Genet. Res., Camb. (1993), **62**, *pp.* 101–109 *With 4 text-figures*

Linkage disequilibrium in the *white* locus region of *Drosophila melanogaster*

NAOHIKO T. MIYASHITA[1]*, MONTSERRAT AGUADÉ[2] AND CHARLES H. LANGLEY[3]

[1] *Laboratory of Genetics, Faculty of Agriculture, Kyoto University, Kyoto 606, Japan*
[2] *Departament de Genètica, Universitat de Barcelona, Barcelona 08071, Spain*
[3] *Department of Genetics, University of California, Davis, California 95616, USA*

(*Received 24 March 1993 and in revised form 24 May 1993*)

Summary

Linkage disequilibrium between molecular polymorphisms in a 10 kb region in the *white* locus of *Drosophila melanogaster*, revealed with a battery of four-cutter restriction enzymes, was investigated in 266 lines sampled from seven natural populations around the world. A total of 73 (35 restriction site, 37 insertion/deletion and 1 inversion) polymorphisms were detected, of which 55 non-unique polymorphisms were analysed for linkage disequilibrium. Clustering of significant linkage disequilibrium was observed in the transcriptional unit of the *white* locus as in Miyashita & Langley (1988). It was shown that about two thirds of the 2-locus combinations showing significant linkage disequilibrium have similar degree and direction of association over different populations. Despite lower divergence in allelic frequencies of molecular polymorphisms among populations, an increase in the proportion of 2-locus pairs showing significant linkage disequilibrium is observed in the transcriptional unit. Large values of Ohta's D measure ratio (1982*a*, *b*) cluster in the transcriptional unit, and correspond to significant linkage disequilibria. Although the exact molecular mechanism is not clear, these results suggest that epistatic selection is responsible for significant linkage disequilibrium in the transcriptional unit of this locus.

1. Introduction

Linkage disequilibrium can be an informative measure which can discriminate among the forces maintaining genetic variations in natural populations. Population genetics theory has shown that there are two main causes of linkage disequilibrium (Kimura & Ohta, 1971; Lewontin, 1974; Kimura, 1983). One is natural selection on epistatically determined fitness (Kimura, 1956; Lewontin & Kojima, 1960) and the other is random genetic drift (Hill & Robertson, 1968; Ohta & Kimura, 1969). In order to investigate the relative contribution of genetic mechanisms which are determining molecular evolution, it is important to know not only how much genetic variation exists in natural populations (frequency) but also how genetic variation is organized in the genome (linkage disequilibrium).

A general conclusion drawn from extensive protein electrophoresis studies during the 1970s was that little linkage disequilibrium exists between allozyme loci in natural populations of *Drosophila melanogaster* and other *Drosophila* species (Prakash & Lewontin, 1968; Kojima, Gillespie & Tobari, 1970; Mukai, Mettler & Chigusa, 1971; Charlesworth & Charlesworth, 1973; Zouros & Krimbas, 1973; Mukai, Watanabe & Yamaguchi, 1974; Langley, Tobari & Kojima, 1974; Mukai & Voelker, 1977). These results have been taken as an evidence against strong epistatic selection at the genic level in nature. At the chromosomal level, the story is more complicated. It has been shown that non-random association of many allozyme loci with cosmopolitan polymorphic inversions exists in natural populations of *Drosophila* species (Prakash & Lewontin, 1968; Kojima *et al.* 1970; Mukai *et al.* 1971; Voelker *et al.* 1978; Yamaguchi *et al.* 1980; Prevosti *et al.* 1983). Although complete agreement has not been obtained, these observations were explained by some form of selective advantage in the polymorphic inversions, and by reduced recombination rate in the regions proximal to the inversions where particular allozyme loci are located (Dobzhansky, 1970).

During the past 10 years, population genetic study has moved to the DNA level and has shown that there is a considerable amount of linkage disequilibrium between molecular polymorphisms in natural popula-

* Corresponding author.

tions of *Drosophila* (Langley, Montgomery & Quattlebaum, 1982; Kreitman, 1983; Aquadro *et al.* 1986; Langley & Aquadro, 1987; Langley *et al.* 1988; Schaeffer, Aquadro & Langley, 1988; Miyashita & Langley, 1988; Aguadé, Miyashita & Langley, 1989; Riley, Hallas & Lewontin, 1989; Eanes *et al.* 1989; Miyashita, 1990; Macpherson, Weir & Leigh Brown, 1990). Most notable are allozyme locus regions (*Adh*, *Amy* and *Zw*) where strong non-random associations between electromorphs and some molecular polymorphisms (nucleotide substitution and/or insertion/deletion) were detected. Although a direct link has not been found yet, the involvement of some molecular polymorphisms with enzyme activity variation was also suggested (Aquadro *et al.* 1986; Laurie-Ahlberg & Stam, 1987; Miyashita, 1990; Laurie, Bridgham & Choudhary, 1991).

Strong linkage disequilibrium at the DNA level might easily be explained by the phylogenetic relationship and the relatively close linkage between molecular polymorphisms, since molecular study usually reveals variation in only a very short segment of DNA. A distance effect on linkage disequilibrium was examined in the *white* locus region, and it was shown that polymorphisms separated more than 2 kb seem to have no detectable level of non-random association (Miyashita & Langley, 1988). This result suggested that a small amount of recombination is sufficient to randomize polymorphisms at the DNA level in nature, and that linkage disequilibrium might not be explained simply by the small distance between polymorphisms. Lack of a distance effect on linkage disequilibrium, even between very close polymorphisms (< 600 bp or less), was also noted in the *Xdh* region of *D. pseudoobscura* (Riley *et al.* 1989). However, it should be mentioned that reduced recombination rate clearly influences overall level of linkage disequilibrium. Aguadé *et al.* (1989) reported that the proportion of pairs showing significant linkage disequilibrium in the *yellow–achaete–scute* region (at the tip of the X chromosome where recombination is reduced) of *D. melanogaster* is much larger than those observed in regions with normal level of recombination. Similar observations of high level of linkage disequilibrium in the regions where recombination rate seems to be reduced were obtained in different samples of *D. melanogaster* and *D. simulans* (Macpherson *et al.* 1990; Miyashita, 1990; Begun & Aquadro, 1992; Martín-Campos *et al.* 1992). In these regions the level of molecular variation is extremely low. In order to explain both low variation and high level of linkage disequilibrium in these regions, a model of the fixation of favourable mutation and subsequent hitch-hiking of neutral polymorphisms in the surrounding region has been proposed (Maynard Smith & Haigh, 1974; Kaplan, Hudson & Langley, 1989). The high level of linkage disequilibrium was explained as the secondary consequence of hitch-hiking due to advantageous mutations. Taken together these result suggest that linkage disequilibrium at the molecular level is not a simple function of distance nor recombination rate, and that a careful examination is required in order to infer the mechanism for linkage disequilibrium.

Miyashita & Langley (1988) have also reported an unexpected feature of molecular polymorphism that there is a clustering of significant linkage disequilibrium between DNA polymorphisms in the transcriptional unit, especially 3′ end of the first intron, of the *white* locus of *Drosophila melanogaster*. The clustering was not detected in the flanking region of the locus despite non-heterogeneous allele frequency distribution and similar level of recombination rate between those two functionally different regions (Green, 1959; Judd, 1964). A more detailed reexamination of this curious observation was the goal of the present study. The clustering of highly significant linkage disequilibrium was confirmed again in new population samples.

2. Materials and methods

(i) *Lines*

A total of 266 X chromosome lines were used. The four newly analysed populations comprise 50 lines from Tanoshumaru, Fukuoka, Japan; 53 from Barcelona, Spain; 50 from Raleigh, North Carolina; 49 from Florida City, Florida. The other 64 lines included for analysis (20 North Carolina, 27 Texas and 17 Fukuoka) were described in Miyashita & Langley (1988). The data table will be available upon request from the first author.

(ii) *Restriction map analysis*

The '4-cutter analysis' of Kreitman & Aguadé (1986) was used. In this experiment six different 4-cutter restriction enzymes were used: *Alu* I, *Dde* I, *Hha* I, *Hinf* I, *Msp* I, and *Taq* I. Procedures for DNA extraction, digestion, blotting, probe preparation, hybridization, and washing were the same as in Miyashita & Langley (1988).

(iii) *Analyses of linkage disequilibrium*

Two different analyses were performed to test the significance of linkage disequilibrium. The first analysis was the same as in Miyashita & Langley (1988), where χ^2 values calculated from each population sample are assumed to be independent; then the pooled χ^2 value over populations is tested for significance of linkage disequilibrium with the corresponding degrees of freedom (Analysis I). The second analysis follows a standard statistical procedure of testing heterogeneity of correlation coefficients [Steel & Torrie (1980), p. 280; Weir (1990), p. 110; W. G. Hill, pers. comm.]. Linkage disequilibrium

Table 1. *Summary of restriction map polymorphism in the* white *locus region of* Drosophila melanogaster

Population	Sample size	Nucleotide variation (10^{-3}) θ	π	Haplotype diversity	Tajima's D
Japan 90 Tanoshumaru	50	3·78	2·86	0·912 (0·821)	−0·64 (−0·02)
Spain 90 Barcelona	53	3·55	3·58	1·00 (0·986)	0·27 (0·96)
North Carolina 90 Raleigh	50	3·40	3·43	0·999 (0·982)	0·25 (0·23)
Florida 90 Florida City	49	5·08	4·11	0·997 (0·985)	−0·32 (0·20)
North Carolina 88 Durham	20	4·44	4·45	0·994 (0·973)	0·29 (0·99)
Texas 88 Austin	27	4·27	4·43	0·997 (0·988)	0·41 (0·85)
Japan 88 Yame	17	3·34	3·24	0·985 (0·919)	0·08 (−0·08)

θ is an estimate of $4Nu$ according to Ewens, Spielman & Harris (1981) and Hudson (1982). π is the nucleotide diversity by Nei & Li (1979) and Nei & Tajima (1981). Haplotype diversity (upper value) is based on all the 55 polymorphisms and lower value in parentheses on 35 restriction site polymorphisms only. Tajima's D (upper value) is based on 35 restriction site polymorphisms and lower value on 38 insertion/deletion polymorphisms. None of Tajima's D are significant.

is expressed as the correlation coefficient of allele frequencies between two polymorphisms for each population sample. Linkage disequilibrium was tested for the mean and heterogeneity among populations by χ^2 (Analysis II). The subdivision of linkage disequilibrium measure follows Ohta (1982*a*, *b*). For each pair of polymorphisms, the expectation in equations (10)–(14) of Ohta (1982*a*) was taken as the average over populations.

3. Results

(i) *Restriction map variation*

A total of 179 restriction sites were scored, of which 35 were found polymorphic. In addition, 37 insertion/deletion and 1 inversion polymorphisms were detected. The inversion polymorphism, located in the 5′ flanking region, was found uniquely in the new Raleigh sample. Table 1 shows the estimates of nucleotide variation (Nei & Li, 1979; Ewens, Spielman & Harris, 1981; Nei & Tajima, 1981; Hudson, 1982) and haplotype diversity (Nei & Tajima, 1981) of each population. All the populations are generally similar to each other in the level of nucleotide variation and show very high haplotype diversity. Although fewer 4-cutter restriction enzymes were used in the present study, the estimates of nucleotide variation were in good agreement with those obtained previously (Miyashita & Langley, 1988).

Tajima's test (1989) was conducted in order to determine if the assumptions of selective neutrality of molecular polymorphisms and stochastic stationarity apply to the present population samples. Table 1 shows Tajima's D values for both restriction site and insertion/deletion polymorphisms for each population. None of the values deviate significantly as was found for the same region by Tajima (1989). This supports the view that the frequency spectrum of polymorphisms does not deviate from that expected with stationarity under neutral molecular evolution.

Heterogeneity among the seven populations of the frequencies was tested for each of the 55 non-unique polymorphism and 1485 2-locus combinations by χ^2 contingency test. Significant heterogeneity was detected in 36 single-locus tests and 1142 combinations for 2-locus tests. These large numbers of significant cases may be due to the existence of population-specific polymorphisms. There are 231 haplotypes among 266 lines studied. Four haplotypes are found over different populations: one haplotype in Florida 90 and North Carolina 88, and three in Japan 90 and Japan 88. For the following analyses of linkage disequilibrium, each population was assumed to be an independent sample.

(ii) *Linkage disequilibrium*

Fifty five molecular variations (both restriction site and insertion/deletion) which were found more than once in the whole sample were used for linkage analyses, as in Miyashita & Langley (1988). Figures 1 and 2 (below the diagonal) show the significance levels

https://doi.org/10.1017/S0016672300031694 Published online by Cambridge University Press

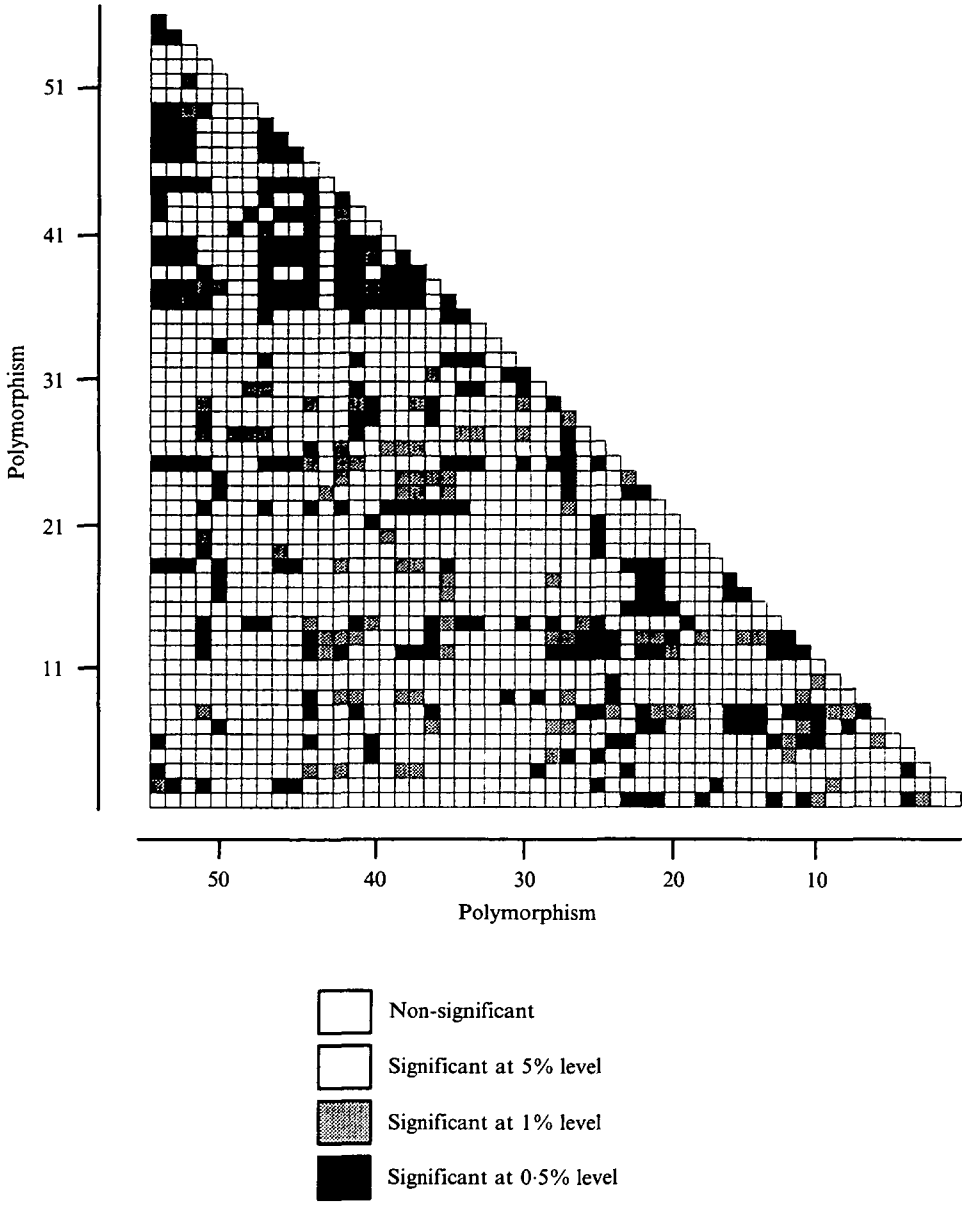

Fig. 1. Significance of linkage disequilibrium detected by Analysis I. Fifty-five non-unique polymorphisms are used for the analysis. Polymorphisms are numbered in the 5′ to 3′ order along the *white* locus region. The transcriptional unit starts at polymorphism no. 29.

of linkage disequilibrium detected by Analyses I and II, respectively. As has been observed, a clustering of highly significant linkage disequilibria in the transcriptional unit was confirmed by both Analyses I and II. It can be noted that there are patches of significant linkage disequilibria scattered in the entire region, and that those patches are disrupted by regions of non-significant linkage disequilibria. In other words, significant non-random associations do not appear to be distributed uniformly or in any continuous pattern. Relationship between linkage disequilibrium (χ^2 of mean effect in Analysis II) and distance between polymorphisms was investigated as before. Pairs of polymorphisms located within 2 kb of one another often display noticeable disequilibrium, while pairs further apart are usually in linkage equilibrium (data not shown). This result is very similar to that obtained previously.

The significance level of heterogeneity in linkage disequilibrium among populations is shown in the upper half of Fig. 2 (above the diagonal). Significant heterogeneity was detected in 201 out of 1140 total combinations. Because some polymorphisms are observed in only one population, the total number of tests for heterogeneity is less than 1485 ($55 \times 54/2$). The probability of observing 201 or more significant cases (at the $P < 0{\cdot}05$ level) among all the combinations if they are independent is less than 10^{-4}. This suggests significant heterogeneity in linkage disequilibrium among the sampled populations. It is of interest to compare the distribution for significant mean linkage disequilibria with that for significant

https://doi.org/10.1017/S0016672300031694 Published online by Cambridge University Press

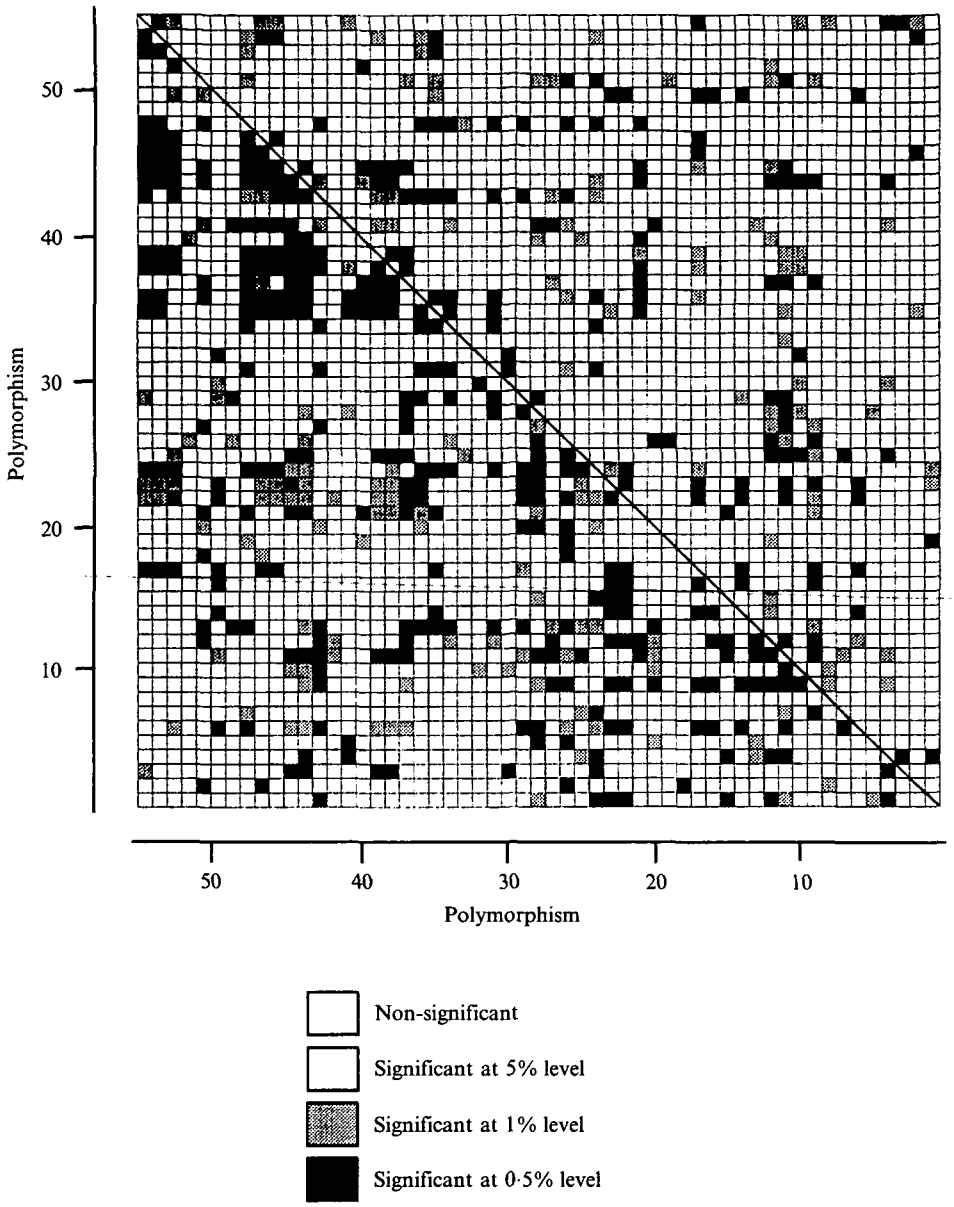

Fig. 2. Significance of mean (below the diagonal) and heterogeneity (above) of linkage disequilibrium detected by Analysis II.

Table 2. *Relationship between linkage disequilibrium and its heterogeneity among populations. The number of pair-wise combinations is shown*

Linkage disequilibrium... Heterogeneity...	Significant Significant	Significant Non-significant	Non-significant Significant	Non-significant Non-significant
Within 5′ flanking region	43 (14·8%)	55 (18·9%)	21 (7·2%)	172 (59·1%)
Within Transcriptional unit	43 (15·4%)	76 (27·2%)	15 (5·4%)	145 (52·0%)
Between 5′ flanking and transcriptional unit	48 (8·2%)	83 (14·6%)	43 (7·5%)	396 (69·5%)
Entire region	134 (11·8%)	214 (18·8%)	79 (6·9%)	713 (62·5%)

heterogeneity of linkage disequilibria. Table 2 shows the number of pairs of polymorphisms showing significant or non-significant mean linkage disequilibria and heterogeneity in various parts of the *white* locus. The proportion of significant pairs is higher in the transcriptional unit (43%) than in the 5′ flanking region (34%). It should be noted that among the pairs showing significant linkage disequilibrium, more than

https://doi.org/10.1017/S0016672300031694 Published online by Cambridge University Press

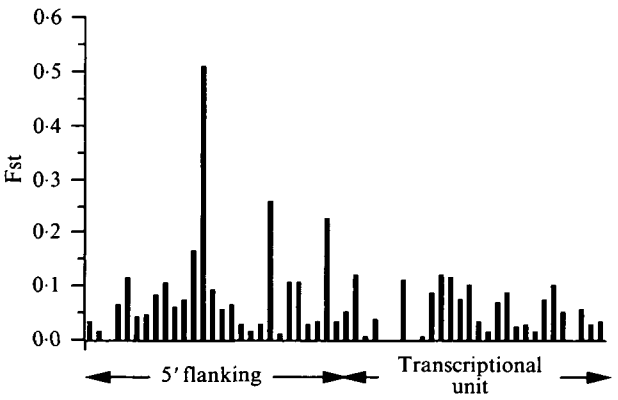

Fig. 3. Estimate of Fst value among populations for each of 55 non-unique polymorphisms used for analyses of linkage disequilibrium.

60% of pairs do not show significant heterogeneity over populations. In other words, when linkage disequilibrium is significant, the degree and direction of linkage disequilibrium tend to be similar among different populations. This tendency is observed in all categories of regions (Table 2).

In order to compare the transcriptional unit and the 5′ flanking region for each population, heterogeneity of polymorphism frequency distribution was tested with contingency χ^2, after classifying polymorphisms in 0·1 frequency intervals. Only the Japan 88 (Yame) population shows marginally significant heterogeneity between the two regions, but none of the other tests are significant. The number of lines in the Yame population is too small (17) for the test result to be meaningful. These results suggest that the two regions are not significantly different with respect to frequency distribution, consistent with the result obtained before (Miyashita & Langley, 1988). It may be concluded that the difference in the frequency distribution is not responsible for the increase (or clustering) of significant linkage disequilibrium in the transcriptional unit.

A simple explanation to account for the increased proportion of significant pairs in the transcriptional unit is geographic differentiation in allele frequency of polymorphisms. If populations have diverged in allele frequency, significant linkage disequilibrium would be detected in the pooled sample (or mean). The expectation was that the transcriptional unit might be more diverged than the flanking region. This possibility was examined by calculating the fixation index (Wright, 1951; Cockerham, 1969, 1973; Nei, 1973). The result is contrary to the above expectation (Fig. 3). Allele frequencies in the transcriptional unit are not diverged compared to those in the 5′ flanking region. This result suggests that the increased proportion of significant linkage disequilibrium in the transcriptional unit can not be explained by geographic differentiation of allele frequency.

Figure 4 shows the ratios of the variance components of linkage disequilibrium, Dis^2/Dst^2 and Dst'^2/Dis'^2, proposed by Ohta (1982*a*, *b*). It was shown that these ratios become larger than one when epistatic natural selection is responsible for linkage disequilibrium, while they are smaller than one when linkage disequilibrium is caused by genetic drift. Most pairs (about 80%) are less than 0·1. Only a small number of pairs have a ratio larger than one. It is evident that relatively large values (including more than one) cluster in the transcriptional unit of the *white* locus, and that this clustering of large values of Ohta's ratio corresponds to that of significant linkage disequilibria detected above. For the combinations of polymorphisms in the 5′ flanking region, and those between 5′ flanking and transcriptional unit, most pairs have very small values. As expected, Dis^2/Dst^2 becomes large more often than Dst'^2/Dis'^2. These results suggest that epistatic natural selection is responsible for significant linkage disequilibria between molecular polymorphisms in the transcriptional unit.

4. Discussion

The clustering of highly significant linkage disequilibria in the transcriptional unit of the *white* locus of *Drosophila melanogaster* was observed again in this report, as detected by Miyashita & Langley (1988). It can be concluded that the clustering is not specific to those population samples in the previous study, and that many molecular polymorphisms in the transcriptional unit of the *white* locus do not associate at random. The clustering was seen as an increase of significant pairs in the transcriptional unit. Forty-three percent of 2-locus combinations between molecular polymorphisms are significant in the transcriptional unit, while 34% are significant in the 5′ flanking region.

Comparisons between the 5′ flanking region and the transcriptional unit suggest that the different level of significant linkage disequilibrium between the two regions can not be attributed to differences between the regions in the frequencies of polymorphisms. First, the frequency spectra of polymorphisms in the two regions are similar. Second, the density of polymorphisms is higher in the flanking region than in the transcriptional unit (see Fig. 1 of Miyashita & Langley, 1988), although the distance effect on linkage disequilibrium between molecular polymorphisms does not seem to be important. Third, no difference in the recombination rate between the two regions was detected (Green, 1959; Judd, 1964). Since polymorphisms (site and insertion/deletion) in the 5′ flanking region are on the average more closely linked, it is unlikely that the clustering of linkage disequilibrium in the transcriptional unit is due to a relative reduction in the rate of recombination between polymorphisms.

In this report it was shown that geographic differentiation in allele frequency of polymorphisms

https://doi.org/10.1017/S0016672300031694 Published online by Cambridge University Press

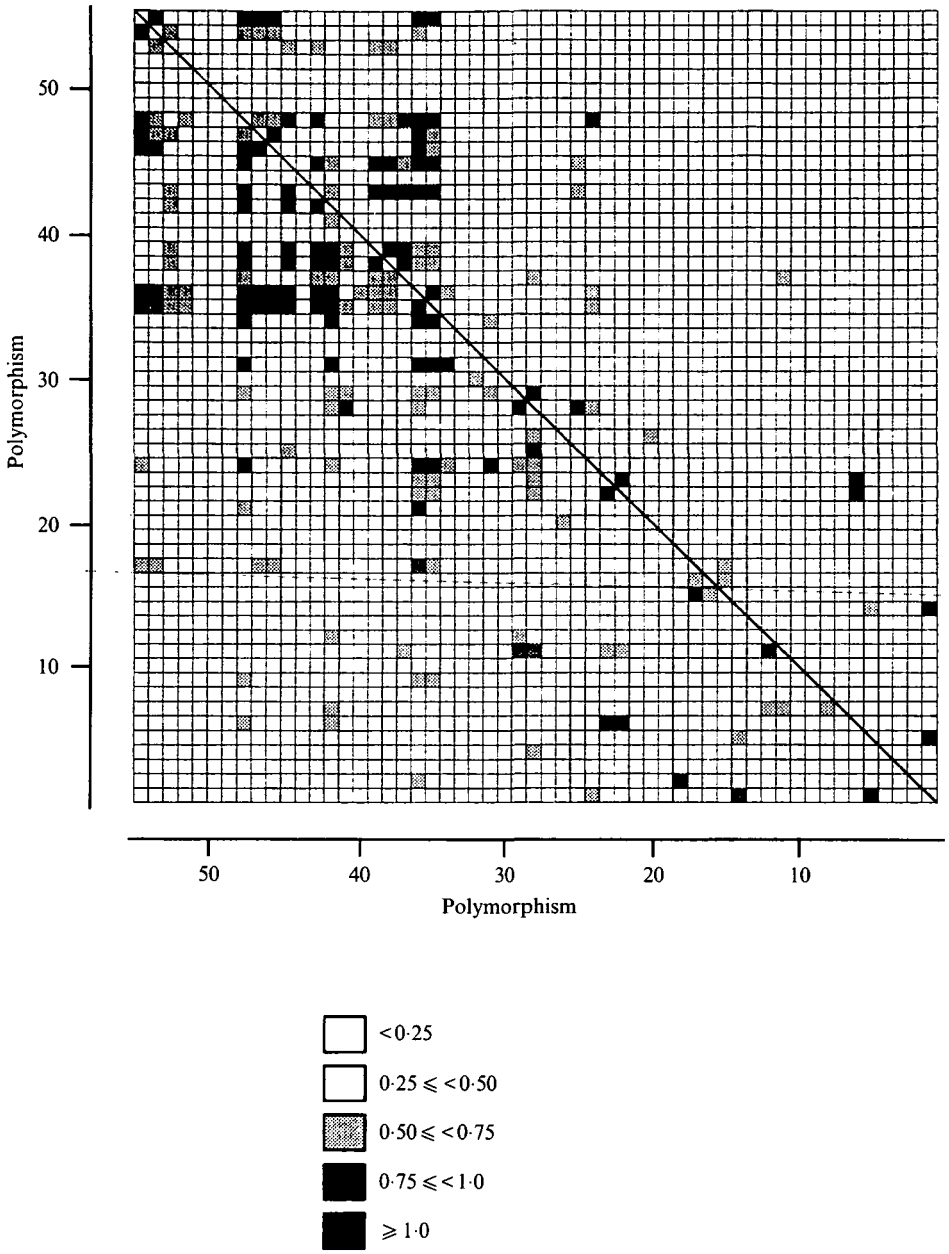

Fig. 4. Ratio of variance components of linkage disequilibrium following Ohta (1982*a*, *b*). Above the diagonal is Dis^2/Dst^2, and below Dst'^2/Dis'^2.

does not account for the different level of significant linkage disequilibrium between the two regions. If the clustering of linkage disequilibria was an artifact of geographic differentiation in the allele frequencies, one might expect that polymorphisms in the transcriptional unit would exhibit more variation in allele frequencies across populations than those in the 5′ flanking region. Contrary to this expectation, the 5′ flanking region is more variable than the transcriptional unit. These results suggest that the clustering of significant linkage disequilibria in the transcriptional unit can not be explained by differences between the two regions in recombination rates or in the frequencies of the polymorphisms.

Heterogeneity of linkage disequilibrium over different populations was investigated in order to characterize the detected linkage disequilibrium further. Significant heterogeneity was detected which suggests that the variation in linkage disequilibrium is not a statistical artifact. Correspondence between linkage disequilibrium and its heterogeneity indicates that when linkage disequilibrium is significant, the degree and direction of linkage disequilibrium tend to be similar among different populations. About two-thirds of pairs of polymorphisms showing significant mean linkage disequilibrium display no significant heterogeneity among populations in linkage disequilibrium. This tendency was observed in both the flanking region and transcriptional unit, and even between those two regions.

It was shown that relatively large values of Ohta's *D* ratio (1982*a*, *b*) are detected in the transcriptional

https://doi.org/10.1017/S0016672300031694 Published online by Cambridge University Press

unit of the *white* locus, and that this clustering corresponds to that of significant linkage disequilibria. The number of pairs with ratio larger than one are very low. Most of them are almost exclusively found in the transcriptional unit, and show significant linkage disequilibrium. This result suggests that significant linkage disequilibrium detected in the transcriptional unit is caused by epistatic natural selection.

Estimated Fst values between populations are not large (Fig. 3) and suggest migration between populations or recent spread of *Drosophila melanogaster* over the world, which would also result in the similarity of linkage disequilibrium between populations. If this is the case, linkage disequilibrium in the transcriptional unit of this locus must have existed in the ancestral population of *D. melanogaster*. It is, of course, impossible to investigate the mechanisms for the ancient linkage disequilibrium from the present data.

Although it is not possible to infer the molecular nature of epistasis, all the analyses conducted in this study suggest the existence of epistatic natural selection for linkage disequilibrium in the transcriptional unit of the *white* locus.

The authors express their thanks to K. Harada, M. Iizuka, F. Tajima and M. Watada for fly collection, to J. M. Comerón for sharing flies and to W. G. Hill, D. Kirby, M. Slatkin, W. Stephan, F. Tajima, R. Terauchi and T. Tokunaga for comments and computer programs. Contribution number 533 from the Laboratory of Genetics, Faculty of Agriculture, Kyoto University, Japan.

References

Aguadé, M., Miyashita, N. & Langley, C. H. (1989). Reduced variation in the *yellow–achaete–scute* region in natural populations of *Drosophila melanogaster*. *Genetics* **122**, 607–615.

Aquadro, C. F., Deese, S. F., Bland, M. M., Langley, C. H. & Laurie-Ahlberg, C. C. (1986). Molecular population genetics of alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics* **114**, 1165–1190.

Begun, D. J. & Aquadro, C. H. (1992). Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitch-hiking of *yellow–achaete* region. *Genetics* **129**, 1147–1158.

Charlesworth, B. & Charlesworth, D. (1973). A study of linkage disequilibrium in populations of *Drosophila melanogaster*. *Genetics* **73**, 351–359.

Cockerham, C. C. (1969). Variance of gene frequencies. *Evolution* **23**, 72–84.

Cockerham, C. C. (1973). Analysis of gene frequencies. *Genetics* **74**, 679–700.

Dobzhansky, Th. (1970). *Genetics of the Evolutionary Process*. New York: Columbia University Press.

Eanes, W. F., Ajioka, J. W., Hey, J. & Wealey, C. (1989). Restriction-map variation associated with the G6PD polymorphism in natural populations of *Drosophila melanogaster*. *Molecular Biology and Evolution* **6**, 384–397.

Ewens, W. J., Spielman, R. S. & Harris, H. (1981). Estimation of genetic variation at the DNA level from restriction endonuclease data. *Proceedings of the National Academy of Sciences, USA* **78**, 3748–3750.

Green, M. M. (1959). Spatial and functional properties of pseudoalleles at the *white* locus in *Drosophila melanogaster*. *Heredity* **13**, 302–315.

Hill, W. G. & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics (Der Zuchter)* **38**, 226–231.

Hudson, R. R. (1982). Estimating genetic variability with restriction endonucleases. *Genetics* **100**, 711–719.

Judd, B. H. (1964). The structure of intralocus duplication and deficiency chromosomes produced by recombination in *Drosophila melanogaster*, with evidence for polarized pairing. *Genetics* **49**, 253–265.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989). The 'hitchhiking' effect revisited. *Genetics* **123**, 887–899.

Kimura, M. (1956). A model of a genetic system which leads to closer linkage by natural selection. *Evolution* **10**, 278–287.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. London: Cambridge University Press.

Kimura, M. & Ohta, T. (1971) *Theoretical Aspects of Population Genetics*. Princeton, New Jersey: Princeton University Press.

Kojima, K., Gillespie, J. & Tobari, Y. N. (1970). A profile of *Drosophila melanogaster* species' enzymes assayed by electrophoresis. I. Number of alleles, heterozygosities, and linkage disequilibrium in glucose-metabolizing systems and some other enzymes. *Biochemical Genetics* **4**, 627–637.

Kreitman, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412–417.

Kreitman, M. & Aguadé, M. (1986). Genetic uniformity in two populations of *Drosophila melanogaster* revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. *Proceedings of the National Academy of Sciences, USA* **83**, 3562–3566.

Langley, C. H. & Aquadro, C. F. (1987). Restriction map variation in natural populations of *Drosophila melanogaster*: *white* locus region. *Molecular Biology and Evolution* **4**, 651–663.

Langley, C. H., Montgomery, E. A. & Quattlebaum, W. F. (1982). Restriction map variation in the *Adh* region of *Drosophila*. *Proceedings of the National Academy of Sciences, USA* **79**, 5631–5635.

Langley, C. H., Shrimpton, A. E., Yamazaki, T., Miyashita, N., Matsuo, Y. & Aquadro, C. F. (1988). Naturally occurring variation in the restriction map of the *Amy* region of *Drosophila melanogaster*. *Genetics* **119**, 619–629.

Langley, C. H., Tobari, Y. N. & Kojima, K. (1974). Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* **78**, 921–936.

Laurie, C. C., Bridgham, J. T. & Choudhary, M. (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-Ahlberg, C. C. & Stam, L. F. (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.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Lewontin, R. C. & Kojima, K. (1960). The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472.

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of favorable gene. *Genetical Research* **23**, 23–35.

Macpherson, J. N., Weir, B. S. & Leigh Brown, A. J. (1990). Extensive linkage disequilibrium in *achaete–scute* complex of *Drosophila melanogaster*. *Genetics* **126**, 121–129.

Martín-Campos, J. M., Comerón, J. M., Miyashita, N. & Aguadé, M. (1992). Intra- and interspecific variation at

the *y–ac–sc* region of *Drosophila simulans* and *melanogaster*. *Genetics* **130**, 805–816.

Miyashita, N. T. (1990). Molecular and phenotypic variation of the *Zw* locus region in *Drosophila melanogaster*. *Genetics* **125**, 407–419.

Miyashita, N. & Langley, C. H. (1988). Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. *Genetics* **120**, 199–212.

Mukai, T., Mettler, L. E. & Chigusa, S. (1971). Linkage disequilibrium in a local population of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA* **68**, 1065–1069.

Mukai, T., Watanabe, T. K. & Yamaguchi, O. (1974). The genetic structure of natural populations of *Drosophila melanogaster*. XII. Linkage disequilibrium in a large local population. *Genetics* **77**, 771–793.

Mukai, T. & Voelker, R. A. (1977). The genetic structure of natural populations of *Drosophila melanogaster*. XIII. Further studies on linkage disequilibrium. *Genetics* **86**, 175–185.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA* **70**, 3321–3323.

Nei, M. & Li, W-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA* **76**, 5269–5273.

Nei, M. & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145–163.

Ohta, T. (1982*a*). Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceedings of the National Academy of Sciences, USA* **79**, 1940–1944.

Ohta, T. (1982*b*). Linkage disequilibrium with the island nodel. *Genetics* **101**, 139–155.

Ohta, T. & Kimura, M. (1969). Linkage disequilibrium due to random genetic drift. *Genetical Research* **13**, 47–55.

Prakash, S. & Lewontin, R. C. (1968). A molecular approach to the study of genic heterozygosity. III. Direct evidence of coadaptation in gene arrangement of *Drosophila*. *Proceedings of the National Academy of Sciences, USA* **59**, 398–405.

Prevosti, A., García, M. P., Serra, L., Aguadé, M., Ribó, G. & Sagarra, E. (1983). Association between allelic allozyme alleles and chromosomal arrangements in European populations and Chilean colonizers of *Drosophila subobscura*. *Isozymes* **10**, 171–191.

Riley, M. A., Hallas, M. E. & Lewontin, R. C. (1989). Distinguishing the forces controlling genetic variation at the *Xdh* locus in *Drosophila pseudoobscura*. *Genetics* **123**, 359–369.

Schaeffer, S. W., Aquadro, C. F. & Langley, C. H. (1988). Restriction-map variation in the *Notch* region of *Drosophila melanogaster*. *Molecular Biology and Evolution* **5**, 30–40.

Steele, R. G. D. & Torrie, J. H. (1980). *Principles and Procedures of Statistics. A Biometrical Approach*. New York: McGraw-Hill.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Voelker, R. A., Cockerham, C. C., Johnson, F. M., Schaffer, H. E., Mukai, T. & Mettler, L. E. (1978). Inversions fail to account for allozyme clines. *Genetics* **88**, 515–527.

Weir, B. (1990). *Genetic Data Analysis*. Sunderland, MA, USA: Sinauer Associates.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics* **15**, 323–354.

Yamaguchi, O., Ichinose, M., Matsuda, M. & Mukai, T. (1980). Linkage disequilibrium in isolated populations of *Drosophila melanogaster*. *Genetics* **96**, 507–522.

Zouros, E. & Krimbas, C. B. (1973). Evidence of linkage disequilibrium maintained by selection in two natural populations of *Drosophila subobscura*. *Genetics* **73**, 659–674.