Extensive Linkage Disequilibrium in the *achaete-scute* Complex of *Drosophila melanogaster*

James N. Macpherson *,[1], B. S. Weir † and Andrew J. Leigh Brown*

**Department of Genetics, University of Edinburgh, Edinburgh EH9 3JN, Scotland, and † Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203*

Manuscript received December 7, 1989
Accepted for publication May 21, 1990

ABSTRACT

We have analyzed the level of gametic association between restriction map variants in a sample of 44 *X* chromosomes from a natural population of *Drosophila melanogaster*. Of 21 pairwise tests involving 7 restriction map polymorphisms in the *yellow-achaete-scute* complex, 17 were found to be significant, including some between restriction sites over 80 kb apart. Three-way linkage disequilibria and their variances were also estimated for all 35 three-way comparisons between these loci. Twelve such tests were found to be significant, again spanning distances of up to 80 kb on the restriction map. Only 9 of a possible 128 haplotypes were represented in the sample and 8 of these could be linked together by changes at a single site. The strength of these associations at *y-ac-sc* is unusual by comparison with studies on other regions of the genome of *D. melanogaster*, and is consistent with the very low level of recombination which has been reported for the complex. However, our estimate of nucleotide diversity in the region is not significantly different from those made for some other loci in this species.

IN natural, outbred populations, it has not proved easy to detect nonrandom associations between alleles at different loci ("linkage disequilibrium"), except in a few well known cases [see reviews by LANGLEY (1977) and HEDRICK, JAIN and HOLDEN 1978). The advent of techniques which could detect allelic substitutions in the DNA sequence of any gene permitted more detailed investigation of this question than was previously possible. Initially, it appeared that disequilibria would be generally observed within the distances over which such molecular analyses extended. At the alcohol dehydrogenase locus (*Adh*) and the 87A heat shock locus of *Drosophila melanogaster*, significant gametic associations were found between restriction map variants separated by as much as 10 kb (LANGLEY, MONTGOMERY and QUATTLEBAUM 1982; LEIGH BROWN 1983; AQUADRO *et al.* 1986). Early work on restriction site variants in the human β-globin gene cluster also detected significant linkage disequilibria over even greater molecular distances (ANTONARAKIS *et al.* 1982).

Studies on restriction map variation at several other genes have now been published, including the *Amylase*, *Notch*, *rosy*, and *white* genes of *D. melanogaster*, the *Adh* gene of *D. pseudoobscura*, the *rosy* gene of *D. simulans* and the *vermilion* and *forked* loci of *D. ananassae* (LANGLEY *et al.* 1988; SCHAEFFER, AQUADRO and LANGLEY 1988; AQUADRO, LADO and NOON 1988; MIYASHITA and LANGLEY 1988; SCHAEFFER, AQUADRO and ANDERSON 1987; STEPHAN and LANGLEY 1989). Significant disequilibria have not been found to be so common in these studies, especially between sites separated by more than about 2 kb on the molecular map. From the observed frequency of recombination between mutations at the *white* locus of *D.melanogaster*, this distance would appear to correspond to a genetic distance of about 4×10^{-3} cM (B. H. JUDD, cited in AGUADE, MIYASHITA and LANGLEY 1989).

In the course of investigating the frequency of transposable element insertions at the *yellow-achaete-scute* gene complex in *D. melanogaster* we previously observed some significant gametic associations between restriction sites up to 30 kb apart and between restriction sites and DNA insertions over 60 kb apart (BEECH and LEIGH BROWN 1989). Two other groups have since confirmed the presence of significant associations between restriction site variants in this gene complex (AGUADE, MIYASHITA and LANGLEY 1989; EANES, LABATE and AJIOKA 1989). All three groups also found that large DNA insertions were rare in this region, per kilobase of DNA screened. However, the estimates of nucleotide diversity differed; that obtained by AGUADE, MIYASHITA and LANGLEY being markedly lower than the other two. We have now screened more polymorphic markers in this region in a larger number of lines and in this paper we present the results of a systematic survey of pairwise and three-way disequilibrium over 86 kb of DNA in this gene cluster, together with a new estimate of nucleotide diversity in a North Carolina population of *D. melanogaster*. A high proportion of pairwise comparisons

[1] Present address: Wessex Regional Cytogenetics Unit, Salisbury General Infirmary, Salisbury SP2 7SX, England.

are significant and we have detected several significant three-way disequilibria. Our results indicate that the entire complex, comprising a dozen or more transcripts of diverse functions (VILLARES and CABRERA 1987) must be evolving as a single unit in this species.

MATERIALS AND METHODS

Extraction of *X* chromosomes: Wild-inseminated female *Drosophila melanogaster* were collected from a natural population at a fruit market in Raleigh, North Carolina, in 1984 by A. E. SHRIMPTON and allowed to oviposit in separate vials in the laboratory. Single male progeny from these females were mated to attached-*X* females to found a series of *X* chromosome lines as described earlier (LEIGH BROWN and MOSS 1987). Lines 1–27 are the same as the NC lines used in an earlier study (BEECH and LEIGH BROWN 1989) and lines 28–44 were taken from the same collection.

Preparation of fly genomic DNA: DNA was prepared in similar fashion to BEECH and LEIGH BROWN (1989): 500 male flies from each line, stored at −70° since collection, were homogenized in 4 cm^3 of 10 mM Tris pH 7.5 containing 100 mM NaCl, 10 mM EDTA, 0.15 mM spermidine and 0.15 M spermine using a manual glass homogenizer (Kontes Glass Co.). An equal volume of 200 mM Tris pH 9 containing 30 mM EDTA, 2% sodium dodecyl sulfate and 0.2 mg cm^{-3} Pronase-E (Sigma) was added and mixed gently with the homogenate before incubating at 37° for 1 hr. Two equal-volume extractions with phenol, one with phenol-chloroform and one with chloroform were carried out and DNA was precipitated by adding 1/10 volume 7 M ammonium ethanoate and 2 volumes ethanol. The centrifuged pellet was dried thoroughly under vacuum and resuspended in 0.5 cm^3 10 mM Tris-Cl, 0.1 mM EDTA (pH 8.0) (TE).

Resolution of DNA fragments: DNA was digested to completion with one of a range of restriction enzymes recognizing either 4- or 6-bp sequences. Samples digested with six-cutter enzymes were separated by horizontal electrophoresis on agarose gels of between 0.3% and 1.2% concentration depending upon the frequency of cleavage sites, while samples digested with four-cutter enzymes were separated either on high concentration (1.8%) agarose gels or by vertical electrophoresis through 7% polyacrylamide gels to resolve fragments less than 1 kb in length. DNA was transferred to Hybond membrane (Amersham International) from agarose gels by the method of SOUTHERN (1975) or from polyacrylamide gels using a "Novablot" electroblotter (Pharmacia-LKB) at 0.5 A for 30 min. Hybridization and washing was carried out according to KREITMAN and AGUADE (1986); radioactive DNA probes were synthesized by oligo-labeling (FEINBERG and VOGELSTEIN 1983).

Cloned fragments used as probes: Cloned sequences homologous to the region were derived from ten overlapping λCharon 4 clones supplied by J. MODOLELL (CARRAMOLINO *et al.* 1982; CAMPUZANO *et al.* 1985). A map of the 110 kb covered by these clones is shown in Figure 1. A series of subclones into pUC8 (VIEIRA and MESSING 1982) was constructed by ROBIN BEECH and named as described in BEECH and LEIGH BROWN (1989). Those used as probes in this study were plasmids pASC133R1, 101R7, 101R5, 94R4, 64R3 and 53R1. To increase the intensity of hybridization, the cloned inserts were excised from the vectors by digestion with *Eco*RI, separated by agarose gel electrophoresis and recovered by rotating the gel through 90° and electrophoresing the fragment perpendicularly onto NA45 DEAE membrane (Schleicher and Schuell) followed by elution in TE containing 1 M NaCl. Excess NaCl was removed in a Sephadex G50 (Pharmacia-LKB) column before labeling. Some additional fragments which were purified directly from the phage clones in this way were used as probes. These are also shown on the map in Figure 1.

RESULTS

Strategy: Our aim was to study the relationship between molecular map distance and linkage disequilibrium. We surveyed several regions of the *yellow-achaete-scute* complex for restriction site variation additional to that described by BEECH and LEIGH BROWN (1989). The regions chosen were short (<3 kb in length), to allow greater accuracy in localizing the polymorphisms, and widely spaced (Figure 1), so as to span as much of the complex as possible. In order to increase the chance of detecting polymorphisms within or close to the probed regions, we screened fragments generated by a series of enzymes which recognize 4-bp restriction sites (KREITMAN and AGUADE 1986). The initial screen was carried out on a smaller sample than the disequilibrium analysis to save time because only those sites where heterozygosity (H) >0.25 will be useful in the detection of disequilibria in samples of less than 100 chromosomes (BROWN 1975). However, to avoid bias, our estimate of nucleotide diversity was based only on the results of the initial screen.

Heterozygosity at *achaete-scute*: Five probes were used to screen for four- and five-cutter polymorphisms: pASC133R1, pASC101R7, pASC101R5, pASC53R1 and pASC64R3. Eight restriction enzymes were used at all five regions: *Alu*I, *Ban*II, *Cfo*I, *Hae*III, *Mbo*I, *Msp*I, *Rsa*I and *Taq*I (*Acc*II was also used in one gel). We had no prior knowledge of their ability to detect variation in this region. Five polymorphic and 150 monomorphic sites were identified in an average of 16 lines screened for each combination of enzyme and probe. In Table 1 details of the gene frequencies and heterozygosities [$H = 2q(1 - q)$] are given for each of the polymorphisms.

From these data we obtain an estimate of nucleotide diversity (π) for the AS-C in the North Carolina population of 0.0024 (SD = 0.0013; NEI and TAJIMA 1981). We likewise obtain an estimate of θ (EWENS, SPIELMAN and HARRIS 1981; HUDSON 1982) of 0.0014 (SD = 0.0006). We estimated the standard deviation of our estimate of θ by jackknifing over sites. To the extent that sites provide independent replicates of the genetic sampling process, this procedure takes into account both genetic and statistical sampling (WEIR 1990). Our previous simulations to evaluate the utility of jackknifing for linked loci give us some confidence in this approach (REYNOLDS, WEIR and COCKERHAM 1983), but we note that we are likely to be underestimating the variance by using linked sites. These estimates do not differ greatly from those obtained earlier in this region on the same population sample with six-

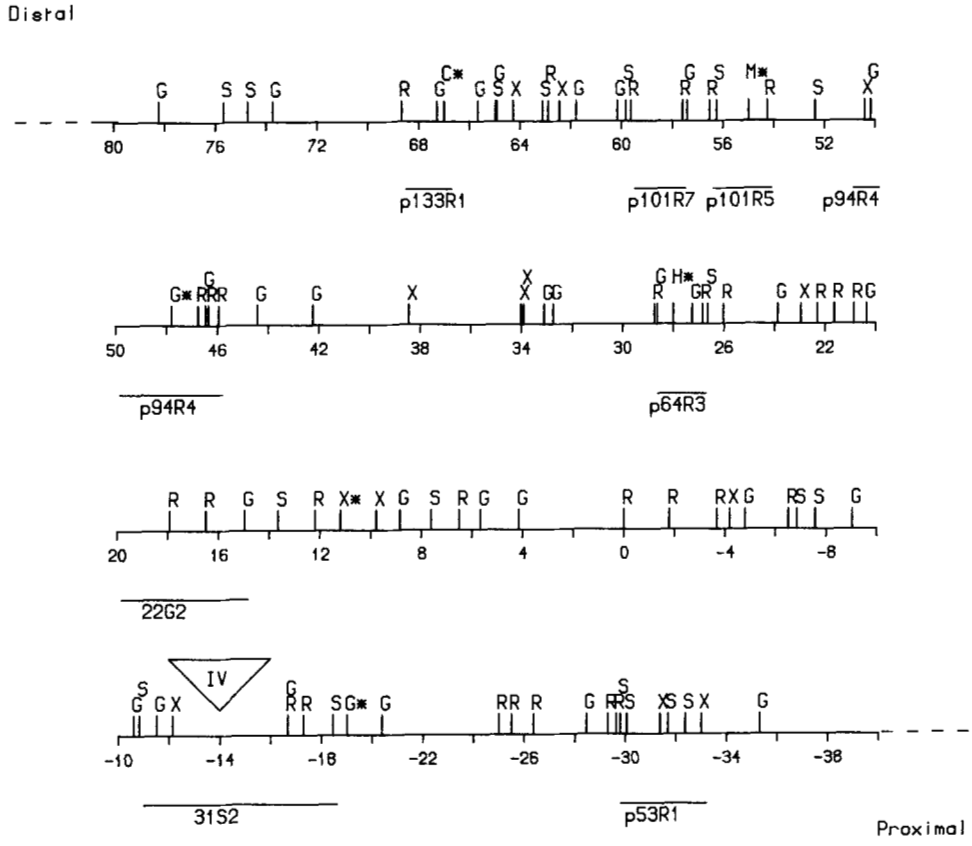

FIGURE 1.—Molecular map of the *achaete-scute* complex. The restriction map is taken from CAMPUZANO *et al.* (1985) and BEECH and LEIGH BROWN (1989). Coordinates are given in kilobases and restriction sites are abbreviated as follows: C, *Cfo*I; G, *Bgl*II; H, *Hae*III; M, *Msp*I; R, *Eco*RI; S, *Sal*I; X, *Xba*I. The polymorphic restriction sites used in the analysis of linkage disequilibrium are indicated (*) as is the polymorphic insertion (IV). Beneath the restriction map are shown the locations of plasmid subclones used in screening for polymorphic variants.

TABLE 1

Polymorphic restriction sites detected with 4-and 5-cutter restriction enzymes

Probe	Enzyme	*q*	*H*	*n*
133R1	*Cfo*I	0.26	0.39	34
133R1	*Msp*I	0.2	0.32	15
101R5	*Msp*I	0.16	0.27	38
53R1	*Hae*III	0.1	0.18	10
64R3	*Hae*III	0.14	0.24	42

Total number of monomorphic sites = 150 [137(4 cutter) + 13(5 cutter)]. Total number of sites screened = 155. Mean sample size (for all sites) = 16. *q*, gene frequency of rarer allele; *H*, exprected heterozygosity ($H = 2q(1 - q)$); *n*, sample size.

TABLE 2

Additional data on six-cutter polymorphisms in the achaete-scute complex (BEECH and LEIGH BROWN 1989)

Probe	Enzyme	*q*	*H*	*n*
101R7	*Xba*I	0.16	0.27	25
λsc17	*Xba*I	0.34	0.45	44
p31S2	*Bgl*II	0.43	0.49	42

Overall nucleotide diversity (including data from BEECH and LEIGH BROWN 1989):

Total polymorphic sites:	11 [5 × 4-cutter, 6 × 6-cutter]
Total monomorphic sites:	213 [137 (4), 13 (5), 63 (6)]
Proportion of shared sites:	0.9856
Mean sample size (weighted by the number of sites screened in each survey):	18.8
π:	0.0033
SD (NEI and TAJIMA, 1981):	0.0012
θ (EWENS, SPIELMAN and HARRIS 1981):	0.0018
SD (jackknifed over sites):	0.0005

Abbreviations as in Table 1.

cutter enzymes (0.0033 and 0.0024; BEECH and LEIGH BROWN 1989). Table 2 presents additional data on gene frequencies and heterozygosities for three of the six polymorphic sites described by BEECH and LEIGH BROWN (1989). Combining the data gives an overall estimate of 0.0033 (SD = 0.0012) for π and 0.0018 (jackknife SD = 0.0005) for θ (Table 2), from a total of 225 restriction sites (approximately 1047 nucleotide sites) with a mean sample size of 18.8 chromosomes.

Linkage disequilibrium at *achaete-scute*: A total of seven polymorphisms, six restriction sites and a polymorphic insertion, were included in the disequilibrium analysis. Three criteria were employed to select those used: (1) high heterozygosity ($H > 0.25$), (2) spacing across the 120 kb of the achaete-scute gene complex and (3) reliability of scoring. For each polymorphism chosen, the number of lines screened from the North Carolina 1984 sample was increased until *n* was at least 30 for all pairwise (and three-way) comparisons. The seven polymorphisms were distributed over approximately 86 kb of DNA and the

TABLE 3

Restriction map variants in the *achaete-scute* gene region

Line	Site and Position							Haplotype
	*Cfo*I 67	*Msp*I 55	*Bgl*II 48	*Hae*III 28	*Xba*I 11	Insertion IV	*Bgl*II −19	
NC1	+	+	−	+	+	−	+	1
2	+	+	−	+	+	−	−	2
3	+	+	−	+	−	−	+	3
4	−	−	+	−	−	+	−	4
5	+	+	−	+	−	−	−	5
6	−	−	+	−	−	+	−	4
7	+	+	−	+	+	−	+	1
8	−	−	+	−	−	+	−	4
9	+	+	−	+	−	−	−	5
10	+	+	−	+	+	−	+	1
11	+	+	−	+	+	−	+	1
12	+	+	+	+	−	−	−	6
13	+	+	−	+	+	−	+	1
14	+	+	−	+	+	−	+	1
15	−	−	+	−	−	+	−	4
16	−	+	+	+	−	−	−	7
17	+	+	−	+	−	−	+	3
18	+	+	−	+	+	−	+	1
19	−	−	+	−	−	+	−	4
20	n	n	+	+	−	−	−	6/7
21	+	+	−	+	−	−	+	3
22	+	+	−	+	−	+	−	8
23	n	n	−	n	+	−	+	1
24	n	n	−	n	−	−	+	3
25	n	n	−	+	−	−	−	5
26	−	+	+	+	−	−	−	7
27	n	n	+	+	−	−	−	6/7
28	+	+	+	+	−	n	n	6
29	−	+	+	+	−	−	−	7
30	+	+	−	+	+	−	−	2
31	+	+	−	+	+	−	−	2
32	+	+	−	+	+	−	−	2
33	n	+	+	+	−	−	−	6/7
34	+	n	−	+	−	−	−	5
35	+	+	−	+	−	−	+	3
36	−	−	+	−	−	n	n	4
37	n	+	−	+	−	−	−	5
38	n	+	−	+	−	−	+	3
39	+	+	+	+	−	−	−	6
40	+	+	−	+	+	−	+	1
41	+	+	+	+	−	−	−	6
42	n	+	−	+	+	+	+	9
43	+	+	−	+	−	−	+	3
44	n	+	−	+	+	−	+	1
Frequency	0.26	0.16	0.36	0.14	0.34	0.17	0.43	

haplotype of each chromosome studied is given in Table 3.

Pairwise disequilibrium coefficients for all two-way comparisons are given in Table 4. Estimates of D, D' (LEWONTIN 1964) and r^2 are given, along with those for the χ^2 statistic. This statistic can be derived in many ways (*e.g.* HILL and ROBERTSON 1968) but is most simply regarded as the squared estimate of D divided by an estimate (under the assumption that D is zero) of the variance of D (WEIR and COCKERHAM 1989). The test statistic has a χ^2 distribution when there is no disequilibrium, so the probability of observing such an association by chance can be found and this is also indicated in Table 3. Of the 21 pairwise comparisons, 17 are significant ($P < 0.05$) and 14 are highly significant ($P < 0.01$). Despite the fact that $n > 35$ for most of the tests, for all but 5 comparisons $D' = \pm 1$, indicating that all four allelic combinations were not present in the sample.

We have also estimated three-way disequilibrium coefficients (BENNETT 1954; THOMSON and BAUR 1984) between all 35 combinations of three loci from

TABLE 4

Disequilibrium in the *achaete-scute* gene region

Comparison	Distance/kb	D	D'	r^2	χ^2		n
G-19 × IV	5	−0.0476	−0.67	0.0667	2.8		42
G48 × M55	7	−0.0997	−1	0.321	12	**	38
C67 × M55	12	0.132	1	0.593	20	**	33
X11 × H28	17	0.0476	1	0.0833	3.5		42
G48 × C67	19	−0.163	−1	0.582	20	**	34
G48 × H28	20	−0.0884	−1	0.271	11	**	42
X11 × IV	25	−0.0357	−0.60	0.040	1.7		42
M55 × H28	27	0.133	1	1	38	**	38
X11 × G-19	30	0.109	0.53	0.211	8.8	**	42
X11 × G48	37	−0.124	−1	0.296	13	**	44
H28 × C67	39	0.130	1	0.595	20	**	34
H28 × IV	42	−0.103	−1	0.673	27	**	40
X11 × M55	44	0.0582	1	0.109	4.2	*	38
H28 × G-19	47	0.05	1	0.0952	3.8		40
X11 × C67	56	0.0934	1	0.196	6.7	**	34
G48 × IV	62	0.0635	0.57	0.131	5.5	*	42
G48 × G-19	67	−0.143	−1	0.375	16	**	42
M55 × IV	69	−0.112	−1	0.668	24	**	36
M55 × G-19	74	0.0617	1	0.129	4.6	*	36
C67 × IV	81	−0.109	−0.78	0.419	13	**	32
C67 × G-19	86	0.102	1	0.228	7.3	**	32

Seventeen of twenty-one comparisons significant * ($P < 0.05$), 14/21 highly significant ** ($P < 0.01$).

Table 3. The three-way coefficient measures the departure from random association between alleles at three loci over and above that due to the pairwise disequilibria. As in the two locus case, it is tested by dividing the squared estimate by an estimate of its variance. The expression for the variance in this case is a little cumbersome and is given by B. S. WEIR and G. B. GOLDING (in preparation).

The three-way disequilibrium coefficients are given in Table 5 with the variance of D and the χ^2 statistic. Twelve of the 35 three-way associations are significant; for seven, $P < 0.01$. In data sets of moderate size, it is less likely that significant three-way associations will be detected (BROWN 1975). The high proportion of significant disequilibria detected in this data set is remarkable.

Association of disequilibrium with molecular distance: In Table 4 the second column lists the distance between the restriction sites in kilobases. In Table 5 the distance spanned by the three sites is given for the comparisons where the three-way disequilibria are significant. In both cases, significant disequilibria are observed between loci separated by over 80 kb on the molecular map and there appears to be no decline in the level of association as the distance between the sites increases. The lack of relationship between the pairwise r^2 values and molecular distance is illustrated in Figure 2.

Phylogenetic relationships among AS-C haplotypes: A consequence of such a high level of disequilibrium is that the number of haplotypes present in the sample is very restricted. There are 128 possible haplotypes, of which in a sample of 44 chromosomes we would expect to see about 26 (based on randomly sampling a population with haplotype frequencies constructed from the allelic frequencies in Table 3 under the assumption of linkage equilibrium). Only 9 different haplotypes are observed. These can be related to one another in a simple tree with an essentially linear topology. In Figure 3 a Wagner network which connects the haplotypes with the minimum number of steps is depicted. The abundance of each haplotype in the sample is indicated by the diameter of the appropriate circle in Figure 3. The frequencies of many of the haplotypes are very similar, haplotype 1 being almost twice as common as haplotypes 2, 3, 4, 6 and 7, while 5, 8 and 9 are relatively rare.

Two points emerge from the phylogenetic analysis:

1. All but one of the haplotypes present is separated from its phylogenetic neighbor(s) by a single step. The exception, haplotype 4, is quite distinct, being three steps away from its nearest neighbor, haplotype 7. However, 4 is not a rare haplotype but is relatively common, being found in 6 out of 44 chromosomes screened (14%). While the structure of the rest of the tree would be consistent with divergence by single-step mutation or recombination events, the absence of haplotypes between 4 and 7 is notable.

2. Insertion IV, a large DNA insertion between coordinates −12 and −17 is unusual in that it reaches appreciable frequencies in the North Carolina population (BEECH and LEIGH BROWN 1989). It is probably equivalent to "Ins11" of AGUADE, MIYASHITA and LANGLEY (1989), which was also found at high fre-

TABLE 5

Three-way linkage disequilibrium in the *achaete-scute* gene complex

Comparison	D	Variance of D	χ^2		n	Maximum distance (kb)
C67 M55 G48	0.0501	0.0151	3.27		33	
C67 M55 H28	−0.0841	0.0086	6.93	**	33	39
C67 M55 X11	−0.0301	0.0106	2.87		33	
C67 M55 IV	0.0785	0.0121	6.78	**	31	80
C67 M55 G-19	−0.0327	0.0115	2.63		31	
C67 G48 H28	0.0513	0.0147	3.50		34	
C67 G48 X11	0.0220	0.0149	1.58		34	
C67 G48 IV	−0.0488	0.0153	3.33		32	
C67 G48 G-19	0.0317	0.0156	2.27		32	
C67 H28 X11	−0.0293	0.0101	2.93		34	
C67 H28 IV	0.0781	0.0128	7.05	**	32	80
C67 H28 G-19	−0.0317	0.0110	2.69		32	
C67 X11 IV	0.0234	0.0120	1.74		32	
C67 X11 G-19	−0.0137	0.0133	1.06		32	
C67 IV G-19	0.0254	0.0129	1.79		32	
M55 G48 H28	0.0682	0.0114	7.21	**	38	27
M55 G48 X11	0.0153	0.0095	1.30		38	
M55 G48 IV	−0.0643	0.0136	7.35	**	36	68
M55 G48 G-19	0.0206	0.0105	1.56		36	
M55 H28 X11	−0.0398	0.0095	4.46	*	38	44
M55 H28 IV	0.0808	0.0157	8.53	**	36	68
M55 H28 G-19	−0.0446	0.0114	4.46	*	36	74
M55 X11 IV	0.0369	0.0105	4.05	*	36	68
M55 X11 G-19	−0.0094	0.0078	1.01		36	
M55 IV G-19	0.0416	0.0112	4.48	*	36	74
G48 H28 X11	0.0113	0.0076	1.12		42	
G48 H28 IV	−0.0591	0.0142	7.58	**	40	61
G48 H28 G-19	0.0150	0.0084	1.31		40	
G48 X11 IV	−0.0108	0.0089	0.86		42	
G48 X11 G-19	0.0147	0.0155	0.96		42	
G48 IV G-19	−0.0113	0.0104	0.74		42	
H28 X11 IV	0.0316	0.0098	3.63		40	
H28 X11 G-19	−0.0038	0.0060	0.33		40	
H28 IV G-19	0.0356	0.0105	4.10	*	40	47
X11 IV G-19	0.0125	0.0087	1.50		42	

Twelve of thirty-five comparisons significant (*); 7 highly significant (**).

quency. In the current sample, the haplotypes in which it is present (4, 8 and 9) are distributed throughout the tree (Figure 3). Thus, haplotypes 8 and 9 can be regarded as haplotype 5 + insertion IV and haplotype 1 + insertion IV, respectively, and each is found only once in the sample. Any closer grouping of these haplotypes on the basis of the shared insertion would increase the total length of the tree considerably which suggests that they do not form a monophyletic group. It seems most probable therefore, that "Insertion IV" does not represent a unique event but has occurred independently on more than one occasion.

DISCUSSION

Linkage disequilibrium between restriction map variants: While significant gametic associations between restriction map variants were a prominent feature of early studies in *D. melanogaster* (LEIGH BROWN 1983; AQUADRO *et al.* 1986; CROSS and BIRLEY 1986), they were not so prevalent in more recent surveys at other loci (reviewed by LEIGH BROWN 1989). Only 6 out of 55 pairwise tests at the *Notch* locus were found to be significant (SCHAEFFER, AQUADRO and LANGLEY 1988). A survey of restriction map variation at the *white* locus by MIYASHITA and LANGLEY (1988) uncovered so many variable sites that over 3000 pairwise comparisons were made. In order to make an appropriate comparison with our data on the *achaete-scute* complex we consider only those sites at *white* for which neither allele had a frequency less than 0.2, a total of 34 sites. Taking the data of MIYASHITA and LANGLEY (1988), a total of 561 pairwise comparisons can be made between such sites; 69 of these are significant at 1% or below. However, for those pairs of sites which are separated by more than 5 kb, 29 out of a total of 390 comparisons (7.4%) show significant disequilibrium. As the lines used for the survey were from more than one population, it is likely that this figure

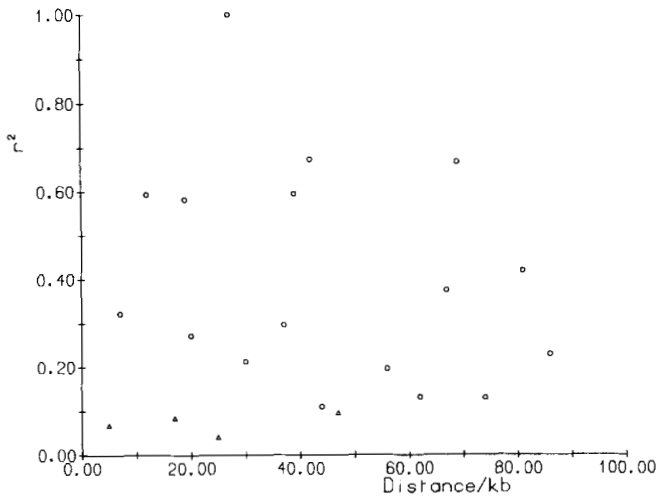

FIGURE 2.—Association between r^2 and molecular distance in the *achaete-scute* complex. The value of r^2 for all pairwise comparisons was obtained from Table 3 and plotted against the molecular distance between the two sites (in kilobases). Significant disequilibria ($P < 0.05$) are shown as circles, nonsignificant disequilibria as triangles.

is an overestimate of the number of significant disequilibria to be found in a single population.

We can conclude that while there are significant associations to be found between molecular variants at the *white* locus, they constitute a small minority of comparisons. Indeed, there are several examples where polymorphic sites separated by only a few hundred base-pairs are apparently in linkage equilibrium. This overall picture also applies to the *rosy* locus, where no significant associations were detected by AQUADRO, LADO and NOON (1988) in *D. melanogaster* and again, some sites only a few hundred base-pairs apart are in linkage equilibrium (J. N. MACPHERSON, unpublished data).

We have been able to extend the earlier results on disequilibrium in the *achaete-scute* complex (BEECH and LEIGH BROWN 1989; AGUADE, MIYASHITA and LANGLEY 1989; EANES, LABATE and AJIOKA 1989) by detecting and including more polymorphic sites in our analysis. However, while our results are in substantial agreement with the earlier work on the AS-C, they contrast strongly with the recent observations from other loci. While we have used only those combinations of sites where the observed gene frequencies render the tests reasonably powerful, the very high proportion of significant pairwise tests is unique so far for any gene in *D. melanogaster*. Even more remarkable is the detection, for the first time in an outbred natural population, of significant three-way disequilibria. As noted earlier, both pairwise and three-way associations extend across more than 80 kb, almost the whole of the achaete-scute complex. The region spanned includes several transcription units, some of which appear to be involved in neurogenesis (JIMENEZ and CAMPOS-ORTEGA 1987), others in epidermal pigmentation (BIESSMAN 1985; CHIA *et al.* 1986), as well as a region (*sis-b*; CLINE 1988) apparently involved in the assessment of the *X*-autosome ratio. From the data we have presented, these genes cannot be considered to be evolving independently in *D. melanogaster*.

It is well known that disequilibrium can be generated in several ways, for example, by sampling in finite populations and by selection. In the case of the *achaete-scute* complex the most significant factor is almost certainly genetic: the virtual absence of recombination in this region of the *X* chromosome. Recombination between *achaete* and *scute* was reported by DUBININ, Sokolov and TINIAKOV (1937) and an estimate of 1.2×10^{-4} cM/kb can be obtained from their data (BEECH and LEIGH BROWN 1989). This is approximately an order of magnitude below that found at the *white* locus (LANGLEY and AQUADRO 1987), but it is almost certainly an overestimate because of the presence of recombination suppressors in other parts of the genome during the experiments. It appears that the distance over which significant gametic associations can be found at *achaete-scute* is almost two orders of magnitude greater than that for *white*.

Nucleotide diversity at *achaete-scute*: The estimate of nucleotide diversity obtained in the course of this work was, at 0.0024, in line with that obtained in our previous survey of the region in this population (BEECH and LEIGH BROWN 1989). Pooling the data from the two surveys leads to an overall estimate of 0.0033 ± 0.0012, based on 12 enzymes and about 1047 nucleotide sites. This estimate is low for *D. melanogaster* but not significantly different from those obtained previously for *Notch*, *rosy* or the 87A heat shock locus (reviewed by LEIGH BROWN 1989). In a second survey of the *achaete-scute* complex, using 7 six-cutter enzymes, all but one of which were different to those used here, AGUADE, MIYASHITA and LANGLEY (1989), reported that the nucleotide diversity of the region was markedly lower than for other regions of the *D. melanogaster* genome. Specifically, they obtained an estimate of π for the same North Carolina population of 0.00057, although their estimate of θ (0.00128) was closer to ours (0.0018; Table 2). This was interpreted as evidence for "hitchhiking" effects (MAYNARD SMITH and HAIGH 1974), which might be expected as a consequence of selection acting in a region of reduced recombination. A third survey of two parts of the complex by EANES, LABATE and AJIOKA (1989) gave a similar estimate to that of BEECH and LEIGH BROWN (1989) and the present work; in their sample $\pi = 0.0022$ and $\theta = 0.0012$ for a New York population.

In view of the common material, approaches and indeed, results, with respect to the frequency of DNA insertions in the region of all three groups, it is surprising to note such a discrepancy in the estimate of

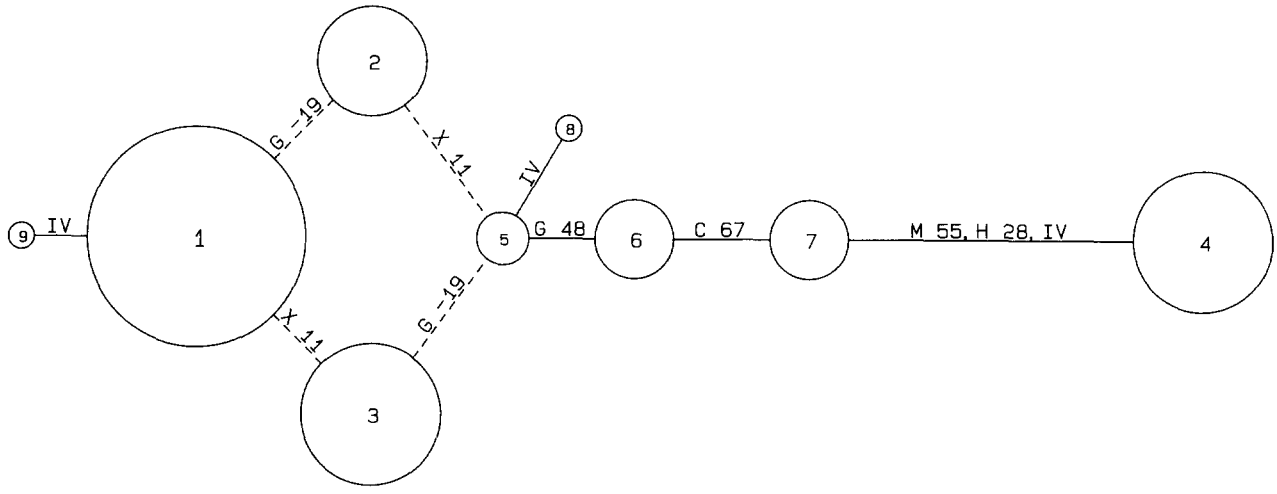

FIGURE 3.—Relationships between restriction-map haplotypes in the *achaete-scute* region. A Wagner network which connects the nine haplotypes with the shortest overall length is presented. The diameter of each circle is proportional to the frequency of that haplotype in the North Carolina population. Alternative paths of equal length are indicated by the dotted lines. The sites at which two adjacent haplotypes differ are given above the line which connects them.

mean nucleotide diversity. One difference lies in the enzymes used, and thus the oligonucleotide sequences surveyed. The distribution of restriction sites in genomic DNA is highly nonrandom, and it is conceivable that there are as yet unknown biases in the distribution of nucleotide sequence variants which might lead to this discrepancy. Second, the allele frequency distribution for restriction sites in this region may be unusual, with a small number of sites for which H is high and a very large number with low or very low heterozygosity. In that case the sampling error of π would be increased, but the estimate of θ would not be affected as a neutral model is assumed (discussed by AGUADE, MIYASHITA and LANGLEY 1989). However, our estimate of θ for *achaete-scute* also does not differ significantly from that obtained for another *X*-linked gene, *forked* ($\theta = 0.002$; C. H. LANGLEY and N. MIYASHITA, unpublished data cited by AGUADE, MIYASHITA and LANGLEY 1989), or from that given for the 87A heat shock locus ($\theta = 0.002$; LEIGH BROWN 1983). As we have indicated above, the jackknifing procedure we have used to estimate the variance of would provide a minimum estimate for linked sites. This, in turn would make us more likely to declare our estimate to be different from those found for other regions–so that we are being conservative in suggesting there is no difference. The absence of recombination in the AS-C appears not to have resulted in a significant reduction in variability below those found for some other regions of the genome, although DNA sequence data may be necessary to resolve the issue.

We are grateful to A. E. SHRIMPTON for collecting flies, to J. MODOLLEL for supplying lambda clones of the *achaete-scute* region and to R. N. BEECH for advice. This work was supported by the Science and Engineering Research Council and by National Institutes of Health grant GM32518 to B. S. W.

LITERATURE CITED

AGUADE, M., N. MIYASHITA and C. H. LANGLEY, 1989 Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122:** 607–615

ANTONARAKIS, S. E., C. D. BOEHM, P. J. GIARDINA and H. H. KAZAZIAN, 1982 Non random association of polymorphic restriction sites in the β-globin cluster. Proc. Natl. Acad. Sci. USA **79:** 137–141.

AQUADRO, C. F., K. M. LADO and W. A. NOON, 1988 The *rosy* region of *Drosophila melanogaster* and *D. simulans*. I. Contrasting levels of naturally occurring restriction map variation and divergence. Genetics **119:** 875–888.

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.

BEECH, R. N., and A. J. LEIGH BROWN, 1989 Insertion-deletion variation at the *yellow, achaete-scute* region in two natural populations of *Drosophila melanogaster*. Genet. Res. **53:** 7–15.

BENNETT, J. H., 1954 On the theory of random mating. Ann. Eugen. **18:** 311–317.

BIESSMAN, H., 1985 Molecular analysis of the yellow gene (*y*) region of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **82:** 7369–7373.

BROWN, A. H. D., 1975 Sample sizes required to detect linkage disequilibrium between two or three loci. Theor. Popul. Biol. **8:** 184–201.

CAMPUZANO, S., L. CARRAMOLINO, C. V. CABRERA, M. RUIZ-GOMEZ, R. VILLARES, A. BORONAT and J. MODOLELL, 1985 Molecular genetics of the *achaete-scute* gene complex of *Drosophila melanogaster*. Cell **40:** 327–338.

CARRAMOLINO, L., M. RUIZ-GOMEZ, M. C. GUERRERO, S. CAMPUZANO and J. MODOLELL, 1982 DNA map of mutations at the *scute* locus of *Drosophila melanogaster*. EMBO J. **1:** 1185–1191.

CHIA, W., G. HOWES, M. MARTIN, Y. B. MENG, K. MOSES and S. TSUBOTA, 1986 Molecular analysis of the *yellow* locus of Drosophila. EMBO J. **5:** 3597–3605.

CLINE, T. W., 1988 Evidence that *sisterless-a* and *sisterless-b* are two of several discrete "numerator elements" of the *X/A* sex determination signal in Drosophila that switch *Sxl* between two alternative stable expression states. Genetics **119:** 829–862.

CROSS, S. R. H., and A. J. BIRLEY, 1986 Restriction endonuclease

map variation in the *Adh* region in populations of *Drosophila melanogaster*. Biochem. Genet. **24:** 415–433.

DUBININ, N. P., N. N. SOKOLOV, and G. G. TINIAKOV, 1937 Crossing over between the genes *yellow*, *achaete* and *scute*. Drosoph. Inform. Serv. **8:** 76.

EANES, W. F., J. LABATE and J. W. AJIOKA, 1989 Restriction-map variation with the *yellow-achaete-scute* region in five populations of *Drosophila melanogaster*. Mol. Biol. Evol. **6:** 492–502.

EWENS, W. J., R. S. SPIELMAN and H. HARRIS, 1981 Estimation of genetic variation at the DNA level from restriction endonuclease data. Proc. Natl. Acad. Sci. USA **78:** 3748–3750.

FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6–13.

HEDRICK, P., S. JAIN and L. HOLDEN, 1978 Multilocus systems in evolution. Evol. Biol. **11:** 101–182.

HILL, W. G., and A. ROBERTSON, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

HUDSON, R. R., 1982 Estimating genetic variability with restriction endonucleases. Genetics **100:** 711–719.

JIMENEZ, F., and J. A. CAMPOS-ORTEGA, 1987 Genes in subdivision 1B of the *Drosophila melanogaster* X-chromosome and their influence on neural development. J. Neurogenet. **4:** 179–200.

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

LANGLEY, C. H., 1977 Non random associations between allozymes in natural populations of *Drosophila melanogaster*, pp. 265–273 in *Measuring Selection in Natural Populations*, edited by F. B. CHRISTIANSEN and T. M. FENCHEL. Springer-Verlag, Berlin.

LANGLEY, C. H., E. MONTGOMERY and W. F. QUATTLEBAUM, 1982 Restriction map variation in the *Adh* region of Drosophila. Proc. Natl. Acad. Sci. USA **79:** 5631–5635.

LANGLEY, C. H., A. E. SHRIMPTON, T. YAMAZAKI, N. MIYASHITA, Y. MATSUO and C. F. AQUADRO, 1988 Naturally occurring variation in the restriction map of the *Amy* region of *Drosophila melanogaster*. Genetics **119:** 619–629.

LEIGH BROWN, A. J., 1983 Variation at the 87A heat shock locus in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **80:** 5350–5354.

LEIGH BROWN, A. J., 1989 Population genetics at the molecular level: a review of the contribution of restriction enzyme studies, pp. 207–242 in *Oxford Surveys in Evolutionary Biology*, Vol. 6, edited by P. HARVEY and L. PARTRIDGE. Oxford University Press, Oxford.

LEIGH BROWN, A. J., and J. E. MOSS, 1987 Transposition of the I element and *copia* in a natural population of *Drosophila melanogaster*. Genet. Res. **49:** 121–128.

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

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

MIYASHITA, N., and C. H. LANGLEY, 1988 Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. Genetics **120:** 199–212.

NEI, M., and F. TAJIMA, 1981 DNA polymorphism detectable by restriction endonucleases. Genetics **97:** 145–163.

REYNOLDS, J., B. S. WEIR and C. C. COCKERHAM, 1983 Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics **105:** 767–779.

SCHAEFFER, S. W., C. F. AQUADRO and W. W. ANDERSON, 1987 Restriction-map variation in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. Mol. Biol. Evol. **4:** 254–265.

SCHAEFFER, S. W., C. F. AQUADRO and C. H. LANGLEY, 1988 Restriction-map variation in the *Notch* region of *Drosophila melanogaster*. Mol. Biol. Evol. **5:** 30–40.

SOUTHERN, E. M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517.

STEPHAN, W., and C. H. LANGLEY, 1989 Molecular genetic variation in the centromeric region of the *X* chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. Genetics **121:** 89–99.

THOMSON, G., and M. P. BAUR, 1984 Third order linkage disequilibrium. Tissue Antigens **24:** 250–255.

VIEIRA, J., and J. MESSING 1982 The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:** 259–268.

VILLARES, R., and C. V. CABRERA, 1987 The *achaete-scute* gene complex of *Drosophila melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to *myc*. Cell **50:** 415–424.

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

WEIR, B. S., and C. C. COCKERHAM, 1989 Analysis of disequilibrium coefficients, pp. 45–51 in *Evolution and Animal Breeding*, edited by W. G. HILL and T. F. C. MACKAY. C.A.B. International, Wallingford.

Communicating editor: M. TURELLI