applicability (NEI and MARUYAMA 1975; ROBERTSON 1975). Previous applications of this approach employed polymorphic allozyme markers: for example, CHRISTIANSEN and FRYDENBERG (1974) found two clinal and two non-clinal loci in Danish populations of the eelpout, *Zoarces viviparus*. Because all protein variants might be visible to selection, we cannot determine which loci, if any, are selected and which, if any, are unselected. We therefore need to study the distribution of variation at both the putatively selected loci and at neutral loci which serve as indicators of population structure; discrepancies between the two distributions imply selection at the study loci.

We have studied the distribution of DNA variation in and around the *Adh* locus in *D. melanogaster* populations throughout the east coast. Of our 19 highly variable restriction fragment length polymorphism (RFLP) markers, only one (the *Fast/Slow* site) corresponds to an amino acid polymorphism; all the others are effectively silent. While this does not necessarily imply that they are not under selection, we assume they are more likely to be neutral indicators of population history than allozymes. Previous geographic studies of *Adh* RFLP polymorphism (KREITMAN and AGUADÉ 1986; SIMMONS *et al.* 1989) found frequencies of these effectively silent markers to be generally uniform over large geographic distances, suggesting that they are indeed neutral markers of population history.

In addition to treating each polymorphic restriction site in the *Adh* region as a separate marker, we can also combine all polymorphic sites to yield a single haplotype (*i.e.*, the particular pattern of presence or absence of a variant at each polymorphic site) for each line. Having a large number of different haplotypes (113) improves the sensitivity of population structure analysis. As European and African populations are differentiated (AGUADÉ 1988; A. BERRY and M. KREITMAN, unpublished data), the historical explanation of the cline predicts haplotypic differentiation between northern (Europe-derived) and southern (Africa-derived) populations. On the other hand, with sufficient migration, there will be little or no differentiation of haplotypes, implying that the allozyme cline is subject to selection.

We confined the study to the *Adh* region, rather than additionally studying variation in other regions, in order to answer questions if we found selection to be responsible for the cline: for example, whether selection is actually acting on the *Adh-F/Adh-S* polymorphism or whether the polymorphism serves merely as a marker for other sites under selection. If selection were acting only on the *F/S* polymorphism and all other markers are neutral, then all haplotypes within each allelic class are selectively equivalent. Thus we can explicitly test hypotheses about clinal selection on specified sites: if clinal selection was acting only on the *F/S* site, then the distributions of haplotypes within both the *Adh-F* and *Adh-S* allelic classes would be similar, with both being attributable to the same forces of drift and migration. On the other hand, selection on sites other than the *F/S* site would result in selective heterogeneity between the *Adh-F* and/or *Adh-S* classes, implying that selection is discriminating between different haplotypes within each class.

Our survey revealed patterns of polymorphism consistent with an adaptive explanation of the cline: extensive clinal differentiation occurs only at one variable site, the *F/S* polymorphism. However, although

Adh-S-bearing haplotypes (*S*-haplotypes) appear to be selective equivalents, we detected geographic heterogeneity in the distribution of *F*-haplotypes, suggesting that polymorphism(s) segregating within the *Adh-F* class are subject to natural selection. The second part of this paper comprises a survey of additional polymorphisms for a subset of the original dataset in an attempt to identify the factor(s) responsible for the within-*Adh-F* heterogeneity.

Linkage to selected inversions is not likely to be responsible for the among-*F*-haplotypes heterogeneity. The only inversion known to be in linkage disequilibrium with *Adh-F/S* is *In-2Lt* (VOELKER *et al.* 1978), but this is entirely associated with *Adh-S*. How, therefore, can we home in on the source of this heterogeneity? AQUADRO *et al.* (1986), and others, have noted that the *Adh-F* allele produces on average a two- to threefold higher concentration of ADH protein than the *Adh-S* allele. By site-directed *in vitro* mutagenesis, CHOUDHARY and LAURIE (1991) have shown that the *F/S* amino acid polymorphism is not responsible for this difference. By means of *P* element transformation of genetically uniform flies with varying *Adh* haplotypes, LAURIE-AHLBERG and STAM (1987) showed earlier that all the variability in ADH concentration maps to within a 2.3-kb fragment including all the *Adh* coding sequence and some intron and 3′-flanking sequence. It is here, therefore, that we looked for polymorphisms causing the heterogeneity within *F*-haplotypes.

Two candidate heterogeneity-causing polymorphisms were the "*Château Douglas*" site and the ∇*1* length polymorphism. *Château Douglas* is a cosmopolitan thermostable *Adh-F* allele (LEWIS and GIBSON 1978; SAMPSELL 1977; THÖRIG, SCHOONE and SCHARLOO 1975) occurring at low frequencies in natural populations (WILKS *et al.* 1980; SAMPSELL and STEWARD 1983). Its *Adh-S*-like properties are apparently conferred by a single amino acid substitution (COLLET 1988; EISSES *et al.* 1990). We studied ∇*1* because LAURIE, BRIDGHAM and CHOUDHARY (1991), in their study of the factors affecting ADH-protein concentration, suggested that it might have a major effect. It is polymorphic in both populations in which it has been studied (Raleigh, North Carolina, and Putah Creek, California; KREITMAN and AGUADÉ 1986) and, significantly, is in linkage disequilibrium with *Adh-F*. The fact that both of these sites are known to be variable within *Adh-F* and not to occur on *Adh-S* alleles suggests that they may be responsible for the observed heterogeneity among *F*-haplotypes, coupled with a lack of heterogeneity among *S*-haplotypes.

MATERIALS AND METHODS

Flies: Table 1 lists the 25 populations sampled. Whenever possible, we structured our sampling as follows: resampling populations over a number of years to ascertain the extent to which gene frequencies vary between years; sampling several populations within a single locality (*i.e.*, all populations at approximately the same latitude) to gauge the amount of local differentiation; and sampling localities throughout the east coast to investigate the effect of latitude. The entire sample was split into two blocks, north and south, each encompassing about 10° of latitude. These divisions (between blocks; among localities within blocks; among populations within localities; among years within populations) form the basis of the hierarchical analyses of population subdivision outlined below. In southern areas, flies were collected in the winter or spring, but northern populations were only sufficiently abundant to permit collection at the end of the summer. It was impossible to control for the type of fruit over which the flies were collected, but, in general, we collected from a mixture of rotting material in garbage containers behind roadside fruit stands.

We established approximately 100 isofemale lines immediately after making each collection and made each line homo- or hemizygous for the *Adh* gene region of a single wild-caught chromosome using a marked chromosome carrying a deletion of the *Adh* region (*A178*, supplied by M. ASHBURNER).

RFLP analysis: All procedures followed KREITMAN and AGUADÉ (1986) except for populations collected since 1987. For these, restriction digests were carried out in micro-titre plates and the DNA precipitated using isopropanol rather than ethanol (COYNE, AULARD and BERRY 1991; KREITMAN 1991). We used the following enzymes: *Alu*I, *Dde*I, *Hae*III, *Hha*I, *Msp*I, *Sau*3AI and *Taq*I. *Dde*I digests were combined with *Bam*HI to reveal an extra polymorphic site. Southern blots were probed with a 2.7-kb *Sal*I-*Cla*I fragment incorporating the *Adh* structural locus (KREITMAN 1983). From the known sequence, we find that this method detects virtually all insertion/deletion variation in the region and about 20% of all possible substitutions. Scoring of autoradiographs was based on the published sequences for eleven *Adh* genes (KREITMAN 1983), which include many of the observed restriction site polymorphisms. Further polymorphisms were identified in the studies of KREITMAN and AGUADÉ (1986) and SIMMONS *et al.* (1989) and additional ones could be ascribed to a single base in the case of a site gain, to any one of four bases in the case of a site loss, and to a minimal length restriction fragment in the case of insertion/deletion variation. Table 2 lists the sites scored. Our analysis includes most of the sites scored by previous authors. However, one notable exception is an *Alu*I cut site at position 1068, which is highly polymorphic in this and previous studies but was omitted because it was difficult to score in two of our populations. Nevertheless, this polymorphism is in total linkage disequilibrium with a number of sites which together comprise the "*Washington-Slow*" haplotype (*Wa-S*; see below), so its exclusion results in no loss of information.

Scoring the *F/S* polymorphism: This polymorphism was not detected with the restriction enzymes used in this study. We therefore assayed four to five individuals from each isogenic line for *Adh* allele by means of agarose gel electrophoresis of total body homogenate and staining for ADH activity (according to SIMMONS *et al.* 1989). We confirmed this scoring by probing the Southern blot filters with allele-specific oligonucleotides (CCACCCTGGTGCACACGTTC for *Adh-F*, and CCACCCTGGTGCACAAGTTC for *Adh-S*) under conditions sufficiently stringent to permit only the hybridization of the full 20-bp match. *Sau*3AI filters were probed for *Adh-S* and *Hha*I filters for *Adh-F*. Oligonucleotides were end-labeled by means of a kinasing reaction with [γ-^{32}P]dATP and hybridized overnight at 46° in 10% PEG

TABLE 1

Hierarchical sampling scheme for both general and ∇*1* surveys

Block	Locality	Population: Name	Population: Abbreviation	Latitude (°N)	Year	No. of lines studied
North	(Me)	(Cherryfield)	Cf	44.36	(84)	98 (0)
					(86)	51 (0)
	Vt	Edgewood	Ew	43.47	88	60
		Orwell	Orw	43.47	88	59
	Mass	Nagog	Ng	42.30	(86)	58 (0)
					88	59
		Idylwilde	Iw	42.30	88	59
		Very Fine	Vf	42.30	88	60
	Ct	Wesleyan	Wes	41.34	(86)	54 (0)
					88	60 (59)
		Lymers	Ly	41.34	88	59
	NJ	(Sansones)	Ss	40.29	(86)	59 (0)
		Orchardside	Orch	40.21	87	36
		Terhunes	Ter	40.21	86	59
					87a[a]	69 (67)
					(87b[a])	57 (0)
	Md	Maryland	Md	39.00	88	56
South	NC	Raleigh	Ral	35.46	83	60
					84	97
					87	48
	Jax	Jacksonville	Jax	30.20	87	60 (59)
	FP	Fort Pierce	Fp	27.28	87	60 (59)
	Mec	Mecca	Mec	26.42	87	60
	FlaC	Florida City	Fc	25.27	(84)	(75) (0)
					87	60
Totals: 2	11 (10)	17 (15)			25 (18)	1533 (1076)

If a locality, population or year sample is excluded in the ∇*1* survey, the entry is in parentheses. Totals for the ∇*1* survey are given in parentheses where they differ from general survey totals.

[a] Terhunes was sampled twice in 1987: in the spring (a) and in the fall (b).

6000, 253 mM NaCl, 83 mM Na_2HPO_4 (adjusted to pH 7.2 with H_3PO_4), 7% sodium dodecyl sulfate (SDS) (G. CHURCH, personal communication) to the filters. Three 15-min washes were carried out at 60° in 71 mM Na_2HPO_4 (adjusted to pH 7.2 with H_3PO_4), 1% SDS.

***Château Douglas*:** We surveyed individuals from two populations from the extremes of the cline: Edgewood, Vermont (n = 57) and Florida City, Florida (the 1987 sample; n = 60). As *Château Douglas* is associated with *Adh-F*, we sampled Edgewood, a northern population with a high frequency (52%) of *Adh-F*. However, because of its *Adh-S*-like properties, we hypothesize that *Château Douglas* is accordingly disproportionately (with respect to other *F*-haplotypes) represented in southern populations such as Florida City.

By polymerase chain reaction (PCR), we amplified a small segment of DNA encompassing the nucleotide substitution causing the *Château Douglas*-specific amino acid difference. That fragment was then digested with *Bsr*I, a restriction enzyme cutting at the *Château Douglas*-specific site and failing to cut the standard sequence.

PCR was performed using 20-mer primers whose 5′ ends correspond to positions 1475 (forward primer) and 1694 (reverse primer) in KREITMAN's (1983) sequence. Amplification was carried out in 50-μl reaction volumes with one unit of *Taq* polymerase (BRL) under standard conditions (SAIKI *et al.* 1988) with a 2 mM magnesium concentration. The PCR product was phenol/chloroform extracted twice, chloroform extracted, ethanol precipitated and digested overnight in the presence of RNAse with one unit of *Bsr*I in 50-μl reaction volumes. Digests were ethanol precipitated and electrophoresed on an 8% native acrylamide gel before being visualized under UV by ethidium bromide staining. Determinations of 10 lines were confirmed by direct sequencing of PCR product from an amplification with primers whose 5′ ends correspond to positions 1012 (forward primer) and 1759 (reverse primer) in KREITMAN's (1983) sequence. Single-stranded sequencing template was prepared according to the λ-exonuclease method of HIGUCHI and OCHMAN (1989) and the DNA was dideoxy-sequenced (SANGER, NICKLEN and COULSON 1977) using 1012 as the sequencing primer; methods followed BERRY, AJIOKA and KREITMAN (1991).

∇*1*: In order to survey as many chromosomes as possible for ∇*1*, we re-probed the Southern blots with a small fragment containing ∇*1*. ∇*1* consists of a coupled insertion and deletion whose net product is a 4-bp insertion. This is detectable on our original filters (in a 105-bp *Alu*I fragment) but was seldom easy to score because that fragment does not typically probe strongly as the region is AT-rich and we

TABLE 2

Polymorphisms scored

Site no.		Mutation				
		Substitution		Length polymorphism		
Full dataset	Reduced dataset	Gain or loss	Enzyme[b]	Insertion or deletion	Length	Position[a]
1		Loss	*Hae*III			−457
2		Loss	*Msp*I			−415
3	1	Loss	*Msp*I			−339
4		Loss	*Hha*I			−134
5	2			Insertion	20 bp	−20–235
6				Insertion	Large	0–235
7		Gain	*Dde*I/*B*			9–587
8		Loss	*Sau*3AI			79
9	3	Loss	*Hae*III			285
10				Insertion	400 bp	392–409
11				Insertion	150 bp	392–497
11a[c]	3a	Gain	*Alu*I			423
11b[c,d]	3b			Insertion and deletion	34 bp 30 bp	447
11c[c]	3c	Loss	*Alu*I			497
12	4	Loss	*Msp*I			505
13	5			Insertion	35 bp	551
14	6	Loss	*Hha*I			573
15		Loss	*Msp*I			583
16		Loss	*Dde*I/*B*			587
17				Insertion	Large	587–609
18	7	Loss	*Hae*III			686
19	8	Loss	*Hae*III			817
19a[c]	8a	Loss	*Alu*I			1068
20	9	Loss	*Msp*I			1235
21	10	Loss	*Sau*3AI			1351
22	11	Loss	*Hae*III			1421
23	12	*F*/*S*				1490
24	13	Loss	*Dde*I/*B*			1518
25	14	Loss	*Dde*I/*B*			1527
26		Gain	*Dde*I/*B*			1551
27		Loss	*Hae*III			1566
28	15	Loss	*Alu*I			1597
29		Gain	*Dde*I/*B*			1680–2186
30	16	Loss	*Hae*III			1923
31				Deletion	290 bp	1987–2219
32				Insertion	15 bp	1987–2112
33		Loss	*Taq*I			2143
34				Deletion	50 bp	2185
35	17			Deletion	30 bp	2220–2509
36		Gain	*Dde*I/*B*			2259
37	18	Gain	*Taq*I			2331
38	19	Loss	*Taq*I			2348
39		Loss	*Hae*III			2509
40		Loss	*Dde*I/*B*			2548
41		Gain	*Alu*I			2553
42		Gain	*Alu*I			2764
43		Gain	*Dde*I/*B*			2843
44				Duplication		?

[a] Numbering of positions follows KREITMAN (1983). Wherever possible, an exact position has been ascribed to the polymorphism; however, for some insertions/deletions and site gains, the polymorphism has only been mapped to a minimum size fragment and the range of that fragment is given.

[b] *Dde*I/*B* refers to the combination of *Dde*I and *Bam*HI enzymes.

[c] Four sites were only scored in the second part of the study; these were not included in the initial general survey.

[d] $\nabla 1$ site.

were labelling our probe with [α-^{32}P]dCTP. For this reason, we did not score this site in the original study. To overcome these problems, we probed the *Alu*I filters with a PCR-amplified fragment corresponding to the region in the immediate vicinity of $\nabla 1$ labeled with both [α-^{32}P]dCTP and [α-^{32}P]dATP. Amplification was performed as above but with primers whose 5′ ends correspond to positions 364 (forward primer) and 708 (reverse primer) in KREITMAN's (1983) sequence and the product gel-purified using GeneClean (Bio101). Subsequent procedures followed the previous study except for the double labeling of the probe. Using this method, we scored four polymorphic sites, including $\nabla 1$, in addition to those surveyed in the original screen (Table 2). We surveyed 18 populations (Table 1). Despite the exclusions of some population samples, we retained the same sampling structure as that outlined in Table 1; thus all populations south of about 36°N were designated southern and those north of 36°N, northern.

Determinations of $\nabla 1$ were confirmed by direct sequencing of PCR product for 14 lines from Edgewood and 5 from the 1987 Florida City collection. The same region as that amplified for the probe was amplified and sequenced for each line in the same way as above, using 364 as the sequencing primer.

Analysis: We analyzed the data on both site-by-site and haplotype-by-haplotype bases, carrying out G_{ST} and linear regression analyses. G_{ST} (NEI 1973), a modification of WRIGHT's (1951) fixation index, F_{ST}, measures the contribution of differences between samples to overall genetic diversity. G_{ST} does not take into account differences in sample size. However we were interested only in the *relative* amount of differentiation between populations at different polymorphic restriction sites (*e.g.*, whether there is greater differentiation at the *F/S* site than at flanking silent sites) and the sample sizes are the same for all polymorphisms. To decompose the total genetic differentiation into the separate contributions of different levels in our sampling hierarchy (Table 1), we also performed a hierarchical G_{ST} analysis (CHAKRABORTY 1980).

G_{ST} analysis does not incorporate information on the spatial distribution of populations. We therefore also carried out regression analysis on the grounds that, if environmental factors affecting *Adh-F/S* frequencies correlate with latitude, then a latitudinal cline will conform broadly with a linear model.

The statistical significance of our results was determined by the Monte Carlo and randomization procedures described below. These analyses entailed the generation of a simulated dataset for which the test statistic was then calculated; this was iterated 1000 times and the values sorted to determine the appropriate confidence limits.

Monte Carlo simulation of hierarchical G_{ST} analysis for polymorphic sites: To gauge the confidence intervals of the site-by-site hierarchical G_{ST} analysis under the sampling scheme outlined in Table 1 (CHAKRABORTY 1980), G_{ST} values were calculated for simulated datasets allowing for binomial sampling at each site in each population.

Regression Monte Carlo simulation: After calculating the squared correlation coefficient, r^2, for the regression on latitude for each site, we investigated the extent to which linkage disequilibrium between one site and another could affect the regression of the second site. The premise of this analysis is that the frequencies of all sites are governed solely by their linkage relationships over the whole sample with a single specified site. Although linkage disequilibrium values are only meaningful within a single population, our results (see below) show gene frequencies to be similar among populations so that it is reasonable to treat the whole sample as being derived from a single population for the purpose of estimating linkage disequilibrium. By pooling all the populations, we gain considerable statistical power (*i.e.*, we are increasing our sample size from about 60 (a single population) to over 1500). In this analysis, we kept constant both the observed linkage disequilibrium value and the frequency in each population of the "governing" locus. The frequency of the "affected" locus in each population is simulated with binomial sampling about its expected value. If we take site A to be governing the frequencies of all other sites, and, across the entire pooled sample, site B is present in 20% of individuals in which A is present and in 30% of individuals in which A is absent, and A is present in 40 out of 60 individuals in population 1, the expected frequency of B in population 1 is $(0.2 \times 40) + (0.3 \times 20) = 14$. We simulated an entire set of new frequencies for every site in every population and computed the r^2 value of the linear regression of the frequency of each site in each population on population latitude. After 1000 iterations the values were sorted to ascertain the confidence limits for the expected r^2 value. Each site in turn was assumed to be the "governing" locus, so the analysis was repeated for each successive site.

Randomization analysis of hierarchical G_{ST} analysis for haplotypes: Having calculated the hierarchical G_{ST} values for all haplotypes, we investigated the statistical significance of the observed values against an expectation of panmixis in the form of a randomization analysis carried out on pooled data appropriate to each level of the hierarchy. Therefore, a separate analysis was carried out at each level of the hierarchy. At the highest level, the north-south dichotomy, all individuals were pooled and each population resampled without replacement. G_{ST} values were then computed (1000 replicates) and sorted to derive confidence intervals. At each lower level, the appropriate populations were resampled from the applicable subset of the data. For example, at the lowest level, that of resampling of populations between years, the samples for population A that had been multiply sampled n times were pooled and the n samples reconstituted at random from the pooled data. This method leaves G_{ST} values for levels higher in the hierarchy unchanged.

Sample shuffling of hierarchical G_{ST} analysis for haplotypes: The above randomization test is a progressively less stringent test as we descend the sampling hierarchy. It assumes that haplotypes are identically distributed in the two compared samples, so at the top level (north *vs.* south) we assume homogeneity from Maine to Florida, whereas, lower down the hierarchy, we assume homogeneity between samples from, say, Terhunes and Sansones (both Princeton, New Jersey, area farms). In order to reduce the stringency of our test of differentiation between north and south (*i.e.*, our test of the cline) we carried out a "sample-shuffling" analysis in which entire localities were shuffled at random (without replacement) without altering the constitution of each locality. In other words, everything is fixed except for the latitude of the localities. G_{ST} values for the north-south dichotomy were calculated (1000 replicates) and confidence intervals derived. A problem with this analysis is the differing number of individuals in each locality: if, in a simulation, all the localities with relatively few individuals ended up by chance in one block, we might expect to see an enhanced G_{ST} value because of the imbalance between sample sizes. We tested for this by plotting the G_{ST} value against the difference in sample size for each block, and found no trend toward higher G_{ST} values with greater differences between block sample sizes (data not shown). This suggests that our sample sizes are sufficiently large that the most skewed possible distribution of sample sizes between blocks has no significant effect on the resultant G_{ST} value.

We also used this approach to investigate the effect of geography on population structure without the confounding effects of local variation, by shuffling populations at random (without replacement) between localities within the north and south blocks. To do this, we removed the lowest level of the hierarchical sampling scheme by combining all year-to-year samples of a population into a single pooled sample for that population. We computed the G_{ST} values and carried out 1000 replicates to determine the confidence limits of those values.

Significance testing of regressions on latitude for individual haplotypes: We calculated the r^2 values for the linear regressions on latitude of each haplotype within an allelic class (*e.g.*, *Adh-S* or *Adh-F*). To test the statistical significance of these regressions, we chose to carry out simulations because of changes in sample size correlated with latitude (*e.g.*, for haplotypes carrying the *Adh-F* site because of the decline in frequency of *Adh-F* with decreasing latitude). Given our finding of a lack of differentiation among populations and assuming that all haplotypes within the *Adh-F* allelic class are neutral equivalents, we expect to see no cline of haplotypes within *Adh-F*. A Monte Carlo simulation (1000 replicates) yields an estimate of binomial sampling error on the computed statistic, the r^2 value of the linear regression of haplotype frequency on latitude. The results were sorted to provide confidence limits.

RESULTS

We will first present the results of the general survey of all 25 samples before moving on to the *Château Douglas* and $\nabla 1$ data.

General survey

This study includes data from four populations previously published by KREITMAN and AGUADÉ (1986) (Raleigh 83) and SIMMONS *et al.* (1989) (Raleigh 84, Florida City 84, Cherryfield 84). From 1533 lines from 25 populations, we identified 113 different haplotypes from 44 variable sites (see APPENDIX). A number of the polymorphisms were rare (Table 3) and their presence yields unique or very low frequency haplotypes. In some of the ensuing analysis, all sites that vary in fewer than 10 of the 25 populations have been excluded on the basis of their minimal information content. The only excluded variable site of overall frequency greater than 1%, loss *Msp*I 583, occurs 18 times (*i.e.*, frequency 1.17%) in eight populations. Its occurrence in both Vermont and Florida populations demonstrates its lack of regional endemism. The reduced dataset consists of 19 sites (Table 2) yielding 87 haplotypes. We specify for each result which dataset was used; henceforth site numbers refer to the reduced dataset.

Although we cannot be certain that a given RFLP site loss corresponds to the appropriate mutation detected by KREITMAN'S (1983) sequencing study, we assume that they are the same. We also do not know the exact positions of a number of rare polymorphisms (Table 2) but all of the 19 highly variable (and therefore informative) sites correspond to polymorphisms identified in the previous studies listed above. None of these polymorphisms, except the *Fast/Slow* site, affect the amino acid sequence of the protein and they are therefore effectively silent. Although it is possible that some of the previously undescribed low frequency polymorphisms are not synonymous, we conclude that the vast majority of polymorphism detected, other than *Fast/Slow*, is effectively silent. Three other amino acid polymorphisms have been documented for *D. melanogaster* ADH (COLLET 1988; LAURIE, BRIDGHAM and CHOUDHARY 1991), but the enzymes used in this study did not detect them. We show below that the most common of these variants, *Château Douglas*, is rare in two surveyed populations.

Figure 1 shows the distribution of the 13 haplotypes most common in the whole sample in each of the populations. Visually striking is not only the similarity between closely related or geographically contiguous populations (*e.g.*, Florida City 84 and Florida City 87) but also the consistency throughout the entire sample: haplotypes that are common in the south are typically also common in the north. There is little suggestion of endemism of haplotypes, although certain haplotypes (such as haplotype *Wa-S* (see below) in Cherryfield 84) are apparently overrepresented in some populations. Figure 2 shows the *F/S* cline on latitude. Consistent with VIGUE and JOHNSON'S (1973) original description of the cline, we see an increase in variance of frequency among populations in the north.

The results for the first part of this study, the general RFLP survey, are presented as follows: in order to investigate the structure of the data *vis-à-vis* overall neutrality we perform a Watterson test of haplotype frequencies. Next we analyze first the individual polymorphic sites and then the distribution and relatedness of haplotypes. Our analysis has three major themes. First we want to compare patterns of differentiation at different polymorphic sites: if clinal selection is acting on only one site, then we expect to see clines only at that site and possibly at sites in linkage disequilibrium with it. Second, we are interested in decomposing the total genetic variation to determine, for example, the contribution of among-population differences to the total. This yields information on population structure pertinent to the selection issue: a neutral cline cannot persist in the face of extensive gene flow. Finally, having found evidence of selection, we apply the "selective equivalence" approach outlined earlier to investigate the possible targets of selection.

Neutrality of the data–Watterson test: We carried out a Watterson test (WATTERSON 1978; algorithm from MANLY 1985) of neutrality on the observed overall distribution of haplotypes (Figure 3) and found a departure from neutrality (observed homozygosity = 0.0753; $P < 0.01$). On the basis of HUDSON'S (1983)

TABLE 3

Frequencies of polymorphic sites in each population

	Site no.																					
Population	1	2	3	4	5	6	7	8	9	10	11	11a	11b	11c	12	13	14	15	16	17	18	19	19a
cf84	1	0	0.93	1	0.05	0	0	1	0.7	0	0				0.99	0.12	0.96	1	1	0	0.97	0.23	
cf86	1	0	0.8	1	0.04	0	0.02	1	0.82	0	0				1	0.04	1	1	1	0	1	0.24	
orw88	1	0	0.81	1	0	0	0	0.97	0.97	0	0	0.07	0.51	1	0.97	0.2	0.97	1	1	0	0.97	0.34	0.97
ew88	1	0	0.83	1	0	0	0	1	0.98	0	0	0.08	0.48	0.98	0.95	0.12	0.98	0.9	1	0	0.98	0.3	0.98
ng86	1	0	0.91	1	0.02	0	0	1	0.97	0	0				1	0.33	0.98	1	1	0	1	0.5	
ng88	1	0	0.56	1	0.05	0	0	1	0.97	0	0	0	0.32	1	0.97	0.07	1	1	1	0	1	0.19	0.95
iw88	1	0	0.73	1	0.03	0	0	1	1	0	0	0.02	0.44	0.97	0.98	0.08	0.97	0.97	1	0	0.97	0.24	0.97
vf88	1	0	0.8	1	0.02	0	0	1	0.95	0	0	0	0.43	0.92	0.98	0.07	0.95	1	1	0	0.95	0.25	0.92
wes86	1	0	0.85	1	0.09	0	0	1	0.87	0	0				1	0.06	0.91	0.91	1	0.02	0.93	0.3	
wes88	1	0	0.88	1	0.07	0	0	1	0.95	0	0	0.03	0.53	0.95	0.95	0.27	0.98	1	0.98	0	0.98	0.45	0.95
ly88	1	0	0.86	1	0.05	0	0	1	0.95	0	0	0.03	0.29	0.93	1	0.17	0.97	1	1	0	0.98	0.32	0.95
ss86	1	0	0.83	1	0.07	0	0	1	0.98	0	0				1	0.2	0.98	1	1	0	0.98	0.39	
orch87	1	0	0.86	1	0.11	0	0	1	0.94	0	0	0.03	0.17	0.92	1	0.11	0.92	1	1	0	0.92	0.47	0.94
ter86	1	0	0.75	1	0.05	0	0	1	0.95	0	0	0.05	0.27	0.97	1	0.12	0.97	1	1	0	0.95	0.37	0.97
ter87a	1	0	0.58	1	0.22	0	0	1	0.99	0	0	0.01	0.16	0.93	1	0.07	0.93	1	1	0	0.93	0.25	0.98
ter87b	1	0	0.63	1	0.12	0.02	0	1	1	0	0				1	0.11	0.93	1	1	0	0.93	0.42	
md88	1	0	0.84	1	0	0	0	1	0.98	0.02	0	0.02	0.3	0.96	1	0.16	0.98	0.98	1	0	0.98	0.34	0.98
ra83	1	0	0.85	1	0.15	0	0	1	1	0	0	0.03	0.23	0.9	0.97	0.05	0.92	0.98	1	0	0.93	0.42	0.98
ra84	1	0	0.78	0.99	0.05	0	0	1	0.95	0	0	0.02	0.22	0.94	0.99	0.16	0.96	0.99	1	0	0.96	0.43	0.96
ra87	0.98	0	0.71	0.98	0.02	0	0	1	0.96	0	0	0.04	0.12	0.94	1	0.06	0.94	0.98	1	0	0.94	0.35	0.94
jax87	1	0.02	0.7	1	0.07	0	0	1	0.9	0	0	0.02	0.1	0.92	1	0.05	0.97	1	1	0	0.97	0.37	0.88
fp87	1	0	0.75	1	0.08	0	0	1	0.92	0	0.02	0	0.03	0.92	1	0	0.95	1	1	0	0.95	0.4	0.93
mec87	1	0	0.83	1	0.05	0	0	1	0.9	0	0	0	0.05	0.98	1	0.03	0.93	0.98	1	0	0.93	0.28	0.9
fc84	1	0	0.65	1	0.08	0	0	1	0.85	0	0				0.87	0.03	0.97	1	1	0	0.97	0.28	
fc87	1	0	0.87	1	0.05	0	0	1	0.85	0	0	0	0.05	0.97	1	0.05	0.98	1	1	0	0.98	0.38	0.83

Frequency given is that of the "1" allele (*i.e.*, site presence).

simulation study, we consider this unlikely to be due to recombination. Our data show an apparent excess of intermediate to high frequency haplotypes, suggesting the possible action of balancing selection, and a concomitant reduction in the expected number of low frequency haplotypes, suggesting purifying selection against low frequency variants. However, as the test assumes population genetic equilibrium and, as discussed by KEITH (1983), departures from the neutral expectation are attributable to a wide range of causes, our result is merely suggestive.

Site-by-site analysis: Whereas we expect historical aspects of population structure to have the same effect on patterns of geographic variation at all variable sites, we expect selection only to affect those sites specifically targeted by selection (and those linked to them). Site-by-site analysis therefore permits us to distinguish between selective and historical explanations of the cline.

Analysis of differentiation among populations: G_{ST} analysis of all the sites (Figure 4) shows a considerable range in the extent to which sites are differentiated between populations. Immediately striking is that, of all the 44 sites tested, the *F/S* polymorphism shows the most differentiation between populations.

At what levels are the populations differentiated? For example, is the observed pattern a product of extensive gene frequency differences between years, or is it a function of geographic distance? To answer these questions, we carried out a hierarchical G_{ST} analysis (CHAKRABORTY 1980) using the sampling structure outlined in Table 1. As can be seen from Figure 5, the relatively high G_{ST} value for the *F/S* site is largely attributable to the north/south dichotomy, as would be expected if it is varying clinally. Only one other site, site 5 (a 35-bp insertion corresponding to KREITMAN's (1983) ∇2), shows a similar pattern of extensive north/south differentiation, though this is not as marked as for the *F/S* site. At the next level, between localities within respective north/south blocks, we see differentiation of the sites that uniquely comprise haplotype *Wa-S* (see below), which is probably largely due to its high frequency in the Cherryfield 84 sample. Intuitively, this might be expected to result in north/south differentiation of these sites (Cherryfield is the most northerly population sampled). However, *Adh-S* haplotypes, of which *Wa-S* is one, are proportionally more common in the south (see below), so *Wa-S*'s commonness in Cherryfield is counterbalanced in the south by its generally high frequencies, leading to no overall discrepancy in its frequencies between north and south. At lower levels in the hierarchy (*i.e.*, between populations within localities, and between years for single populations) the sampling variances, as revealed by Monte Carlo simulation, are large and do not permit the detection of any significant differentiation.

Regression analysis–variation and latitude: If environmental factors affecting *Adh-F/S* frequencies correlate with latitude, then it is reasonable to expect a latitu-

20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
1	0.69	0.7	0.33	0.55	0.44	0	1	0.89	0	0.95	0	0	1	0	0.3	0	0	0.96	1	1	0	0	0	0
1	0.82	0.82	0.27	0.69	0.31	0	1	0.88	0	0.92	0	0	1	0	0.18	0	0	0.92	1	1	0	0	0	0
1	0.97	0.97	0.54	0.76	0.24	0	1	0.83	0	1	0	0	1	0	0.03	0	0.02	0.98	1	1	0	0	0	0
0.97	0.98	0.98	0.52	0.82	0.18	0	0.98	0.85	0	0.98	0	0	1	0	0.02	0	0.03	0.95	1	1	0	0	0	0
0.98	0.9	0.97	0.45	0.71	0.29	0	0.98	0.86	0	0.93	0	0	1	0	0.03	0	0	0.93	1	1	0	0	0	0
0.98	0.95	0.95	0.36	0.54	0.46	0	1	0.58	0	1	0	0	1	0	0.05	0	0.03	0.97	1	1	0	0	0	0
0.98	0.97	0.97	0.47	0.66	0.34	0	0.98	0.75	0	0.98	0	0	1	0	0.03	0	0	0.98	1	1	0	0	0	0
1	0.92	0.92	0.43	0.68	0.32	0	1	0.77	0	0.92	0	0	1	0	0.08	0	0.02	0.9	0.98	1	0	0	0	0
0.98	0.87	0.87	0.17	0.63	0.37	0	0.98	0.8	0.02	0.89	0	0.02	1	0	0.15	0	0.02	0.91	1	1	0.02	0.02	0	0
1	0.95	0.95	0.53	0.83	0.17	0	1	0.9	0.02	0.95	0.02	0.02	1	0	0.05	0	0.02	0.95	1	1	0.02	0.02	0	0
1	0.95	0.95	0.34	0.76	0.24	0	1	0.8	0.02	0.92	0	0.02	1	0	0.05	0	0.03	0.9	1	1	0.02	0.02	0	0
0.97	0.98	0.98	0.42	0.83	0.17	0	1	0.88	0	0.98	0	0	1	0	0.02	0	0.02	0.98	1	1	0	0	0	0
1	0.94	0.94	0.22	0.81	0.19	0	0.97	0.92	0	0.92	0	0	1	0	0.14	0	0	0.92	1	1	0	0	0	0
0.98	0.97	0.97	0.29	0.64	0.36	0	1	0.73	0.02	0.95	0	0.02	0.98	0	0.05	0	0	0.98	1	1	0.02	0.02	0	0
0.99	0.99	0.99	0.19	0.65	0.35	0	1	0.71	0	0.91	0	0	1	0	0.09	0	0	0.93	1	1	0	0	0	0
1	1	1	0.25	0.68	0.32	0	0.98	0.74	0	0.91	0	0	1	0	0.07	0	0	0.91	1	1	0	0	0	0
0.91	0.98	0.98	0.39	0.73	0.27	0.02	1	0.79	0	0.95	0	0	1	0	0.02	0	0.04	0.93	1	0.98	0	0	0	0
0.97	1	1	0.3	0.8	0.2	0	0.97	0.85	0	0.88	0.03	0	1	0	0	0	0	0.9	1	1	0	0	0	0
0.99	0.96	0.96	0.32	0.67	0.32	0	0.99	0.78	0	0.9	0	0	1	0.01	0.04	0	0.01	0.95	1	1	0	0	0	0
1	0.94	0.94	0.15	0.52	0.48	0	0.98	0.65	0	0.9	0	0	1	0	0.06	0	0	0.94	1	1	0	0	0	0
1	0.88	0.88	0.17	0.58	0.42	0	1	0.7	0	0.92	0	0	1	0	0.12	0	0.03	0.95	1	1	0	0	0	0
1	0.92	0.92	0.15	0.63	0.37	0	1	0.78	0	0.92	0	0	1	0	0.08	0	0	0.95	1	1	0	0	0.02	0.02
1	0.9	0.9	0.1	0.6	0.4	0	1	0.75	0	0.95	0	0	1	0	0.08	0.02	0	0.97	1	1	0	0	0	0
1	0.87	0.87	0.16	0.56	0.44	0	1	0.72	0	0.96	0	0	1	0	0.13	0	0	1	1	1	0	0	0	0
1	0.83	0.83	0.15	0.63	0.37	0	1	0.85	0	0.97	0	0	1	0	0.17	0	0.02	0.97	1	1	0	0	0	0

dinal cline to conform broadly to a linear model. Hierarchical G_{ST} analysis revealed north-south differentiation of the *F/S* site; this is consistent with its varying clinally. As a measure of "clinality" (for want of a better term), we have taken the r^2 value of the regression of frequencies on latitude; this is essentially a measure of the proportion of the total variance in frequencies that is explained by latitude. How does the r^2 value for *F/S* compare to that for other sites? Figure 6 is similar to the north-south comparison in the hierarchical G_{ST} in showing marked "clinality" at the *F/S* site, some at site 5, and very little at other sites.

What is striking about both Figure 6 and the top panel of Figure 5 is the apparent lack of hitchhiking. We expect physically close sites to be in strong linkage disequilibrium with the *F/S* site, so, if the *F/S* site is under clinal selection, as this analysis suggests, why do we not see the same "clinality" at contiguous sites? We speculated initially that the sites might, despite their proximity, be in linkage equilibrium.

Linkage disequilibrium analysis: We elected to use LEWONTIN's (1964) measure of linkage disequilibrium, D', because it corrects for differences in frequencies at the two loci under study and yields a measure between -1.0 and 1.0 where 0.0 implies linkage equilibrium. In view of the contribution of different gene frequencies in subdivided populations to a measure of linkage disequilibrium, it should strictly only be calculated for individual populations. For this reason, we calculated D' for all pairs of sites for each population and then calculated an unweighted average for each site pair. As the probability of recombination between two loci increases with the distance between them, we expect to see a decline in linkage disequilibrium with distance. Figure 7 shows no such trend.

We increased the statistical power of our analysis by pooling all populations and treating the whole sample as a single population. This is justified by the observations that (i) the D' values calculated are close to those derived from the above analysis which took the mean value for each of the separate populations (data not shown) and (ii) the G_{ST} values for the data are low, suggesting that all *D. melanogaster* on the east coast can reasonably be considered to belong to a single population. We excluded all pairwise comparisons in which one of the variants at one of the sites had a frequency of less than five in order to remove a large number of comparisons which often yield high values of D' because of their low frequencies. Furthermore, we included only those values of D' which differ significantly (*i.e.*, $P < 0.05$ by G test of 2×2 contingency table) from 0.0. As before, linkage disequilibrium does not decline with physical distance (data not shown). The data do not permit us to distinguish between the two alternative explanations for the lack of a decline with distance of linkage disequilibrium: either the disequilibrium is selectively maintained or the region studied is too small for us to detect the effect of increasing recombination distance on disequilibrium. There is extensive linkage disequilibrium between the sites in the region in general and the *F/S* site is no exception.

Hitchhiking and latitudinal clines: Given the high observed levels of linkage disequilibrium between the

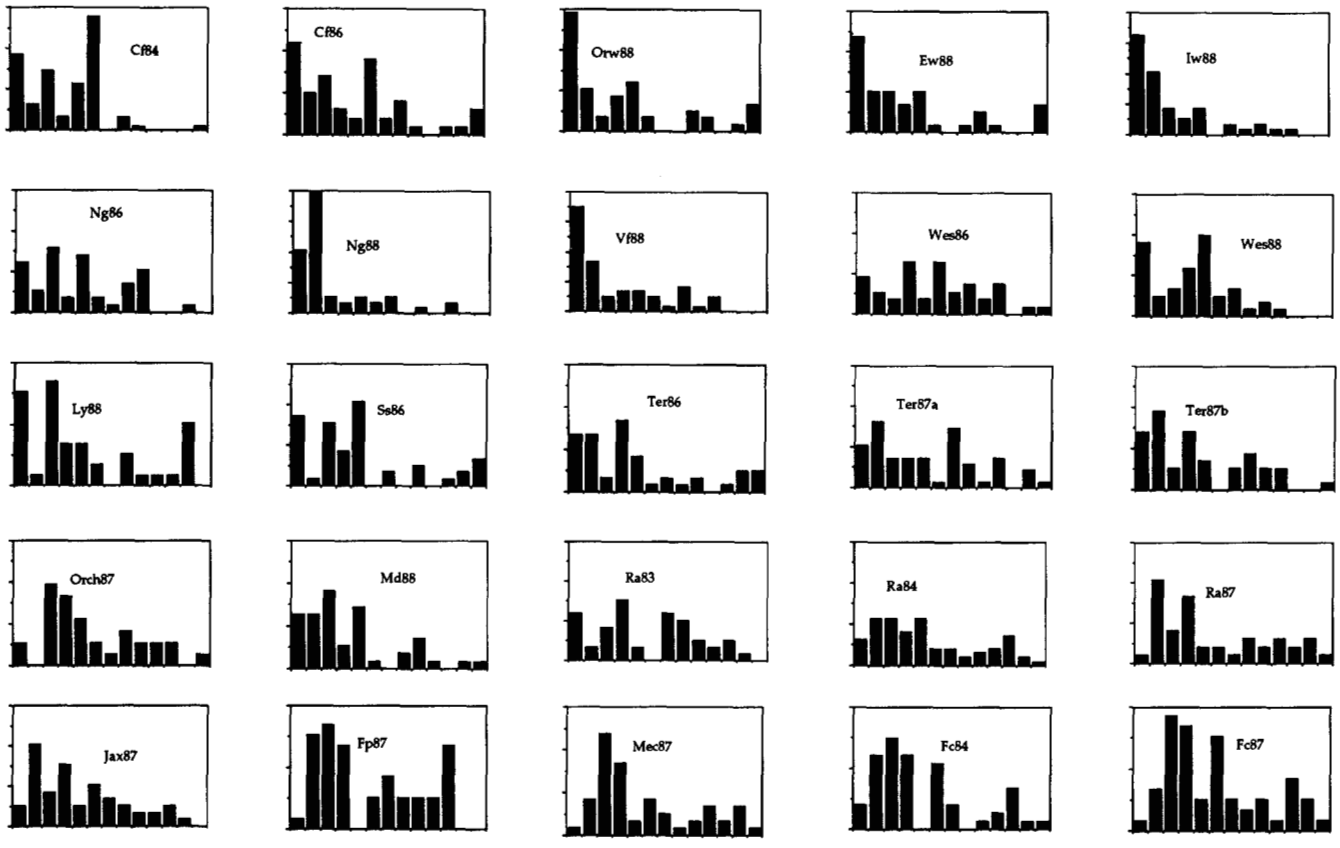

FIGURE 1.—Frequencies of each of the 13 most common haplotypes in the pooled sample in each separate collection. Plots are placed from left to right and down the page in order of descending latitude (*i.e.*, the top left plot is for the most northerly population).

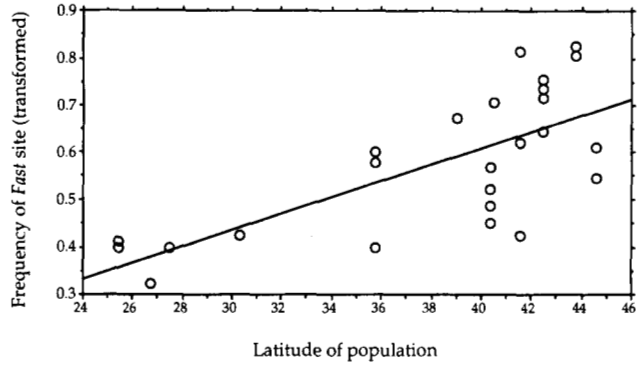

FIGURE 2.—Regression of the frequency of *Adh-F* (square-root, arcsine transformed) on the latitude of each sample; $r^2 = 0.511$.

F/S site and flanking polymorphism, why do we see a lack of hitchhiking in the regression analysis? To determine whether the observed amounts of "clinality" at each site can be explained in terms of linkage disequilibrium to the *F/S* site (or other sites) we carried out a Monte Carlo simulation in which the frequencies of each site are governed by its linkage disequilibrium with a specified site, the observed frequencies of the specified site in each population, and a binomial sampling process. Figure 8a shows that when a typical "nonclinal" site (in this case, site 1) is specified to govern the frequencies of all other sites, it does not, as expected, explain the observed "clinality" at the *F/S* site, nor, we see in Figure 8b, does the only other clinal site (site 5). However, when the *F/S* site is set to govern the frequencies of all the other sites (Figure 8c), we see that it can explain the observed clinality (or lack of it) in all sites included in this survey. Note that there may be polymorphic sites in the *Adh* region not included in this study whose latitudinal variation cannot be explained by linkage to the *F/S* site (see below). This result implies that, contrary to intuition based on linkage disequilibrium, the observed latitudinal pattern can be explained by assuming that all frequencies are driven by the *F/S* site frequencies.

Haplotype-by-haplotype analysis: Because of the high haplotypic diversity of our sample, haplotype

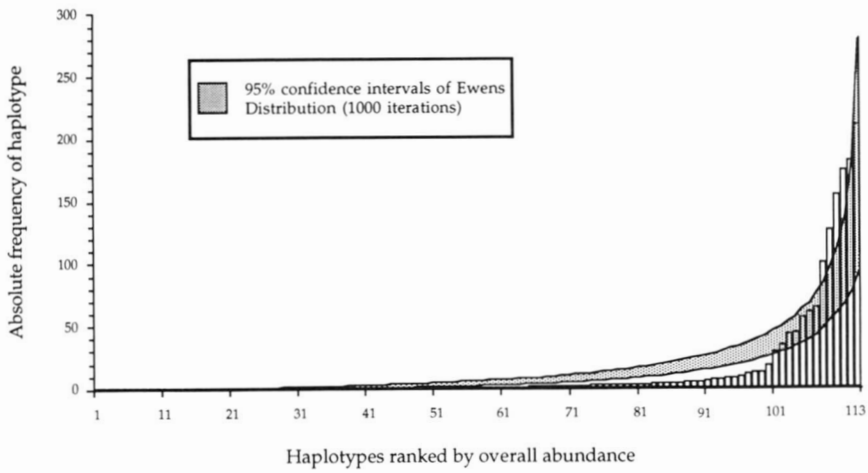

FIGURE 3.—Observed distribution of haplotypes and 95% confidence limits of the Ewens distribution for 113 haplotypes distributed among 1533 individuals.

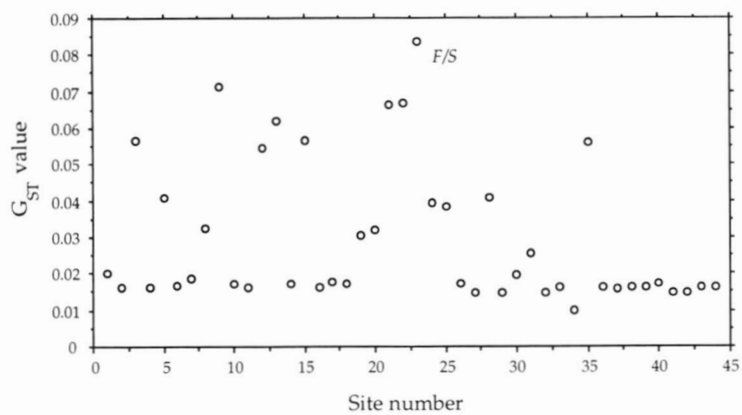

FIGURE 4.—G_{ST} values (NEI 1973) between all samples for each polymorphic site.

analysis is a powerful way to investigate population structure and can also provide information on which polymorphisms are under selection.

Hierarchical G_{ST} analysis of haplotypes–differentiation among populations: Under the sampling scheme outlined in Table 1, we calculated hierarchical G_{ST} values (CHAKRABORTY 1980) from all haplotypes by treating each haplotype as a separate allele at a single locus. We also carried out the same calculations for haplotypes carrying *Adh-S* and those carrying *Adh-F* (*S*- and *F*-haplotypes respectively) because, if selection is acting on the *F/S* site as the above regression analysis suggests, we expect to see no selection-derived latitudinal structure to the distribution of haplotypes within an allelic class as all members of each class are neutral equivalents. Only history-derived differentiation will remain. Therefore, if selection is acting only on the *F/S* site, we expect similar patterns of haplotype differentiation between allelic classes. Out of the several possible ways of testing the significance of the distributions, we have chosen to test each distribution against a stringent null hypothesis of panmixis because this not only performs the required comparison but also yields useful information about population structure.

The results are presented in Table 4. Statistical testing for departure from panmixis was carried out at two levels of stringency. Both tests reveal heterogeneity between north and south within the whole dataset (expected, given the cline) and within the *F*-haplotypes, but not within *S*-haplotypes. Indeed the overall uniformity of the latter is remarkable given that *S*-haplotypes (n = 1061) account for over two thirds of the total sample. Deviations from panmixis also occur at lower levels of the hierarchy, suggesting local substructuring to *D. melanogaster* populations.

How can we reconcile a lack of differentiation among *S*-haplotypes between north and south with evidence of local heterogeneity (*e.g.*, among localities within the north block)? This pattern suggests stochasticity in the founding of individual populations but, overall, a lack of divergence between north and south implying large scale movement. In order to investigate local population structure in more detail, we carried out a sample-shuffling analysis in which we randomly moved populations between localities within the north block (localities were not sampled in the south). As localities differ in latitude whereas all populations within a locality share a similar latitude, the cline predicts homogeneity within a locality and differentiation between localities. Thus the cline obscures the possible effects of population structure.

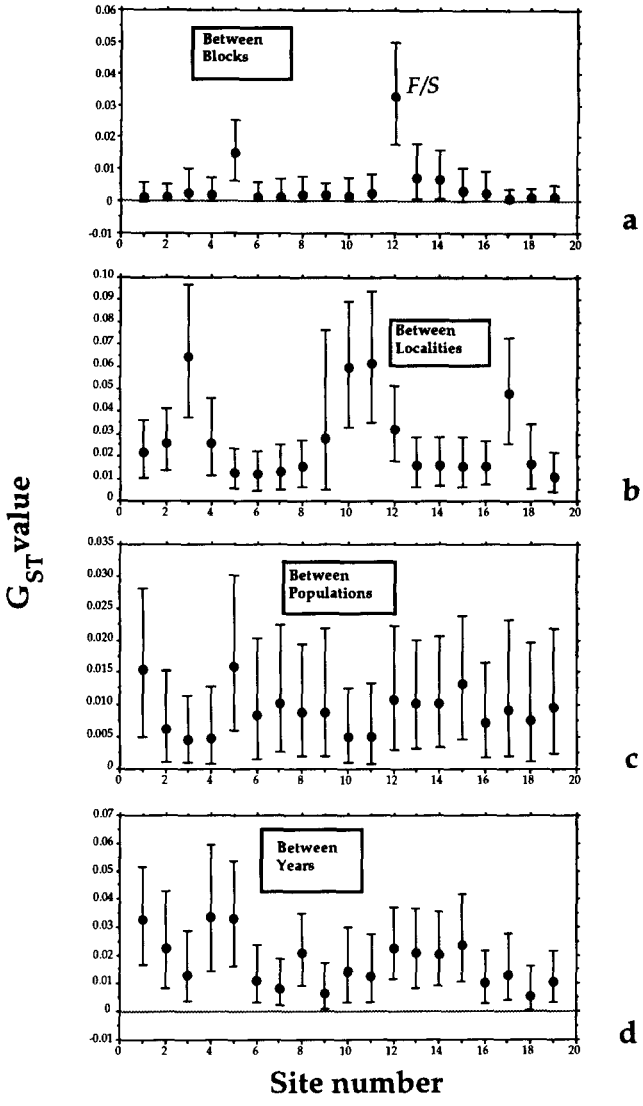

FIGURE 5.—Hierarchical G_{ST} values (CHAKRABORTY 1980) for each polymorphic site (reduced dataset) under the sampling scheme of Table 1. The 95% confidence intervals were derived by Monte Carlo simulation (see text for details).

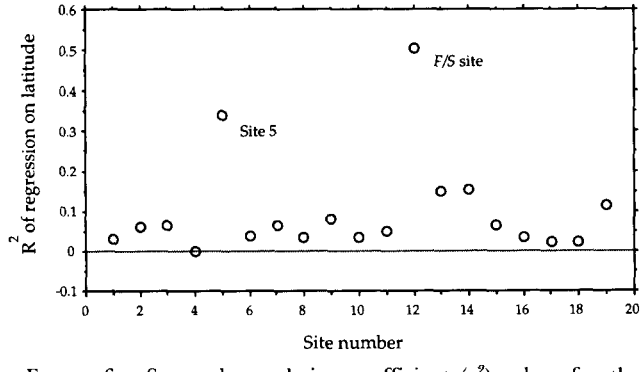

FIGURE 6.—Squared correlation coefficient (r^2) values for the regressions on latitude for the frequency of each site (reduced dataset) (frequency square-root, arcsine transformed) in each sample.

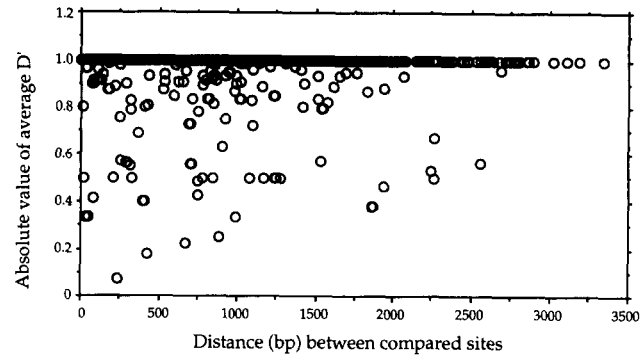

FIGURE 7.—Absolute value of the mean over 25 samples of D' (LEWONTIN 1964) for each pairwise comparison of polymorphic sites plotted against the distance between the compared sites.

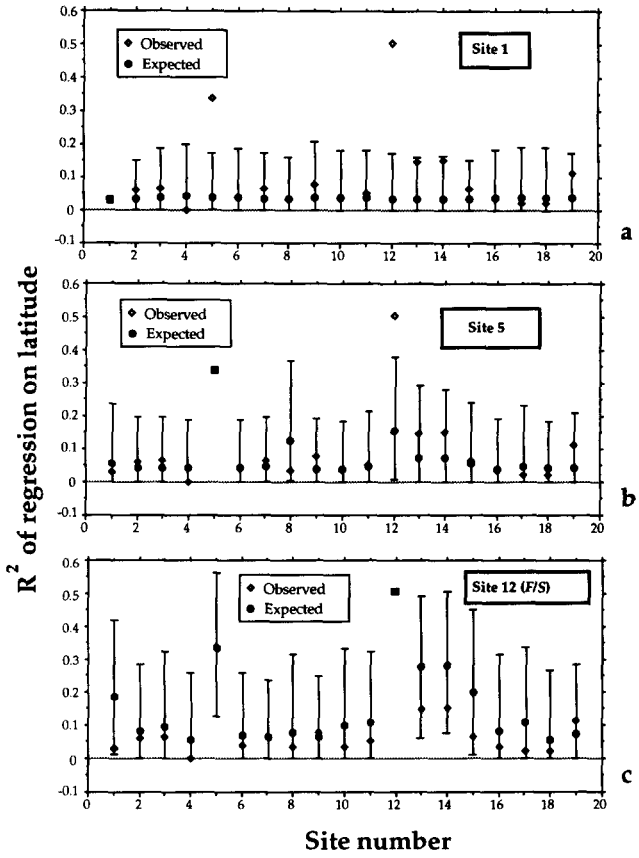

FIGURE 8.—Results of regression simulation (see text for details) for reduced dataset. Each plot shows the observed data and the expected values with 95% confidence intervals when a different site is held constant and is assumed to be governing frequencies at all other sites: (a) site 1, (b) site 5 and (c) site 12 (*F/S* site).

However, the above results show there to be no latitudinal structure to the *S*-haplotype dataset, so it is reasonable to assume that deviations from expectation under the population-shuffling model for this subset of the data are indeed caused by population subdivision independent of the cline. For *S*-haplotypes, we see greater than expected homogeneity within localities (observed G_{ST} value: 0.0122; $P < 0.01$) and greater than expected heterogeneity among localities (observed G_{ST} value: 0.0300; $P < 0.01$). This supports

TABLE 4

Values for hierarchical G_{ST} performed on haplotypes

Dataset analyzed	Between blocks	Between localities	Between populations	Between years
All haplotypes	0.0081**††	0.0161**	0.0079**	0.0094**
F-haplotypes	0.0675**††	0.0482*	0.0178	0.0221
S-haplotypes	0.0016	0.0262**	0.0114*	0.0154**

The sampling scheme is the same as in Table 1. Three separate datasets were used: all haplotypes, just *Adh-F*-bearing haplotypes and *Adh-S*-bearing haplotypes. Statistical significance with respect to a null hypothesis of panmixis was determined by appropriate randomization simulation (* $P < 0.05$; ** $P < 0.01$) or, for the between-blocks comparison only, by sample-shuffling simulation († $P < 0.05$; †† $P < 0.01$). See text for details.

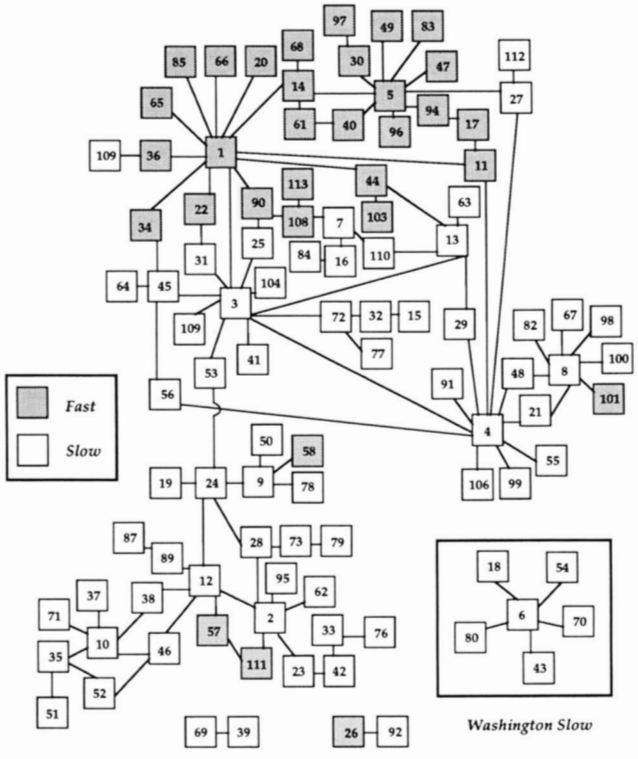

FIGURE 9.—Haplotype network showing all haplotypes related to at least one other haplotype by a single difference. Each haplotype's designated number corresponds to its frequency rank abundance in the total sample (#1 is the most common).

a model in which Northern populations are founded and local migration is sufficient to homogenize gene frequencies within, but not between, localities.

Relationships among haplotypes: Standard methods of phylogenetic reconstruction such as UPGMA are inapplicable to recombining DNA. We have therefore followed SIMMONS *et al.* (1989) in constructing a haplotype network from a difference matrix for all haplotypes (Figure 9). This shows all single step links between the haplotypes. Like SIMMONS *et al.* (1989), we see that *F*- and *S*-haplotypes cluster together, and that *F*-haplotypes are less diverse than *S*-haplotypes, suggesting a relatively recent origin of the *Adh-F* allele (KREITMAN 1983). Most haplotypes are closely connected to the network; even those that are not shown in the network because they lie more than a single step away from their closest relative are usually only two or, at most, three steps away from another haplotype. However, one cluster of haplotypes (numbers 6, 18, 43, 54, 70 and 80) is some at distance from the rest of the network. These carry at least three of a group of four differences (site numbers 3, 10, 11 and 17) which distinguish them from all other haplotypes. Three other haplotypes (numbers 74, 105 and 107) also fall into this class but are not shown on the haplotype network (Figure 9) because each one differs by more than a single polymorphism from its closest haplotype relative. These highly divergent haplotypes are derived from the *Washington-Slow* (*Wa-S*) haplotype originally identified by KREITMAN'S (1983) sequencing study. We also note evidence of recombination from the presence of all four gametic types at many pairs of sites; these show up in the network as squares connecting four haplotypes (SIMMONS *et al.* 1989).

Château Douglas and ∇*1* surveys

***Château Douglas*:** This only occurred twice in the Edgewood population; it did not occur in the Florida City sample. These determinations were confirmed by the sequence analysis. We therefore estimate it to be rarer overall (2%) than was found in previous studies (5%: WILKS *et al.* 1980; SAMPSELL and STEWARD 1983) and it is accordingly unlikely to be responsible for the heterogeneity observed among *F*-haplotypes.

∇*1*: We scored four polymorphic sites (Table 2), including ∇*1*, in addition to those scored in the general survey. This information was added to the available data from the previous study. Our 18 populations (a total of 1076 individuals) yield 94 haplotypes from a total of 48 sites. However, the exclusion of seven populations has resulted in three of the sites being monomorphic among our 18 populations (sites 6, 7 and 17) so in fact only 45 sites contribute. As before, for convenience, we have excluded many of the low frequency sites from much of the analysis: by removing all sites excluded in the general survey and adding the four new sites, we have 23 sites and 75 haplotypes (data available from authors).

Figure 10 shows ∇*1* to vary clinally with latitude. In fact, for the ∇*1* dataset, latitude explains more of the variation in frequencies of ∇*1* than of the *F/S* site (r^2 values for the linear regression on latitude are 0.786 and 0.655 for ∇*1* and *F/S*, respectively). Regression simulations like the above show that the cline in ∇*1* frequencies cannot be explained in terms of its linkage to the *F/S* site (Figure 11b), whereas the converse is feasible: the observed cline at *F/S* can be accounted for by hitchhiking in response to the clinal ∇*1* site (Figure 11a).

On the basis of the very limited population structure

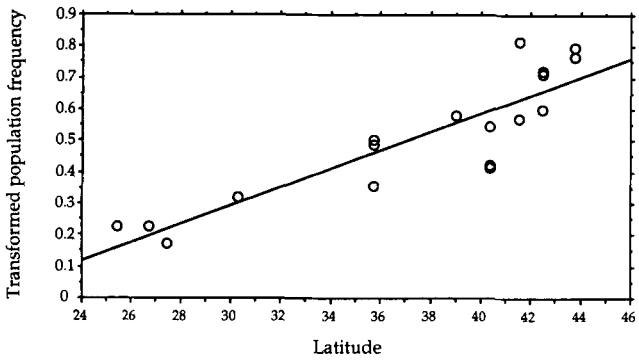

FIGURE 10.—Regression against population latitude of the square-root, arcsine transformed frequencies of $\nabla 1$ in each population; $r^2 = 0.786$.

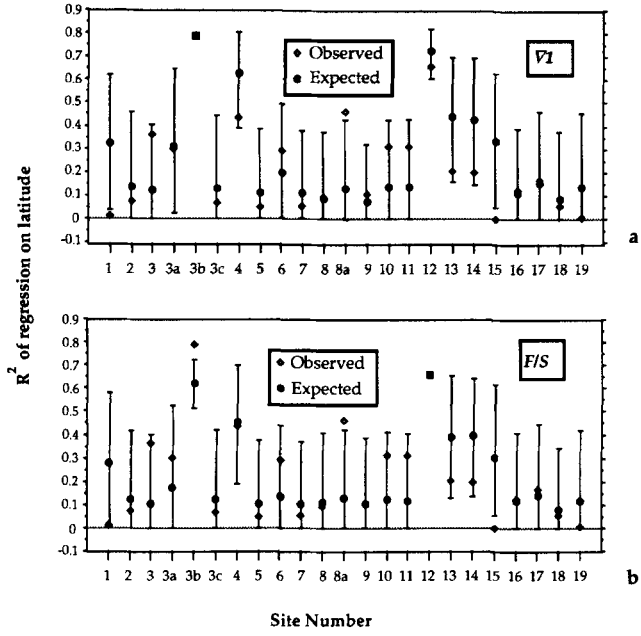

FIGURE 11.—Results of regression simulation (see text for details) for reduced $\nabla 1$ dataset. Each plot shows the observed data and the expected values with 95% confidence intervals when a different site is held constant and is assumed to be governing frequencies at all other sites: (a) $\nabla 1$ and (b) *F/S* site.

identified in the general survey, we have pooled all populations to study linkage disequilibrium. Measures of linkage disequilibrium, D' (LEWONTIN 1964) (data not shown), show the four previously unscored sites to fit into the general pattern identified above: high levels of linkage disequilibrium which do not decline over this study's maximum inter-site distances, about 3 kb. $\nabla 1$ is no exception: for its 24 pairwise comparisons in which the frequency of each variant is at least 5 in the pooled sample, 21 have an absolute value greater than 0.9. The three remaining comparisons are: *Hae*III 817, $D' = 0.09$; *Msp*I 1235, $D' = 0.32$; *Hae*III 1923, $D' = 0.77$.

We performed the same selective equivalence test as discussed above on the between north/south-block level of the hierarchical G_{ST} analysis of haplotypes (CHAKRABORTY 1980) (Table 5). We repeated our

TABLE 5

Hierarchical G_{ST} values for between north and south blocks

Dataset analyzed	G_{ST} value for between north/south blocks
All data	0.0082**†
$\nabla 1$-haplotypes only	0.0253
non-$\nabla 1$-haplotypes only	0.0035**
F-haplotypes only	0.0512**††
S-haplotypes only	0.0031*

The sampling scheme was as in Table 1. Statistical significance was gauged by two different simulation methods (see text for details). * $P < 0.05$ and ** $P < 0.01$ for the randomization simulation. † $P < 0.05$ and †† $P < 0.01$ for the sample-shuffling simulation.

earlier analysis by looking at the amount of within *F*- and within *S*-haplotype differentiation between north and south as well as for the whole dataset. As before, we tested the observed values for a departure from homogeneity against two different null hypotheses. The first, more stringent, one is a randomization simulation in which all individuals are pooled and each sample reconstituted at random from the pool; this effectively simulates panmixis. The second is less stringent and involves shuffling at random the relative positions of the localities. This preserves all population structure below the level of locality.

We see essentially the same results as before for this reduced dataset, except that there is evidence of some differentiation between North and South for the *S*-haplotypes dataset when tested against the stringent null hypothesis of panmixis. This may reflect the reduced sample size. We continue to see evidence of extensive heterogeneity within *F*-haplotypes against both the randomization and sample-shuffling tests. We have also tested both the $\nabla 1$-haplotype and the non-$\nabla 1$-haplotype datasets. We see no evidence of differentiation within $\nabla 1$-haplotypes; within non-$\nabla 1$-haplotypes, we see differentiation relative to the stringent randomization test but the there is no departure from homogeneity under the sample-shuffling scheme. This suggests that the extent of differentiation within non-$\nabla 1$-haplotypes is limited.

A haplotype network (not shown) shows the 22 $\nabla 1$-haplotypes clustering within the *F*-haplotype group; only 3 $\nabla 1$-haplotypes, represented by a total of four chromosomes, carry *Adh-S*.

DISCUSSION

There were two phases to this study: first, the general survey of RFLP variation in the *Adh* region, and, second, the specific analysis of *Château Douglas* and $\nabla 1$ frequencies. The second phase was prompted by the results of the first. Our discussion will follow this dichotomy.

General survey

Our initial aim was to distinguish between selective and non-selective explanations of the east coast *D.*

melanogaster Adh allozyme cline. Our data on the geographic distribution of polymorphisms in and around the *Adh* region can be broken into two classes: sites, in which we treat each polymorphism as an independent locus, and haplotypes, in which we treat an entire array of variable sites as a single multisite genotype. Our initial site analysis revealed a latitudinal cline at only the *F/S* site and haplotype analysis demonstrated a homogeneous distribution of *S*-haplotypes throughout the east coast but latitudinal heterogeneity among *F*-haplotypes. Does this information permit us to draw conclusions about the possible role of selection in establishing and/or maintaining the cline?

We argue below that the critical test of whether or not the cline is adaptive is supplied by the site-by-site analysis. The haplotype analysis provides support for the conclusions of the site-by-site analysis as well as addressing the issue of what site(s) is/are the target of selection. Before discussing each class of data, we will point out some of the problems associated with the analysis of nucleotide polymorphism in the presence of recombination.

Intermediate recombination and population genetic analysis: We find evidence of extensive, but not necessarily total, linkage disequilibrium in pairwise comparisons between sites throughout the *Adh* region. Thus separate polymorphic sites are correlated because they are in linkage, and haplotypes are not allelic because they may be related through recombination as well as through mutation. We lack a body of theory of neutral mutations in subdivided populations when there is intermediate recombination.

We have elected to apply conventional population genetic analyses inappropriately: for example, we calculate G_{ST} values for haplotypes as though each one were a different allele at a single locus (which they are not) and for individual polymorphic sites as if the presence and absence of the site constitute two alleles at a single locus (but each locus is not independent). Fortunately, both analyses yield, qualitatively, the same results. For our data, with essentially the same haplotypes appearing in all samples, the time scale for population mixing may be substantially shorter than that for mutation and recombination. Thus the rate of origination of new alleles may not be important in the analysis of population structure. In addition, the dataset is large enough to permit extensive statistical testing against simple null models, such as, in the case of population structure, panmixis, by means of randomization simulations. Such techniques allow us to avoid applying models in which no recombination or free recombination is assumed. We accordingly do not present estimates of migration rates and other parameters whose significance cannot be evaluated under available methods (*e.g.*, SLATKIN 1987).

One way of avoiding these problems would be to use a number of unlinked putatively neutral loci scattered throughout the genome. This, however, could be misleading: some of these sites may be in linkage disequilibrium with nearby selected sites such that their distributions reflect the action of selection on those selected sites rather than the influence of neutral drift/migration processes. Heterogeneity among the neutral loci cannot be ascribed to a single cause: either neutral processes or hidden linked selection may be responsible. Confining the study to a well defined region is preferable.

Site-by-site analysis–selection or history? We can resolve whether the cline is the by-product of historical processes or is adaptive by comparing the geographical distributions of variable sites. History should affect all sites uniformly whereas selection only affects the specific targets of selection (and linked sites).

What information do we need to demonstrate selection? In order to distinguish between the influences of selection and drift, we must identify a subset of our data (*i.e.*, a group of polymorphisms) that is evolving neutrally, serving as an indicator of the effects of population structure and history.

Can we identify such a subset of neutrally evolving markers? We have *a priori* reasons to expect many of our markers to be neutral: all the highly variable polymorphic sites in this study, except for the *F/S* site itself, are effectively silent in that they do not affect the amino acid sequence of the ADH protein. These sites occur in flanking sequence or introns, or are synonymous substitutions in coding sequence. Although these classes of mutation may be subject to selection (*e.g.*, for codon usage among synonymous codons (SHIELDS *et al.* 1988); see also ∇I below), silent sites generally evolve at high rates, suggesting that their evolution is approximately neutral. In addition, all the polymorphisms in this study are included in a test of neutral evolution at the *Adh* locus and flanking regions (KREITMAN and HUDSON 1991). Except for a small region encompassing the *F/S* site, there is generally a good correspondence between polymorphism and divergence as expected under neutrality.

We cannot demonstrate unequivocally that the studied polymorphisms are not subject to selection; indeed we argue below that one of them, ∇I, is subject to selection. However, two features of the data suggest that, other than ∇I and *F/S*, all the sites studied are not under selection. First Figure 6 shows a uniform pattern (*i.e.*, a lack of "clinality") among all putatively neutral sites, except site 5, whose cline can be explained in terms of hitchhiking with the *F/S* site. Such uniformity is predicted for neutral "indicator" loci: population structure and history should affect all loci equally whereas selection only affects the loci (and any linked to them) under selection. Second, the homogeneity of the distribution of *S*-haplotypes (Tables 4

and 5) implies that all *S*-haplotypes are selective equivalents. Note that $\nabla 1$ is scarcely polymorphic within the *Adh-S* class; variation among *S*-haplotypes may therefore be neutral.

Our null hypothesis, no selection, is that the same forces of migration and drift determine patterns of variation at all variable sites. Selection may cause a discrepancy between the distributions of site frequencies. For example, a site whose frequency is governed by selection may vary clinally while the frequencies of neutral markers are determined solely by drift/migration.

Selection on F/S: In order to determine whether different polymorphic sites show different patterns of geographic variation, we treat each site as a separate locus, initially ignoring the possible correlation among sites caused by linkage disequilibrium among them. Both the site by site hierarchical G_{ST} and the latitudinal regression analyses (Figures 5 and 6) show clinal variation at the *F/S* site, and, to a lesser extent, site 5. Frequencies at all other sites do not vary systematically with latitude. This pattern is surprising in view of the high levels of linkage disequilibrium throughout the region and, in particular, between the clinal *F/S* site and other markers. We expect to see "clinality" at other sites through hitchhiking. However, our simulations (Figure 8) reveal the observed lack of "clinality" at these sites to be nevertheless consistent with levels of linkage of disequilibrium with the *F/S* site, and that site 5's "clinality" can be explained in terms of hitchhiking in response to variation in *F/S* site frequencies. The *F/S* site cline cannot be explained in terms of hitchhiking in response to any of the scored sites, though *F/S* may yet be in linkage disequilibrium with an unsurveyed site (see below).

These results vindicate our approach. Although we studied a small region exhibiting high levels of linkage disequilibrium, hitchhiking does not obscure the differences in "clinality" among markers. What accounts for the "clinality" at the *F/S* site? Selection may be acting on a linked locus (see below), but, at this stage, we are concerned with the question of whether any explanation other than selection can account for the cline. A nonselective scenario requires founder populations to be similar in frequencies at every surveyed site except the *F/S* site. Studies of European (AGUADÉ 1988) and African (A. BERRY and M. KREITMAN, unpublished) populations, however, reveal considerable variation at a number of sites other than *F/S*. Frequencies would therefore have differed between founder populations at a range of sites, resulting, under the historical model, in parallel clines for each of these sites. We conclude that Figures 6 and 8 provide strong evidence in support of a selective basis for the *Adh* cline.

Haplotype-by-haplotype analysis–selection on what? Given that the cline is selectively maintained, what specifically is under selection? Of the sites studied in the initial survey, *F/S* is the most clinal (Figure 6). Furthermore, if we assume that the frequencies of all other sites are determined by their linkage relations with just *F/S*, we can explain all observed frequencies (Figure 8c) whereas no other site can account for the "clinality" at *F/S* (Figure 8, a and b). This suggests that *F/S* itself (or an unscored locus (or loci) to which it is linked) is under selection.

To test whether or not selection is acting solely on the *F/S* site, we carried out a "selective equivalence" test. A single site is assumed to be under selection, all other variation being neutral. Thus all *S*-haplotypes are selective equivalents as are all *F*-haplotypes. Under this hypothesis, we expect the distributions of two sets of haplotypes, *F* and *S*, to reflect the same forces of drift and migration. Using G_{ST} to gauge population subdivision and comparing between blocks (north *vs.* south), Table 4 shows the distributions of *F* and *S* haplotypes to differ. The G_{ST} for *S* alleles is very small; indeed there is no evidence for departure from panmixis. However, there is considerable "clinality" for *F*-haplotypes. Why do we see this discrepancy? If frequencies within each allelic class were governed solely by neutral evolutionary forces (*i.e.*, population history and structure), then we would not expect the G_{ST} values to be different. Selection must therefore be affecting the distribution of haplotypes within one or both allelic classes.

S-haplotypes: These exhibit remarkable uniformity between north and south. If this homogeneity were derived from a very recent expansion of the species, we cannot account for the lack of homogeneity we see at lower levels of the sampling hierarchy (Table 4). Moreover, under a recent expansion model, the observed uniformity of *S*-haplotypes is explicable only in terms of the expansion of a single population; such a scenario cannot account for the varying overall frequencies of all *F*-haplotypes and all *S*-haplotypes (*i.e.*, the cline at the *F/S* site). We must therefore reject a single-population expansion model as an explanation of the homogeneity within *S*-haplotypes. There are two alternative models: balancing selection for constant within-*S* frequencies of haplotypes or high levels of gene flow. Both of these models can account for the combined observation of a lack of north-south differentiation coupled with limited local population structure by assuming some local stochasticity causing differentiation in the face of global homogenizing processes.

In allozyme studies, it has been difficult to discriminate between these alternative explanations of homogeneity. This study's use of effectively silent markers, however, implies strongly that gene flow is responsible for the homogeneity. Whereas selection may

plausibly distinguish between the two or a few allozyme variants at a locus, it is inconceivable that selection could distinguish among, and set the frequencies of, the 78 *S*-haplotypes identified in this study. In fact, it is highly implausible that selection could act in this way to maintain the frequencies of those eight *S*-haplotypes that occur more than 50 times in the global sample. That each of these haplotypes is distinguished from other *S*-haplotypes solely by effectively silent differences further reduces the plausibility of selection seeing, and acting upon, those differences. We conclude, therefore, that the uniformity within *S*-haplotypes on a large geographic scale is caused by gene flow.

F-haplotypes: These show evidence of "clinality." If, as shown above, levels of gene flow are sufficient to homogenize frequencies of haplotypes within *S*-haplotypes, the "clinality" within *F*-haplotypes may be attributed to selection. Alternatively, perhaps some *F*-haplotypes have arisen so recently that there has not been enough gene flow to homogenize their distributions throughout the east coast. However, assuming that the same recombinational/mutational process generates new haplotypes within both *Adh-F* and *Adh-S* alleles, we do not expect *F*-haplotypes to show a more pronounced pattern of endemism than *S*-haplotypes. In fact, we expect to see *more* endemism among *S*-haplotypes because, by virtue of their greater abundance (n = 1061) relative to *Adh-F* alleles (n = 472), the overall rate of generation of new *S*-haplotypes will be higher. Thus, if rates of migration (*i.e.*, of population homogenization) are too low to prevent differentiation, *S*-haplotypes are more likely than *F*-haplotypes to exhibit regional endemism. Yet we see no heterogeneity among *S*-haplotypes. We cannot therefore account for the heterogeneity among *F*-haplotypes in terms of their recent origin coupled with limited migration. In fact, migration is undoubtedly the dominant factor determining population structure: indeed, it is possible that gene flow is extensive on the scale of the entire North American continent as KREITMAN and AGUADÉ (1986) found that a population from Putah Creek, California, is composed of the same major haplotypes as we see on the east coast. Thus an unselected haplotype with a distribution restricted to its area of origin must necessarily have arisen very recently and would accordingly be at a low frequency whose effect on our calculations would be trivial. To explain the within-*F*-haplotype heterogeneity, we are left with selection, which is discriminating among *F*-haplotypes with the result that they are not all selective equivalents. This result does not preclude selection on the *F/S* site, but it implies that selection is also acting on at least one other site which is polymorphic within *F*-haplotypes but apparently not so within *S*-haplotypes.

There are a number of conclusions we can draw from the first part of this study, the general RFLP survey. The site-by-site analysis shows clinal selection to be responsible for the *Adh* cline and the haplotype-by-haplotype analysis indicates that selection is not acting solely on the *F/S* site but is apparently also affecting a locus (or loci) in linkage disequilibrium with *Adh-F*. These findings are consistent with CHOUDHARY and LAURIE's (1991) observation that the *F/S* difference is responsible for the difference in catalytic efficiency between the allozymes but does not cause the two- to threefold difference in ADH protein abundance. Selection may therefore be acting on both the *F/S* site and an unidentified protein abundance-determining factor. The second part of this study was an attempt to identify that factor.

Château Douglas and ∇*1* surveys

Our limited survey for variation at the *Château Douglas* polymorphic site reveals this to be a low frequency variant unrelated to within-*F*-haplotype heterogeneity. However, ∇*1* shows clinal differentiation more marked even than that at the *F/S* site. In view of levels of gene flow among east coast *D. melanogaster* populations high enough to homogenize the distributions of unselected markers, we conclude that the ∇*1* site's cline is either the product of direct selection or hitchhiking in response to selection on linked sites.

Can variation at the ∇*1* site explain the heterogeneity within *F*-haplotypes? That ∇*1* is polymorphic within *F*-haplotypes but scarcely so within *S*-haplotypes (only four individuals in the entire sample have the ∇*1*/*Adh-S* combination) suggests this to be possible. Polymorphism for ∇*1* within *F*-haplotypes may be responsible for selective differences among *F*-haplotypes, and the lack of significant polymorphism for ∇*1* within *S*-haplotypes may therefore account for the selective equivalence of *S*-haplotypes. The simulation results shown in Figure 11a show that virtually all the variation in "clinality" observed among the surveyed sites, including the cline at the *F/S* site, can be explained by assuming that frequencies of each site are governed by their linkage disequilibrium with the ∇*1* site. Note that the converse does not hold: the "clinality" at ∇*1* cannot be explained by hitchhiking in response to the *F/S* site (Figure 11b). Now, if selection were acting on only ∇*1*, we expect to see heterogeneity within *F*-haplotypes because selection pressures vary between ∇*1*-bearing *F*-haplotypes and non-∇*1*-bearing *F*-haplotypes. Conversely, we predict homogeneity within *S*-haplotypes because virtually all of them (789 out of 793) are non-*1*-bearing.

To test whether selection may plausibly be acting on only ∇*1*, we performed a selective equivalence test as before. As a measure of clinal differentiation, we again used the G_{ST} value for the comparison between

the north-south blocks in the sampling scheme outlined in Table 1 (Table 5). The results are equivocal. We have shown that levels of gene flow are high enough to homogenize gene frequencies between north and south in the absence of differentiating selection; we therefore expect to see no evidence of north-south differentiation among selective equivalents. However, within non-$\nabla 1$-haplotypes, we see a departure from uniformity when tested against a stringent null hypothesis of effective panmixis. There is no such departure in a test against a less stringent sample-shuffling model. There is no significant heterogeneity within $\nabla 1$-haplotypes under either test. Consistent with the previous study, we see evidence of considerable heterogeneity within *F*-haplotypes. Overall, therefore, the supposition that selection is acting on the $\nabla 1$ site alone is better supported by this analysis than an alternative, that selection is acting on the *F/S* site alone.

To further investigate variation within haplotype classes, we calculated the regressions on population latitude of the relative frequencies (square root, arcsine transformed) of each haplotype within (1) the appropriate *Adh-F* or *Adh-S* class and (2) the appropriate $\nabla 1$ or non-$\nabla 1$ class. Thus, for a $\nabla 1$-*Adh-F* haplotype, we calculated the regressions on latitude of its frequencies in each population as a proportion of (1) all *Adh-F*'s in each population and (2) all $\nabla 1$'s in each population. We tested the statistical significance of the regressions by simulation, as described above. For the within *Adh-F/S* survey, two *F*-haplotypes yielded significant regressions (*i.e.*, $P < 0.01$. A value of 0.05 is an inappropriate significance level because a unique haplotype located in a population at the extreme of the geographic distribution yields a significant regression on this criterion). The within $\nabla 1$/non-$\nabla 1$ survey yielded just one significant regression, for a non-$\nabla 1$-haplotype. These results accord with the G_{ST} data for haplotypes (Table 5): geographical homogeneity of *S*-haplotypes and $\nabla 1$-haplotypes, heterogeneity among *F*-haplotypes (two significant regressions) and less pronounced heterogeneity among non-$\nabla 1$-haplotypes (one significant regression).

Can we explain the observed heterogeneity within *Adh-F* and non-$\nabla 1$ classes solely in terms of the distributions of these few aberrant haplotypes? It turns out that a single haplotype (corresponding to haplotype 11 in Figure 9) yields a significant regression for both within *Adh-F* and within non-$\nabla 1$ classes. Remarkably, the regression on latitude of its transformed frequencies yields a negative slope ($y = -0.015x + 0.686$; $r^2 = 0.602$; $P < 0.001$): haplotype 11 is a *F*-haplotype with a *S*-haplotype-like distribution. Is haplotype 11 peculiar in some way? It is the highest frequency *F*/non-$\nabla 1$-bearing haplotype ($n = 31$). Perhaps an interaction between these two sites endows haplotype 11 with its "quasi-*S*" properties. If this were true, however, we would expect all *F*/non-$\nabla 1$-bearing haplotypes to vary in this way. Now, overall frequencies of individual haplotypes, other than haplotype 11, in this class are too low to permit detection of a trend over our 18 populations. However, if we pool all these haplotypes, excluding haplotype 11, we generate a single composite haplotype of overall frequency 29. The regression on latitude of the transformed frequencies of this composite haplotype shows no significant latitudinal trend. In fact, the regression slopes in the opposite direction to that of haplotype 11 ($y = 0.003x + 0.019$; $r^2 = 0.048$; $P > 0.05$). Thus we cannot explain haplotype 11's aberrant behavior solely in terms of the combination of the non-$\nabla 1$ and *Adh-F* sites.

Linkage disequilibrium between $\nabla 1$ and *Adh-F*—evidence of epistatic selection? As expected from the patterns of linkage disequilibrium, the haplotype network for the $\nabla 1$ dataset (not shown) shows the $\nabla 1$-haplotypes clustering within the *F*-haplotypes. Is this linkage disequilibrium between the $\nabla 1$ and *Adh-F* sites explicable solely in terms of insufficient recombination to erode the disequilibrium, or need we invoke another factor, such as epistasis, to account for the linkage disequilibrium? A recent mutation is typically in disequilibrium with other markers on the chromosome on which it arose because there has not been enough time for disequilibrium-eroding recombination to have occurred. Also, unless the mutation is selectively favored so that its frequency increases deterministically under selection, its frequency will remain low for an initial period until, by chance, it drifts to a higher frequency. Thus we expect a correspondence between levels of linkage disequilibrium and the frequency of a mutation. In addition, high levels of linkage disequilibrium imply that the mutation-bearing chromosome has not undergone much recombination, so we expect to see a corresponding low diversity of mutation-bearing chromosomes. $\nabla 1$ does not fulfil these expectations: it is in strong linkage disequilibrium with most of the sites surveyed and yet over 26% of chromosomes studied carry it. Also, it is has undergone considerable diversification: 25 haplotypes (27% of the total) carry $\nabla 1$.

Particularly striking is the lack of *Adh-S*/$\nabla 1$ haplotypes given the extensive diversification of $\nabla 1$-haplotypes that has occurred within *F*-haplotypes. Has recombination generated this diversity within *F*-haplotypes? If the main cause of diversification is mutation, then the linkage disequilibrium between the *F/S* and $\nabla 1$ sites is expected, but that disequilibrium would be broken down by recombination. We note that 12 of the 15 sites that differentiate between $\nabla 1$-haplotypes differing by only a single polymorphism also differentiate between similarly related non-$\nabla 1$-haplotypes. If mutation is responsible for the diversification of

haplotypes, then we have to invoke many instances of parallel mutation to explain this distribution. Recombination, on the other hand, promotes the sharing of mutations between haplotypes and is presumably, therefore, the primary cause of diversification of haplotypes. How can we explain why extensive recombinational diversification has taken place for $\nabla 1$ on *F*-haplotypes but not on *S*-haplotypes? Unfortunately, we do not know enough about the dynamics of recombination over evolutionary time to be able to dismiss simple chance as an explanation. In addition, it is always possible to invoke historical scenarios to explain the observed pattern: for example, perhaps $\nabla 1$ originated, and subsequently diversified via recombination, in a population consisting entirely of *F*-haplotypes. Such a population is unknown. It would then have had no chance to recombine onto *S*-haplotypes. Nevertheless, the pattern is also consistent with a model in which the association between $\nabla 1$ and *Adh-F* is selectively maintained, possibly by selection against $\nabla 1$/*Adh-S* haplotypes. Such selection would necessarily be epistatic.

How could selection, whether epistatic or not, act on $\nabla 1$? ADH is transcribed from different promoters at different life history stages. $\nabla 1$ is located in an intron in the adult transcript of ADH but is upstream of the site of initiation of larval transcription (BENYAJATI *et al.* 1983). The region in which $\nabla 1$ is located has been implicated by a number of studies in the regulation of transcription from the larval (or proximal) promoter (HEBERLEIN, ENGLAND and TJIAN 1985; POSAKONY, FISCHER and MANIATIS 1985; KAPOUN *et al.* 1990). It is possible therefore that the presence of $\nabla 1$ affects regulation of larval transcription. However, LAURIE and STAM (1988) report a lack of correlation between ADH-protein levels and mRNA levels, which suggests that differences between *Adh* genotypes for ADH-protein activity are determined by differences either in translation rates or in rates of degradation of the protein. It is difficult to see how $\nabla 1$ could affect either of these processes as it is never part of the larval transcript and is spliced out of the adult mRNA. But further developments may provide $\nabla 1$ with a selectable role. The difficulty in attributing selectable function to $\nabla 1$ can be avoided if we invoke selection on another site in linkage disequilibrium with the $\nabla 1$ and *Adh-F* sites. This is possible, but LAURIE-AHLBERG and STAM's (1987) transformation analysis (i) implies that that site must lie within the region surveyed in this study and (ii) excludes many of the candidate polymorphisms. At this stage, however, we can rule out neither direct selection on $\nabla 1$ nor selection on linked loci.

Conclusion: The success of our approach in detecting natural selection requires the population subdivision of the neutral markers to be distinguishable from the distribution of variation at site(s) under selection. For *D. melanogaster*, with its high rates of migration, the lack of clinality of virtually all silent sites provides a favorable backdrop against which to test for selection. Another situation in which selection may be detected arises from a complementary scenario: if populations are clearly subdivided then polymorphisms whose frequencies are constant across the subdivided populations may also be identified as being under selection. KARL and AVISE (1992) deduce selection on allozymes in the American oyster on this basis: populations from the Atlantic and Gulf of Mexico coasts are distinct for both nuclear and mitochondrial DNA markers, but exhibit uniform allozyme frequencies.

The lack of genetic differentiation of silent polymorphisms among east coast *D. melanogaster* populations is evidence of a selective basis of the cline. Although, for reasons discussed above, we do not estimate the migration rate and cannot therefore directly calibrate the erosion of heterogeneity between populations, the simple observation of effective panmixis among *S*-haplotypes between north and south blocks demonstrates that rates of gene flow are high enough to homogenize gene frequencies over large geographic areas. Evidence of such extensive gene flow is not surprising in a vagile insect: RAYMOND *et al.* (1991) found evidence of rapid worldwide spread of insecticide resistance alleles in the mosquito, *Culex pipiens*. KEITH *et al.* (1985), RILEY, HALLAS and LEWONTIN (1989) and SCHAEFFER and MILLER (1992) have used a variety of techniques to demonstrate high levels of gene flow among *Drosophila pseudoobscura* populations. SINGH and RHOMBERG (1987) and HALE and SINGH (1991), using nuclear and mitochondrial genetic markers respectively, and COYNE and MILSTEAD (1987), using a direct mark-release-recapture approach, all found evidence of extensive gene flow between *D. melanogaster* populations. It is inconceivable that a historical cline could be maintained in the face of such gene flow, even allowing for a very recent introduction of *D. melanogaster* to North America. This implies that other *D. melanogaster* east coast clines [*e.g.*, for wing length (COYNE and BEECHAM (1987)] also have a selective basis.

Our conclusions regarding the specific targets of selection are less clear-cut. We find that $\nabla 1$ explains some, but not all, of the heterogeneity within *F*-haplotypes. That neither the *F/S* nor $\nabla 1$ sites can fully account for the distribution of haplotypes (especially haplotype 11) suggests that selection is affecting other linked sites or is acting epistatically on the combination of the *F/S* and $\nabla 1$ sites. The high levels of linkage disequilibrium between $\nabla 1$ and other sites, in particular *F/S*, may indicate epistatic selection. It is also possible that selection is not acting directly on the

F/S site and that the observed cline is the product of hitchhiking in response to the cline at $\nabla 1$. However, KREITMAN and HUDSON's (1991) finding of a peak of polymorphism around the *F/S* site suggests that balancing selection is indeed acting on that site. The lack of such an excess polymorphism around the $\nabla 1$ site may indicate that, contrary to what our findings suggest, $\nabla 1$ is not under selection, or, alternatively, that $\nabla 1$ arose too recently to permit a build-up of polymorphism. We note finally that the heterogeneity among *F*-haplotypes, whether it is caused by $\nabla 1$ or by other factors, may account for the consistently equivocal results of population cage experiments (CHAMBERS 1988). If, for example, the flies studied were collected from a southern population, a large proportion of the *Adh-F* alleles under study could well be the "quasi-S" haplotype 11. Studies concentrating on *F/S* differences may well be confounded by these within-allele differences.

We thank J. H. MCDONALD, M. TAYLOR, D. STRATTON, D. MCCUNE, M. FORNEY, M. CHARLES, C. NEWLIN, C. LAURIE, the National Science Foundation and the National Institutes of Health.

LITERATURE CITED

AGUADÉ, M., 1988 Restriction map variation at the *Adh* locus of *Drosophila melanogaster* in inverted and non-inverted chromosomes. Genetics **119:** 135–140.

AQUADRO, C. F., S. F. DEESE, M. M. BLAND, C. H. LANGLEY and C. C. LAURIE-AHLBERG, 1986 Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. Genetics **114:** 1165–1190.

ÅRNASON, E., 1991 Perturbation-reperturbation test of selection vs. hitchhiking of the two major alleles of *Esterase-5* in *Drosophila pseudoobscura*. Genetics **129:** 145–168.

BENYAJATI, C., N. SPOEREL, H. HAYMERLE and M. ASHBURNER, 1983 The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5′ end in different developmental stages. Cell **33:** 125–133.

BERRY, A., J. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

CAVALLI-SFORZA, L. L., 1966 Population structure and human evolution. Proc. R. Soc. B **164:** 362–379.

CHAKRABORTY, R., 1980 Gene-diversity analysis in nested subdivided populations. Genetics **96:** 721–723.

CHAMBERS, G. K., 1988 The *Drosophila* alcohol dehydrogenase gene-enzyme system. Adv. Genet. **25:** 39–107.

CHOUDHARY, M., and C. C. LAURIE, 1991 Use of *in vitro* mutagenesis to analyze the molecular basis of the difference in *Adh* expression associated with the allozyme polymorphism in *Drosophila melanogaster*. Genetics **129:** 481–488.

CHRISTIANSEN, F. B., and O. FRYDENBERG, 1974 Geographical patterns of four polymorphisms in *Zoarces viviparus* as evidence of selection. Genetics **77:** 765–770.

COLLET, C., 1988 Recent origin for a thermostable alcohol dehydrogenase allele of *Drosophila melanogaster*. J. Mol. Evol. **27:** 142–146.

COYNE, J. A., S. AULARD and A. BERRY, 1991 Lack of underdominance in a naturally-occurring pericentric inversion in *Drosophila melanogaster* and its implications for chromosome evolution. Genetics **129:** 791–802.

COYNE, J. A., and E. BEECHAM, 1987 Heritability of two morphological characters within and among natural populations of *Drosophila melanogaster*. Genetics **117:** 727–737.

COYNE, J. A., and B. MILSTEAD, 1987 Long-distance migration of *Drosophila*. 3. Dispersal of *D. melanogaster* alleles from a Maryland orchard. Am. Nat. **130:** 70–82.

DAVID, J. R., and P. CAPY, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106–111.

EISSES, K. TH., A. J. ANDRIESSE, A. DOUWE DE BOER, G. E. W. THÖRIG and P. J. WEISBEEK, 1990 Analysis of the gene encoding the multifunctional alcohol dehydrogenase allozyme ADH-71K of *Drosophila melanogaster*. Mol. Biol. Evol. **7:** 459–469.

EWENS, W. J., 1979 *Mathematical Population Genetics*. Springer-Verlag, Berlin.

HALE, L. R., and R. S. SINGH, 1991 A comprehensive study of genic variations in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history *vs.* selection in the genetic structure of geographic populations. Genetics **129:** 103–117.

HEBERLEIN, U., B. ENGLAND and R. TJIAN, 1985 Characterization of *Drosophila* transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. Cell **41:** 965–977.

HEINSTRA. P. W. H., W. SCHARLOO and G. E. W. THÖRIG, 1987 The physiological significance of the alcohol dehydrogenase polymorphism in larvae of Drosophila. Genetics **117:** 75–84.

HIGUCHI, R. G., and H. OCHMAN, 1989 Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17:** 5865.

HUDSON, R. R., 1983 Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. **23:** 183–201.

JOHNSON, C. W., 1913 The distribution of some species of *Drosophila*. Psyche **20:** 202–205.

KAPOUN, A. M., B. W. GEER, P. W. H. HEINSTRA, V. CORBIN and S. W. MCKECHNIE, 1990 Molecular control of the induction of alcohol dehydrogenase by ethanol in *Drosophila melanogaster* larvae. Genetics **124:** 881–888.

KARL, S. A., and J. C. AVISE, 1992 Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science **256:** 100–102.

KEITH, T. P., 1983 Frequency distributions of *Esterase-5* alleles in two populations of *Drosophila pseudoobscura*. Genetics **105:** 135–155.

KEITH, T. P., L. D. BROOKS, R. C. LEWONTIN, J. C. MARTINEZ-CRUZADO and D. L. RIGBY, 1985 Nearly identical distributions of xanthine dehydrogenase in two populations of *Drosophila pseudoobscura*. Mol. Biol. Evol. **2:** 206–216.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus in *Drosophila melanogaster*. Nature **304:** 412–417.

KREITMAN, M., 1991 Protocols for 4-cutter blots and PCR sequencing, pp. 357–367 in *Molecular Techniques in Taxonomy* (NATO ASI Series H: Cell Biology, Vol. 57), edited by G. M. HEWITT, A. W. B. JOHNSTON and J. P. W. YOUNG. Springer-Verlag, Berlin.

KREITMAN, M., and M. AGUADÉ, 1986 Genetic uniformity in two populations of *Drosophila melanogaster* as revealed by filter hybridization of 4-nucleotide-recognizing restriction enzyme digests. Proc. Natl. Acad. Sci. USA **83:** 3562–3566.

KREITMAN, M. and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

LAURIE, C. C., J. T. BRIDGHAM and M. CHOUDHARY,

1991 Associations between DNA sequence variation and variation in expression of the *Adh* gene in natural populations of *Drosophila melanogaster*. Genetics **129:** 489–499.

LAURIE, C. C., and L. F. STAM, 1988 Quantitative analysis of RNA produced by Slow and Fast alleles of *Adh* in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **85:** 5161–5165.

LAURIE-AHLBERG, C. C., 1985 Genetic variation affecting the expression of enzyme-coding genes in Drosophila: an evolutionary perspective. Isozymes Curr. Top. Biol. Med. Res. **12:** 33–88.

LAURIE-AHLBERG, C. C., and L. F. STAM, 1987 Use of *P*-element mediated transformation to identify the molecular basis of naturally-occurring variants affecting *Adh* expression in *Drosophila melanogaster*. Genetics **115:** 129–140.

LEWIS, N., and J. B. GIBSON, 1978 Variation in amount of enzyme protein in natural populations. Biochem. Genet. **16:** 159–170

LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49:** 49–67.

LEWONTIN, R. C., and J. KRAKAUER, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics **74:** 175–195.

MANLY, B. F. J., 1985 *The Statistics of Natural Selection*. Chapman & Hall, New York.

NEI, M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70:** 3321–3323.

NEI, M., and T. MARUYAMA, 1975 Lewontin-Krakauer test for neutral genes. Genetics **80:** 395.

OAKESHOTT, J. G., J. B. GIBSON, P. R. ANDERSON, W. R. KNIBB, D. G. ANDERSON and G. K. CHAMBERS, 1982 Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. Evolution **39:** 86–96.

PIPKIN, S. B., E. FRANKLIN-SPRINGER, S. LAW and S. LUBEGA, 1976 New studies of the alcohol dehydrogenase cline in *Drosophila melanogaster* from Mexico. J. Hered. **67:** 258–266.

POSAKONY, J. W., J. A. FISCHER and T. MANIATIS, 1985 Identification of DNA sequences required for the regulation of *Drosophila* alcohol dehydrogenase gene expression. Cold Spring Harbor Symp. Quant. Biol. **50:** 515–520.

RAYMOND, M., A. CALLAGHAN, P. FORT and N. PASTEUR, 1991 Worldwide migration of amplified insecticide resistance genes in mosquitoes. Nature **350:** 151–153.

RILEY, M. A., M. E. HALLAS and R. C. LEWONTIN, 1989 Distinguishing forces controlling genetic variation at the *Xdh* locus in *Drosophila pseudoobscura*. Genetics **123:** 359–369.

ROBERTSON, A., 1975 Remarks on the Lewontin-Krakauer test. Genetics **80:**396.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. G. HIGUCHI, G. T. HORN, K. B. MULLIS and H. A. ERLICH, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SAMPSELL, B., 1977 Isolation and genetic characterization of alcohol dehydrogenase thermostability variants occurring in natural populations of *Drosophila melanogaster*. Biochem. Genet. **15:** 971–988.

SAMPSELL, B., and E. STEWARD, 1983 Alcohol dehydrogenase thermostability variants in *Drosophila melanogaster*: comparison of activity ratios and enzyme levels. Biochem. Genet. **21:** 1071–1088.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHAEFFER, S. W., and E. L. MILLER, 1992 Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. Genetics **132:** 471–480.

SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS and F. WRIGHT, 1988 "Silent" sites in *Drosophila* genes are not neutral: evidence for selection among synonymous codons. Mol. Biol. Evol. **5:** 704–716.

SIMMONS, G. M., M. E. KREITMAN, W. F. QUATTLEBAUM and N. MIYASHITA, 1989 Molecular analysis of the alleles of alcohol dehydrogenase along a cline in *Drosophila melanogaster*. I. Maine, North Carolina, and Florida. Evolution **43:** 393–409.

SINGH, R. S, and L. R. RHOMBERG, 1987 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. I. Estimates of gene flow from rare alleles. Genetics **115:** 313–322.

SLATKIN, M., 1987 The average number of sites separating DNA sequences drawn from a subdivided population. Theor. Popul. Biol. **32:** 42–49.

SMITH, M. R., G. K. CHAMBERS, L. D. BROOKS, F. M. COHAN and S. C. COHAN, 1984 How many *Adh* clines on the west coast of North America? Drosophila Inform. Serv. **60:** 188–190.

THÖRIG, G. E. W., A. A. SCHOONE and W. SCHARLOO, 1975 Variation between electrophoretically identical alleles at the alcohol dehydrogenase locus in *Drosophila melanogaster*. Biochem. Genet. **13:** 721–731.

VAN DELDEN, W., 1982 The alcohol dehydrogenase polymorphism in *Drosophila melanogaster*: selection at an enzyme locus. Evol. Biol. **15:** 187–222.

VIGUE, C. L., and F. M. JOHNSON, 1973 Isozyme variability in species of the genus Drosophila. VI. Frequency-property-environment relationships of allelic alcohol dehydrogenases in *D. melanogaster*. Biochem. Genet. **9:** 213–227.

VOELKER, R. A., C. C. COCKERHAM, F. M. JOHNSON, H. E. SCHAFFER, T. MUKAI and L. E. METTLER, 1978 Inversions fail to account for allozyme clines. Genetics **88:** 515–527.

WATTERSON, G. A., 1978 The homozygosity test of neutrality. Genetics **88:** 405–417.

WILKS, A. V., J. B. GIBSON, J. G. OAKESHOTT and G. K. CHAMBERS, 1980 An electrophoretically cryptic alcohol dehydrogenase variant in *Drosophila melanogaster*. II. Post-electrophoresis heat-treatment screening of natural populations. Aust. J. Biol. Sci. **33:** 575–585.

WRIGHT, S., 1951 The genetical structure of populations. Ann. Eugenet. **15:** 323–354.

Communicating editor: A. G. CLARK

APPENDIX

The haplotypes and their frequencies in each sample are given in Tables 6 and 7, respectively.

TABLE 6

Haplotypes

Haplotype	Site no. 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
1	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
2	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
3	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
4	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
5	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
6	1	0	1	1	0	0	0	1	0	0	0	1	0	1	1	1	0	1	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0
7	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
8	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0
9	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
10	1	0	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	0	0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
11	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
12	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
13	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
14	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
15	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	1	1	0	0	0	0
16	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0
17	1	0	1	1	0	0	0	1	1	0	0	0	0	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
18	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0
19	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0
20	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	0	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
21	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0
22	1	0	1	1	0	0	0	1	1	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
23	1	0	0	1	0	0	0	1	1	0	0	0	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
24	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
25	1	0	1	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
26	1	0	1	1	0	0	0	1	1	0	0	1	0	1	0	1	0	1	1	0	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
27	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
28	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
29	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
30	1	0	0	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
31	1	0	1	1	0	0	0	1	1	0	0	1	0	1	0	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
32	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
33	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0
34	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	0	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
35	1	0	1	1	1	0	0	1	1	0	0	1	0	0	1	1	0	0	0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
36	1	0	1	1	0	0	0	1	1	0	0	0	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
37	1	0	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	0	0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
38	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	0	0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
39	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	1	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0
40	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
41	1	0	1	1	0	0	0	1	0	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
42	1	0	0	1	0	0	0	1	1	0	0	0	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0
43	1	0	1	1	0	0	0	1	0	0	0	1	0	1	1	1	0	1	1	1	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0
44	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
45	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
46	1	0	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
47	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0
48	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0
49	1	0	1	1	0	0	0	0	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
50	1	0	1	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
51	1	0	1	1	1	0	0	1	1	0	0	1	0	0	1	1	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
52	1	0	1	1	1	0	0	1	1	0	0	1	0	0	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
53	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
54	1	0	1	1	0	0	0	1	0	0	0	1	0	1	1	1	0	1	0	1	0	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0
55	1	0	1	1	0	0	1	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0

Haplotype	Site no.																																										
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
56	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
57	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
58	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
59	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
60	1	0	1	1	0	0	0	1	1	0	0	1	1	0	1	1	0	1	1	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
61	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	0	1	0	1	1	0	0	1	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
62	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	0	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
63	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0
64	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	1	1	0	1	1	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
65	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	0	0	1	1	0	0	1	1	0	1	0	0	0	0	1	1	1	0	0	0	0
66	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	0	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
67	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0
68	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	1	0	0
69	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	1	1	0	1	0	0	0	0	1	1	1	1	0	0	0
70	1	0	0	1	0	0	0	1	0	0	0	1	0	1	1	1	0	1	0	1	0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0
71	1	0	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	0	1	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
72	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
73	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
74	1	0	1	1	0	0	0	1	0	0	0	1	0	1	1	1	0	1	0	1	0	0	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
75	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	0	1	1	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
76	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0
77	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0
78	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0
79	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0
80	1	0	1	1	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0
81	1	0	1	1	0	0	0	1	0	0	0	1	0	0	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
82	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
83	1	0	1	1	0	1	0	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
84	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0
85	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
86	1	0	1	1	0	0	0	1	1	1	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
87	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	0	1	1	0	0	1	0	1	0	0	1	1	0	1	0	0	0	0	1	1	1	0	0	0	0
88	1	0	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	0	0	1	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
89	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	1	1	0	1	0	0	0	0	1	1	1	0	0	0	0
90	1	0	1	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
91	1	0	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
92	1	0	1	1	0	0	0	1	1	0	0	1	0	1	0	1	0	1	1	0	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
93	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	1	1	1	0	0	0	0
94	1	0	1	1	0	0	0	1	1	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
95	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
96	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
97	1	0	0	1	0	0	0	1	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
98	1	0	1	0	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0
99	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	1	0	0	0	1	1	1	0	0	0	0
100	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0
101	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0
102	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
103	1	1	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
104	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	1	0	1	1	1	0	0	0	0
105	1	0	1	1	0	0	0	1	0	0	0	1	0	1	1	1	0	1	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0
106	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	1	0
107	1	0	1	1	0	0	0	1	0	0	1	1	0	1	1	1	0	1	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	1
108	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
109	1	0	1	1	0	0	0	1	1	0	0	0	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
110	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
111	1	0	0	1	0	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0
112	1	0	1	1	0	0	0	1	0	0	0	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0
113	1	0	0	1	1	0	0	1	1	0	0	0	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0

1 represents the presence of a site and a *Fast* allele at site 23; 0 its absence and a *Slow* allele at site 23.

TABLE 7

Frequencies of haplotypes in each sample

	Population																									
Haplotype	cf84	cf86	orw88	ew88	ng86	ng88	iw88	vf88	wes86	wes88	ly88	ss86	orch87	ter86	ter87a	ter87b	md88	ral83	ral84	ral87	jax87	fp87	mec87	fc84	fc87	Total
1	19	11	18	14	7	12	19	21	5	11	9	10	2	8	7	8	7	7	6	1	3	1	1	3	1	211
2	6	5	6	6	3	23	13	10	3	3	1	1	0	8	11	11	7	2	12	10	13	10	5	9	4	182
3	14	8	2	6	9	3	5	3	2	4	10	10	7	2	5	3	10	5	11	4	5	10	14	11	12	175
4	3	3	5	4	2	3	3	4	7	7	4	5	6	10	6	8	3	9	8	8	9	8	10	10	10	155
5	11	2	7	6	9	3	5	4	2	12	4	12	4	5	5	4	9	2	11	2	3	0	2	0	3	127
6	27	9	2	1	2	2	0	3	7	3	2	0	2	1	1	0	1	0	4	2	6	3	5	8	9	100
7	0	2	0	0	1	3	2	1	3	4	0	2	1	2	10	3	0	7	4	1	4	5	3	3	3	64
8	3	4	0	1	4	0	1	5	4	1	3	0	3	1	4	5	2	6	2	3	3	3	1	0	2	61
9	1	1	3	3	6	1	2	1	2	2	1	3	2	2	1	3	4	3	3	2	2	3	2	1	3	57
10	0	0	2	1	0	0	1	3	4	1	1	0	2	0	5	3	1	2	4	3	2	3	4	2	1	45
11	0	1	0	0	0	2	1	0	0	0	1	1	2	1	0	0	0	3	7	2	3	8	2	5	5	44
12	0	1	1	0	1	0	0	0	1	0	6	2	0	3	3	0	1	1	2	3	1	0	5	1	3	35
13	1	3	4	4	0	0	0	0	1	0	0	4	1	3	1	1	1	0	1	1	0	0	1	1	1	29
14	0	0	3	1	2	1	0	0	1	3	2	0	0	2	0	0	0	1	1	1	0	0	0	0	0	18
15	0	0	1	2	0	2	0	1	1	1	2	1	0	0	0	0	2	0	0	0	0	0	0	0	0	13
16	0	0	0	0	0	0	0	0	1	0	0	1	3	0	5	3	0	0	0	0	0	0	0	0	0	13
17	1	0	2	2	0	2	1	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
18	0	0	0	0	0	1	2	2	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	10
19	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	2	1	0	1	1	0	0	9
20	0	0	0	1	0	1	0	0	0	0	0	1	0	1	1	0	4	0	0	0	0	0	0	0	0	9
21	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	2	1	0	1	0	7
22	0	0	0	5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	7
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	6
24	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	1	5
25	0	0	0	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	5
26	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	5
27	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	4
28	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	4
29	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	4
30	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	4
31	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	3
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	3
34	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	3
35	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3
36	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	3
37	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	3
38	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	3
39	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3
40	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2
43	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	2
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	2
46	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
47	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
48	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
49	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
51	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
52	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
53	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
54	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
55	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
56	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
57	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
58	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

	Population																									
Haplotype	cf84	cf86	orw88	ew88	ng86	ng88	iw88	vf88	wes86	wes88	ly88	ss86	orch87	ter86	ter87a	ter87b	md88	ral83	ral84	ral87	jax87	fp87	mec87	fc84	fc87	Total
59	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
60	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
61	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
62	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
63	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
64	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
65	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
66	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
67	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
69	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
70	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
71	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
72	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
73	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
74	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
75	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
76	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
77	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
78	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
79	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
80	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
81	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
83	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
91	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
96	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
97	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
101	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
102	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
103	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
104	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
105	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
106	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
108	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
109	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
111	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
113	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1