Distribution of Nonrandom Associations Between Pairs of Protein Loci Along the Third Chromosome of *Drosophila melanogaster*

Carlos Zapata,[1] Concepción Núñez and Teresa Velasco

Departamento de Genética, Universidad de Santiago, 15782 Santiago de Compostela, Spain

Manuscript received March 6, 2002
Accepted for publication May 9, 2002

ABSTRACT

The within-chromosome distribution of gametic disequilibrium (GD) between protein loci, and the underlying evolutionary factors of this distribution, are still largely unknown. Here, we report a detailed study of GD between a large number of protein loci (15) spanning 87% of the total length of the third chromosome of *Drosophila melanogaster* in a large sample of haplotypes (600) drawn from a single natural population. We used a sign-based GD estimation method recently developed for multiallelic systems, which considerably increases both the statistical power and the accuracy of estimation of the intensity of GD. We found that strong GD between pairs of protein loci was widespread throughout the chromosome. In total, 22% of both the pairs of alleles and pairs of loci were in significant GD, with mean intensities (as measured by D' coefficients) of 0.43 and 0.31, respectively. In addition, strong GD often occurs between loci that are far apart. By way of illustration, 32% of the allele pairs in significant GD occurred within pairs of loci separated by effective frequencies of recombination (EFRs) of 15–20 cM, the mean D' value being 0.49. These observations are in sharp contrast with previous studies showing that GD between protein loci is rarely found in natural populations of outcrossing species, even between very closely linked loci. Interestingly, we found that most instances of significant interallelic GD (68%) involved functionally related protein loci. Specifically, GD was markedly more frequent between protein loci related by the functions of hormonal control, molybdenum control, antioxidant defense system, and reproduction than between loci without known functional relationship, which is indicative of epistatic selection. Furthermore, long-distance GD between functionally related loci (mean EFR 9 cM) suggests that epistatic interactions must be very strong along the chromosome. This evidence is hardly compatible with the neutral theory and has far-reaching implications for understanding the multilocus architecture of the functional genome. Our findings also suggest that GD may be a useful tool for discovering networks of functionally interacting proteins.

KNOWLEDGE about gametic disequilibrium (GD), the nonrandom association of alleles at different loci, remains limited for loci forming part of the functional genome. Nevertheless, knowledge on the amount and distribution of GD among protein loci along chromosomes and of the underlying evolutionary factors is fundamental to unraveling the multilocus architecture of the functional genome and its evolutionary dynamics. The theory of multilocus genetic systems suggests that epistatic (*i.e.*, nonadditive) interactions in fitness are able to generate GD if linkage among loci is tighter than the value required by the magnitude of the epistasis. If this were so, the study of individual loci would have insufficient dimensionality to explain the empirical observations, because multilocus systems have additional properties over and above those of individual loci (Franklin and Lewontin 1970; Slatkin 1972; Lewontin 1974). Undoubtedly, more information is needed to ascertain the most appropriate levels of analysis. In this connection, many basic aspects of the functional multilocus architecture of chromosomes are still poorly understood, including (1) frequency and intensity of GD, (2) distribution of GD along chromosomes and its relationship with recombination frequency, and (3) the amount of GD among protein loci subject to selection.

Extensive empirical work has been carried out in the past with the aim of assessing the amount of GD between protein loci in natural populations, particularly in Drosophila species. It has been found that there is very little GD in random-mating populations, even between closely linked protein loci (Hedrick *et al.* 1978; Barker 1979; Lewontin 1985; Hartl and Clark 1997; Mitton 1997; Powell 1997; Hedrick 2000). This apparent scarcity of GD suggested that multilocus genetics was of minor importance for understanding the evolution of protein variation (Lewontin 1985). Nevertheless, Zapata and Alvarez (1992) demonstrated that the failure to detect abundant GD between protein loci was in great measure due to the exceedingly low statistical powers resulting from small sample sizes. A metaanalysis of a number of individual studies of GD in Drosophila species showed that, despite the lack of significance in most individual studies, GD is in fact relatively frequent even between loosely linked protein loci. Specifically, 28% of the pairs

[1]*Corresponding author:* Departamento de Genética, Facultad de Biología, Universidad de Santiago, 15782 Santiago de Compostela, Spain. E-mail: bfcazaba@usc.es

of loci analyzed were in significant GD, and the average effective frequency of recombination was 9 cM (Zapata and Alvarez 1992).

The common practice of using only a few (two to six) protein loci per chromosome (see Zapata and Alvarez 1992) does not provide a systematic picture of the within-chromosome distribution of GD. It is certainly paradoxical that the within-chromosome distribution of GD for first-generation molecular genetic markers or protein loci is to date much less well known than for last-generation markers. The past few years have witnessed an extraordinary surge of information concerning the distribution of GD between DNA markers over extensive regions of human chromosomes. This research has been based on the analysis of single-nucleotide polymorphisms (SNPs; Taillon-Miller *et al.* 2000; Goldstein 2001) and microsatellites (Peterson *et al.* 1995; Zapata *et al.* 2001) and has been fueled by the usefulness of GD for locating genes that underlie susceptibility to common diseases. Unfortunately, estimates of GD from DNA markers such as SNPs and microsatellites cannot be used as reliable indicators of background GD over chromosomes, because of the occurrence of marker-dependent GD (Zapata *et al.* 2001). Each type of genetic marker is endowed with particular characteristics that influence its evolutionary dynamics, and thus reliable information on GD among protein loci can be obtained only from the study of the loci themselves.

A procedure that has been used almost routinely in GD studies is the pooling of rarer alleles at each locus into a single class to avoid the statistical problem of small expected frequencies, thereby reducing multiallelic loci to diallelic systems. Such a procedure may be criticized for several reasons. First, a considerable amount of information of value for elucidating the evolutionary forces that generate GD could eventually be lost by the pooling of alleles (Hedrick and Thomson 1986; Klitz and Thomson 1987; Zapata *et al.* 2001). Second, GD may be underestimated because pooling of alleles often reduces the statistical power of the statistical tests used (Weir and Cockerham 1978; Sham and Curtis 1995). Third, the pooling of alleles prevents proper quantification of the amount of GD and reliable detection of the possible occurrence of allele-dependent GD patterns (Zapata *et al.* 2001). Last, the manner in which the alleles are pooled can have important and unknown effects on the inferences drawn, because the allelic combinations probably have different GD levels. However, the theory of estimation of GD for multiallelic systems is nowadays sufficiently developed to make it both unnecessary and undesirable to pool alleles. In addition, the problem of false positives associated with small expectations can be evaluated by comparing the results from conservative and nonconservative statistical tests, taking into account that statistical tests should be used only as tools for the exploration of the hypotheses that need to be substantiated on the basis of other lines of evidence.

A long-standing question at the interpretative level is what evolutionary forces are responsible for the observed GD. This is a very difficult question because, besides natural selection, a number of other forces can generate GD in populations, such as genetic drift or bottlenecks (Hill and Robertson 1968; Avery and Hill 1979), stratification or admixture (Nei and Li 1973; Chakraborty and Weiss 1988), and recent mutations (Zapata *et al.* 2001); thus alternative explanations cannot usually be ruled out. The idea that epistatic interactions are more likely to occur between loci coding for proteins having a functional relationship provides one of the best ways to relate GD to selection. However, the experimental evidence on the occurrence of GD between functionally related protein loci is to date scarce (Mitton and Koehn 1973; Zouros and Krimbas 1973; Zouros and Johnson 1976; Fontdevila *et al.* 1983; Zapata *et al.* 2000) and sometimes contradictory (Zouros and Krimbas 1973; Loukas *et al.* 1979).

Unraveling the forces generating GD among protein loci has acquired new relevance with the emerging field widely referred to as "proteomics." As is well known, a major challenge over the coming years will be to interpret the functional significance of the DNA sequences that are currently being revealed by genome research. The goal of proteomics is not only to catalog the proteins of a given cell, but also to identify their functions and interactions, and thus to understand the biology of the cell (Dove 1999; Williams 1999; Eisenberg *et al.* 2000; Pandey and Mann 2000). Various methodological approaches have been used to identify the functions of unknown proteins and their interactions, including their similarity in structure, sequence, or activity to previously characterized proteins; their phylogenetic profiles; and yeast two-hybrid systems (Eisenberg *et al.* 2000; Pandey and Mann 2000). In addition, we suggest that GD analysis may be a useful approach for detecting networks of functionally interacting proteins, if it is demonstrated that GD is more likely to occur among loci coding for functionally related proteins.

In this article we report the amount (frequency and intensity) and distribution of overall and interallelic GD between pairs of 15 protein loci distributed along the third chromosome of *Drosophila melanogaster*. We used a considerably larger sample size than in previous studies and a sign-based GD estimation method for multiallelic loci, thus improving statistical power. Analysis of existing information on the functional relationships among the protein loci studied provided a basis for identification of evolutionary factors underlying the distribution of GD along the chromosome and for assessment of the potential usefulness of GD analysis in proteomics.

MATERIALS AND METHODS

Sampling and extraction of the third chromosomes: Adult flies of *D. melanogaster* were collected over a 3-day period in October 1998 by net-sweeping over conventional fermented banana baits placed in a large extension of fruit trees in Santa

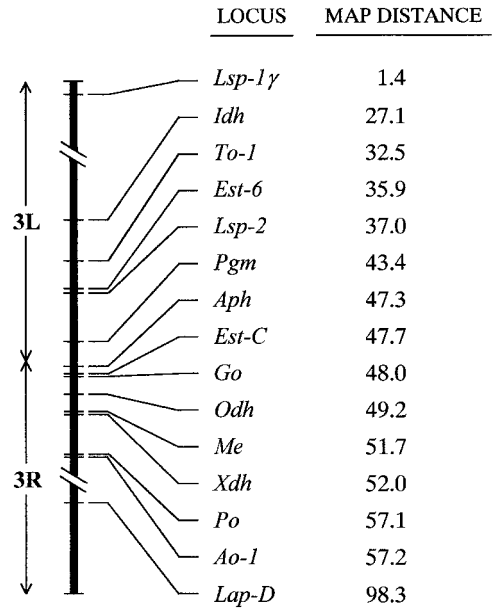

FIGURE 1.—Location of the 15 protein loci studied on the left (3L) and right (3R) arms of the third chromosome of *D. melanogaster*. Map distance (in centimorgans) is shown for each locus.

Cruz de Rivadulla (A Coruña, northwest Spain). Homozygous lines for independent third chromosomes were established from males by standard crosses to TM6B/MKRS balancer stock (LINDSLEY and ZIMM 1992). Each wild male was mated to several virgin TM6B/MKRS females. A single F_1 male of each cross, genotype III^+/TM6B, was backcrossed to the balancer stock. Males and females from these matings, genotypes III^+/TM6B, were mated to initiate the homozygous lines (III^+/III^+), which were maintained in mass cultures. Less viable or lethal wild third chromosomes were maintained balanced (no recombination) with TM6B. A total of 605 lines were eventually obtained and characterized by karyotype and allozyme analyses.

A sample of adults was collected in September 1998 for analyzing the genetic structure of the natural population of Santa Cruz de Rivadulla.

Selection, denomination, recombination frequencies, and functional relationships of the protein loci studied: To characterize GD we considered all previously described protein loci on the third chromosome, with only two basic requirements. First, the map positions of the loci were known; second, the protein loci were polymorphic according to previous screenings of variability in the natural population of Santa Cruz de Rivadulla. Fifteen protein loci distributed along the left and right arms of the third chromosome eventually satisfied those two requirements (Figure 1). The following 15 protein loci in the third chromosome, ordered from left-arm telomere to right-arm telomere, were studied: larval serum protein 1 (*Lsp-1γ*), isocitrate dehydrogenase (E.C. 1.1.1.42; *Idh*), tetrazolium oxidase-1 (E.C. 1.15.1.1; *To-1*), esterase-6 (E.C. 3.1.1.1; *Est-6*), larval serum protein 2 (*Lsp-2*), phosphoglucomutase (E.C. 5.4.2.2; *Pgm*), alkaline phosphatase (E.C. 3.1.3.1; *Aph*), esterase-C (E.C. 3.1.1.1; *Est-C*), glucose oxidase (E.C. 1.1.99.10; *Go*), octanol dehydrogenase (E.C. 1.2.1.1; *Odh*), malic enzyme (E.C. 1.1.1.40; *Me*), xanthine dehydrogenase (E.C. 1.1.1.204; *Xdh*), pyridoxal oxidase (E.C. 1.2.3.8; *Po*), aldehyde oxidase-1 (E.C. 1.2.1.3; *Ao-1*), and leucine aminopeptidase-D (E.C. 3.4.11.1; *Lap-D*) (http://flybase.bio.indiana.edu).

The 15 loci studied span ~87% (96.9 cM, from 1.4 to 98.3 cM) of the total length (110.9 cM) of the third chromosome (LINDSLEY and ZIMM 1992). Effective frequencies of recombination (EFRs) between all possible pairs of the 15 loci were obtained from map distances (http://flybase.bio.indiana.edu), using Kosambi's map function (KOSAMBI 1944) and assuming no recombination in males of Drosophila (Table 1). The resulting values ranged from 0.1 cM (*Est-C* × *Go*, *Me* × *Xdh*, and *Po* × *Ao-1*) to 24 cM (*Lsp-1γ* × *Lap-D*) and averaged 8.7 ± 0.7 cM.

An exhaustive analysis of the available information about the functional relationships of the loci studied (http://flybase.bio.indiana.edu) led us to group them under five major functional categories: hormonal control (HEWITT 1974; O'BRIEN and MACINTYRE 1978; COX-FOSTER *et al.* 1990; KAROTAM and OAKESHOTT 1993; ANTONIEWSKI *et al.* 1995; FARKAS and KNOPP 1998; BURMESTER *et al.* 1999), molybdenum control (HANLY 1980; WARNER and FINNERTY 1981), antioxidant defense system (HILLIKER *et al.* 1992; HUMPHREYS *et al.* 1996), glucose metabolism (CAVENER 1980; METZLER 2001), and reproduction (CAVENER and MACINTYRE 1983; PHILLIPS *et al.* 1989; RICHMOND *et al.* 1990; MASSEY *et al.* 1997). The distribution of the protein loci according to these categories is shown in Table 2. It can be seen that all the loci studied, except *Aph*, *Est-C*, and *Lap-D*, are related by at least one of these functional categories.

Cytological and electrophoretic analyses: Lines were screened for inversions by crossing males derived from each line with virgin Oregon-R (Or-R) females homozygous for the standard gene arrangement (LINDSLEY and GRELL 1968). Larval salivary glands from the F_1 progeny were dissected and squashes were examined after staining the chromosomes with lactic acetic orcein. Break points of the inversions were determined by comparison with standard maps (LEFEVRE 1976).

The 15 protein loci in the third chromosome were studied using starch gel electrophoresis except *Lsp-1γ* and *Lsp-2*, which were analyzed by acrylamide gel electrophoresis. Electrophoresis and histochemical staining methods were essentially as described previously (AYALA *et al.* 1972; LOUKAS *et al.* 1974; LOUKAS and KRIMBAS 1979, 1980; CAVENER 1980; CYPHER *et al.* 1982). Chromosomes in balance with TM6B were assayed for the enzymes described above, using a TM6B/MKRS balancer stock derived by us from a single TM6B chromosome with known allelic constitution at the 15 protein loci.

Genotypes at nine protein loci active in the adult stage (*To-1*, *Est-6*, *Pgm*, *Est-C*, *Go*, *Odh*, *Me*, *Po*, and *Ao-1*) were also determined in the sample of adults collected in September 1998.

Measures of variability and deviations from Hardy-Weinberg proportions: Unbiased expected heterozygosities at single loci were calculated as $h = N(1 - \Sigma_{i=1}^{k} p_i^2)/N - 1$, where p_i is the relative frequency of the allele i in a sample of N haplotypes (NEI 1987). The genetic structure of the natural population was examined by analysis of deviations of observed genotype frequencies from those expected under Hardy-Weinberg. Estimates of departures from Hardy-Weinberg proportions at each locus, in terms of homozygote excess, along with their standard errors, were obtained with the *f*-statistic. Tests of significance for $f = 0$ were performed using the square of the ratio of the estimate to its standard error for $f = 0$, which approximates a chi-square (χ^2) distribution with 1 d.f. (CURIE-COHEN 1982; ROBERTSON and HILL 1984).

Estimation of gametic disequilibrium: Inversion-carrying chromosomes and singletons were excluded from the analysis of GD, and so the number of haplotypes across pairs of loci ranged from 557 to 600. We used a sign-based GD estimation method recently proposed for multiallelic systems (ZAPATA *et al.* 2001). Briefly, let p_i and q_j be the relative frequencies of alleles A_i ($i = 1, \ldots, k$) and B_j ($j = 1, \ldots, l$) at the loci A and B, respectively, and X_{ij} be the relative frequency of the haplotype $A_i B_j$ in N haplotypes sampled from a population. The array of possible two-locus haplotypes can be ordered in a $k \times l$ contingency table and subsequently partitioned into

TABLE 1

Effective frequencies of recombination (in centimorgans) between the pairs of protein loci studied

	Lsp-1γ	*Idh*	*To-1*	*Est-6*	*Lsp-2*	*Pgm*	*Aph*	*Est-C*	*Go*	*Odh*	*Me*	*Xdh*	*Po*	*Ao-1*	*Lap-D*
Lsp-1γ	—														
Idh	11.8	—													
To-1	13.8	2.7	—												
Est-6	14.9	4.3	1.7	—											
Lsp-2	15.3	4.9	2.2	0.5	—										
Pgm	17.1	7.9	5.4	3.7	3.2	—									
Aph	18.1	9.6	7.2	5.6	5.1	1.9	—								
Est-C	18.2	9.7	7.4	5.8	5.3	2.1	0.2	—							
Go	18.3	9.9	7.5	5.9	5.4	2.3	0.3	0.1	—						
Odh	18.6	10.4	8.0	6.5	6.0	2.9	0.9	0.7	0.6	—					
Me	19.1	11.4	9.2	7.7	7.2	4.1	2.2	2.0	1.9	1.3	—				
Xdh	19.2	11.5	9.3	7.8	7.3	4.3	2.3	2.1	2.0	1.4	0.1	—			
Po	20.1	13.4	11.4	10.0	9.5	6.7	4.8	4.6	4.5	3.9	2.7	2.5	—		
Ao-1	20.1	13.5	11.4	10.0	9.6	6.7	4.9	4.7	4.5	4.0	2.7	2.6	0.1	—	
Lap-D	24.0	22.3	21.6	21.2	21.0	20.0	19.2	19.2	19.1	18.8	18.3	18.2	16.9	16.9	—

$k \times l$ separate 2×2 contingency tables, the frequency in each cell being X_{ij}, $p_i - X_{ij}$, $q_j - X_{ij}$, and $1 - p_i - q_j + X_{ij}$. We then considered the haplotype classes in coupling to be those involving the most or the least frequent alleles, and only instances of interallelic GD with a positive sign were included in the analyses. This strategy increases both statistical power and the accuracy of estimation of the intensity of GD between pairs of alleles.

The intensity of GD between pairs of alleles at the two loci was measured by $D'_{ij}(+) = D_{ij}(+)/D_{\max}$, where $D_{ij}(+) = X_{ij} - p_i q_j(+)$ and $D_{\max} = \min[p_i(1 - q_j), (1 - p_i)q_j]$, and $p_i q_j(+)$ are the expected frequencies of the haplotypes with positive $[D_{ij}(+)]$ deviations from random association. The $D'_{ij}(+)$ coefficient potentially ranges from 0 to 1.0 (Zapata 2000). The null hypothesis of gametic equilibrium for each pair of alleles ($D_{ij} = 0$) *vs.* the alternative hypothesis ($D_{ij} > 0$) was tested by the one-sided $\chi^2_{ij}(+) = ND^2_{ij}(+)/p_i(1 - p_i)q_j(1 - q_j)$ test, which approximates a χ^2 distribution with 1 d.f. The intensity of positive overall GD was measured as $D'(+) = \Sigma_i\Sigma_j p_i q_j(+) D'_{ij}(+)/\Sigma_i\Sigma_j p_i q_j(+)$. The range of $D'(+)$ varies from 0 to a maximum value close to or equal to 1.0 (Zapata 2000; Zapata *et al.* 2001). The null hypothesis of overall gametic equilibrium was tested by individual one-sided $\chi^2_{ij}(+)$ tests with Bonferroni correction. Estimation of sign-based GD for the simple case of pairs of loci with two alleles each was carried out as described by Zapata and Alvarez (1993).

The relationship between the magnitude of GD and the EFR between locus pairs was investigated by calculation of the Pearson product-moment correlation coefficient (r) and Kendall's nonparametric coefficient of rank correlation (τ;

TABLE 2

Functional relationships among the 15 protein loci studied

Locus	Hormonal control	Molybdenum control	Antioxidant defense system	Glucose metabolism	Reproduction
Lsp-1γ	+	−	−	−	+
Idh	−	−	−	+	−
To-1	−	−	+	−	+
Est-6	+	−	−	−	+
Lsp-2	+	−	−	−	+
Pgm	−	−	−	+	−
Aph	−	−	−	−	−
Est-C	−	−	−	−	−
Go	+	−	−	+	+
Odh	+	−	−	−	−
Me	+	−	−	−	−
Xdh	−	+	+	−	−
Po	−	+	+	−	−
Ao-1	+	+	+	−	−
Lap-D	−	−	−	−	−

The signs plus (+) and minus (−) indicate whether the loci are related or not related by the function specified.

TABLE 3

Deviations from Hardy-Weinberg proportions

Locus	*N*	Observed heterozygosity	Expected heterozygosity	$f \pm$ SE	*P* value
To-1	113	0.292	0.341	0.145 ± 0.103	0.123
Est-6	206	0.257	0.273	0.070 ± 0.066	0.153
Pgm	400	0.332	0.350	0.038 ± 0.061	0.185
Est-C	413	0.109	0.115	0.018 ± 0.052	0.598
Go	263	0.232	0.234	0.009 ± 0.063	0.888
Odh	181	0.011	0.011	−0.000 ± 0.003	1.000
Me	272	0.000	0.000	—	—
Po	259	0.069	0.067	−0.034 ± 0.008	0.583
Ao-1	147	0.272	0.291	0.015 ± 0.072	0.752
Mean ± SE	250 ± 35	0.175 ± 0.043	0.187 ± 0.047	0.033 ± 0.019	

Unbiased expected heterozygosities were calculated as $h = 2N(1 - \Sigma_{i=1}^{k} p_i^2)/2N - 1$, where p_i is the frequency of the allele i in a sample of N genotypes drawn from the population (Nei 1987).

Sokal and Rohlf 1995), with assessment of significance by one-sided Mantel's matrix-comparison test (Mantel 1967).

RESULTS

Variability and deviations from Hardy-Weinberg proportions: The 15 protein loci studied were slightly to highly polymorphic. A total of 49 alleles were detected in the sample of haplotypes, and 22 (45%) of 49 had a frequency ≤3%. Rare alleles (frequency ≤3%) were distributed across all loci with the exception of *To-1*, *Go*, and *Lap-D*. The number of alleles varied between 2 (*To-1*, *Lsp-2*, *Go*, and *Me*) and 6 (*Xdh*), with the mean (±SE) being 3.3 ± 0.3. Unbiased estimates of the expected allelic diversity ranged from 0.01 (*Me*) to 0.66 (*Xdh*). The 15 markers had an average estimated heterozygosity of 0.24 ± 0.06. Such levels of variability are in agreement with those expected for these loci in west European populations of *D. melanogaster* (Girard *et al.* 1977; Cabrera *et al.* 1982; David 1982; Singh *et al.* 1982). There is no evidence of recent founder effects, which could induce GD in the natural population of Santa Cruz de Rivadulla.

We found no consistent evidence of deviations from Hardy-Weinberg proportions in the natural population of Santa Cruz de Rivadulla (*i.e.*, in the sample of genotypes collected in September 1998; Table 3). Except at the *To-1* locus, the magnitude of the deviation was in all cases very small and not statistically significant (Table 3). In six of the eight loci, the deviation was positive (*i.e.*, fewer observed homozygotes than expected), but this apparent trend was likewise not statistically significant (sign-test, $P > 0.05$). The natural population of Santa Cruz de Rivadulla thus seems to behave as a panmictic unit with no indication of population subheterogeneity (which might cause GD). Evidence of fit to Hardy-Weinberg proportions is also a quality control of genotyping (Gomes *et al.* 1999).

Interallelic GD: A total of 800 different two-locus haplotypes were detected across all the possible pairwise comparisons (105) of the 15 loci studied. Only 329 allelic combinations with positive deviations from random association were eventually considered in the analyses. The 329 allelic combinations distributed at 79 pairs of loci. The frequency and the mean intensity of significant interallelic GD within each locus pair are shown in Figure 2. It was found that 72 (22%) of the 329 allelic combinations were in significant GD ($P < 0.05$), with a mean intensity as measured by D'_{ij} (+) of 0.43 ± 0.04. All 15 loci were involved in at least one significant effect but cases of significant interallelic GD distributed across 30 (38%) of the 79 pairs of loci. The percentage and the intensity of significant interallelic GD fluctuated markedly among the pairs of loci studied: specifically, the percentage ranged from a minimum of 13% (*Xdh* × *Ao-1*) to a maximum of 100% (*Lsp-1*γ × *Lsp-2*, *Lsp-1*γ × *Me*, *Lsp-1*γ × *Po*, *Idh* × *Est6*, *To-1* × *Go*, *Est6* × *Me*, *Lsp-2* × *Me*, *Lsp-2* × *Po*, *Pgm* × *Me*, *Pgm* × *Ao-1*, and *Po* × *Ao-1*), while mean D'_{ij} (+) ranged from 0.10 (*To-1* × *Xdh*) to 1.00 (*Lsp-1*γ × *Lsp-2*, *Est-6* × *Me*, and *Odh* × *Xdh*). Note that in only two pairs of loci (*Lsp-1*γ × *Lsp-2* and *Est-6* × *Me*) were all the allele combinations in significant GD and with maximal association [D'_{ij} (+) = 1].

Further evidence discussed below argues against the possibility that most of the observed interallelic GD may have arisen as a consequence of type I errors. In fact, however, it is more likely that our estimate of the frequency of GD is an underestimate because the statistical power of the chi-square test to detect significant deviations from random association is low, even with the large sample sizes used in this study (see Zapata and Alvarez 1992).

Relationship of interallelic GD with recombination: Analysis of the relationship between GD and EFR revealed several interesting observations. First, GD was more frequent between more closely linked locus pairs: the frequency of significant GD was 33% (14/42) for locus pairs separated by EFRs between 0 and 2 cM and 20% (58/287) for locus pairs separated by >2 cM. This

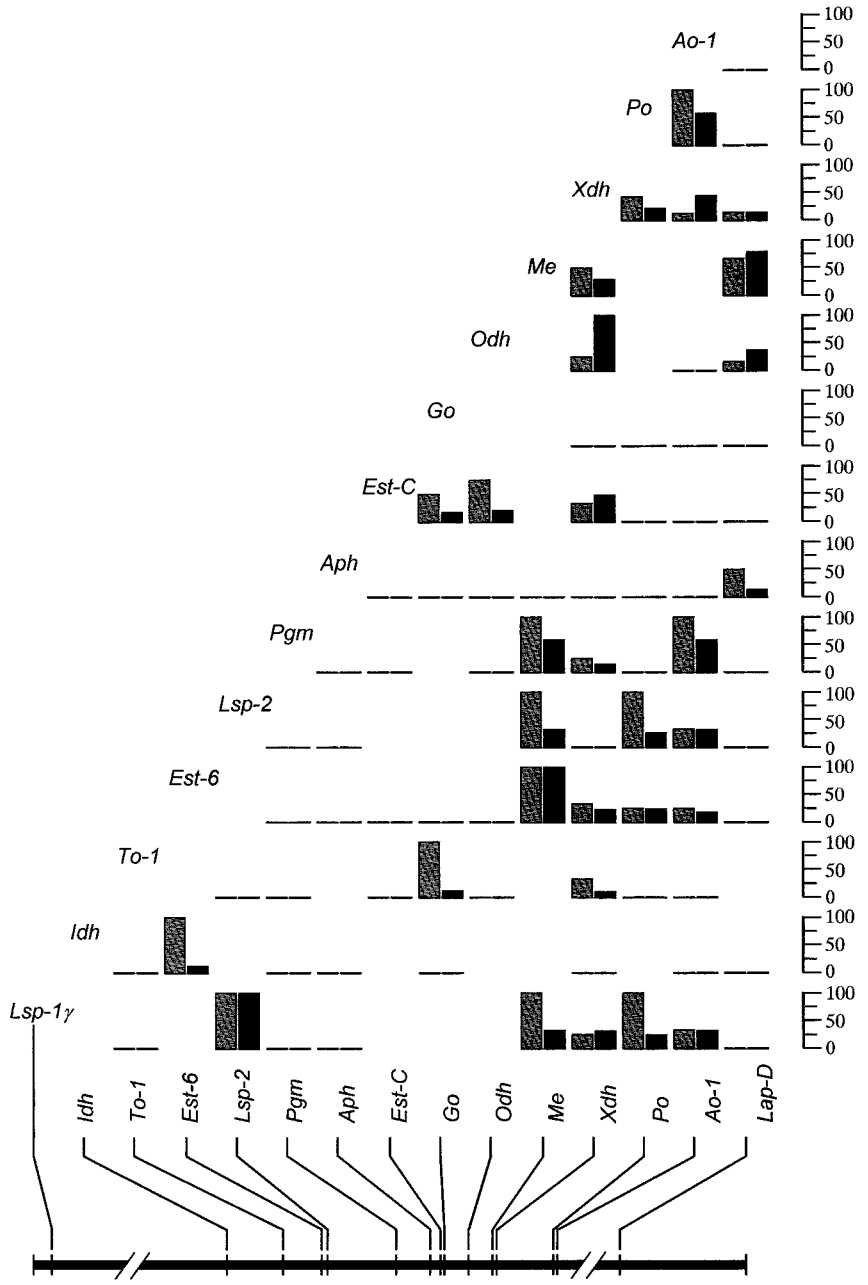

FIGURE 2.—Distribution of interallelic GD between pairs of loci along the third chromosome of *D. melanogaster*. Frequency (hatched bars) and mean intensity (solid bars) of significant interallelic GD within each locus pair are given as percentages. Data for pairs of loci with negative deviations from random association are not shown.

difference was statistically significant based on the one-sided chi-square test for 2 × 2 contingency tables with 1 d.f. ($\chi^2 = 3.69$, $P = 0.03$). However, strong GD often extends over considerable chromosomal distances. The EFR between pairs of alleles in GD ranged from 0.1 cM (*Est-C* × *Go*, *Me* × *Xdh*, and *Po* × *Ao-1*) to 20.1 cM (*Lsp-1*γ × *Po* and *Lsp-1*γ × *Ao-1*), with a mean value of 9.2 ± 0.8 cM. We observed that 32% (23/72) of the significant cases were separated by EFRs within 15–20 cM, while mean D'_{ij} (+) was 0.49 ± 0.07. These EFRs for pairs of loci in GD are remarkably high, since the theoretical maximum value is only 25 cM (assuming no recombination in males of Drosophila). Last, there is no evidence to suggest that GD is a monotonic function of between-locus distance (Figure 3). Indeed, there is no significant decrease in the values of D'_{ij} (+) as EFR increases ($r = 0.021$, $\tau = -0.039$, $P = 0.45$, $n = 72$).

Interallelic GD dependent on the functional relationship among loci: Table 4 shows that GD was about two times more frequent among functionally related locus pairs than among locus pairs without known functional relationships. Specifically, the percentage of significant interallelic GD was 36% (32/90) for related locus pairs and 17% (40/239) for unrelated pairs. This difference was highly significant ($P < 0.001$, one-sided chi-square test for 2 × 2 contingency tables). The intensity values are also consistent with this pattern: considering allele pairs showing significant GD, the mean D'_{ij} (+) value

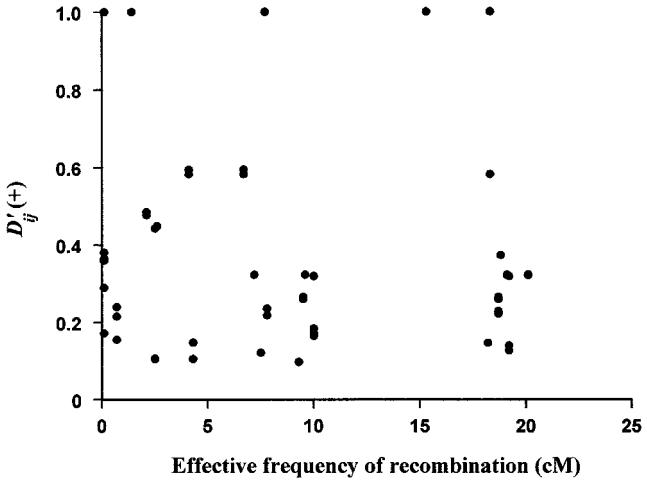

FIGURE 3.—Plot of D'_{ij} (+) values against effective frequency of recombination (EFR) for significant comparisons.

was 0.50 ± 0.06 for functionally related pairs, *vs.* 0.37 ± 0.04 for unrelated pairs. The difference in mean EFR between related and unrelated pairs is too small to account for the substantial difference observed in the frequency of GD (9.1 ± 1.2 cM *vs.* 9.3 ± 1.2 cM).

The above-mentioned general trend was also evident when each particular functional category was considered separately: significant differences in the frequency of GD were detected between unrelated loci and those loci related by the functions of hormonal control ($P < 0.001$), molybdenum control ($P < 0.001$), antioxidant defense system ($P < 0.01$), and reproduction ($P < 0.05$), with the only exception being glucose metabolism ($P = 0.14$). In addition, the mean of D'_{ij} (+) for each particular functional category ranged from 0.40 ± 0.09 (antioxidant defense system) to 0.82 ± 0.18 (reproduction), whereas that of unrelated pairs was 0.37 ± 0.04. Note that 95% bootstrap confidence intervals of the means of D'_{ij} (+) between loci related by hormonal control and reproduction functions do not overlap with that of the mean D'_{ij} (+) between unrelated pairs.

Allele frequency-dependent GD: We next investigated whether GD depends on allele frequency. Following the same arbitrary criterion used in a previous study (ZAPATA *et al.* 2001), we subdivided the sample of two-locus haplotypes (329) into RR, RC, and CC, where R and C indicate rarer (frequency ≤3%) and more common (frequency >3%) alleles. Table 5 shows that GD is significantly more frequent in RR haplotypes (89%) than in RC (22%) or CC (18%) haplotypes ($P \ll 0.001$), with no significant difference between RC and CC ($P = 0.37$). An increase in type I errors when one or more expectations in 2 × 2 contingency tables are small may be a source of spurious allele frequency-dependent patterns. In this case, however, the same allele frequency-dependent pattern was maintained when we considered only interallelic GD shown to be significant by the chi-square test with Yates's correction (Table 5). The extremely conservative behavior of this test for small expectations (*i.e.*, it maintains the null hypothesis too frequently; ZAPATA *et al.* 2001) makes it very unlikely the pattern observed is attributable to type I error.

In addition, the pattern cannot be explained by variations in the amount of recombination, since mean EFR values were even higher for RR haplotypes (10.2 ± 2.7 cM) than for RC (8.4 ± 1.2 cM) and CC (9.9 ± 1.3 cM) haplotypes.

Overall two-locus GD: The relative contributions of allelic combinations in GD within each locus pair as a whole were evaluated by estimating overall two-locus GD (Table 6). Significant overall GD was detected in 17 (22%) of the 79 pairs of loci studied, despite application of the highly conservative Bonferroni correction. All loci studied were involved in at least 1 locus pair in significant overall GD, except *Aph* and *Lap-D*. It is worth noting that *Aph* and *Lap-D* are two loci without apparent functional relationships with any of the other loci studied (see Table 2). The loci involved in the highest rela-

TABLE 4

Interallelic GD dependent on the functional relationships among protein loci

Functional relationships among loci	No. of pairwise comparisons	% of significant comparisons	χ^2 value[a]	*P* value[b]	Mean D'_{ij} (+) (95% C.I.)[c]
Any function	90	35.6	13.5	0.0001	0.504 ± 0.062 (0.383, 0.631)
Hormonal control	47	40.4	13.5	0.0001	0.592 ± 0.082 (0.442, 0.750)
Molybdenum control	21	47.6	11.9	0.0003	0.457 ± 0.098 (0.299, 0.665)
Antioxidant defense system	29	41.3	8.5	0.0018	0.397 ± 0.091 (0.242, 0.578)
Glucose metabolism	10	0	1.2	0.1369	—
Reproduction	12	41.7	4.8	0.0140	0.824 ± 0.176 (0.473, 1.000)
Functionally unrelated loci	239	16.7			0.366 ± 0.038 (0.293, 0.435)

[a] Chi-square tests for 2 × 2 contingency tables with the numbers of significant and nonsignificant pairwise comparisons for unrelated loci and for loci related by the function indicated.

[b] One-tailed probability.

[c] Each 95% confidence interval was obtained from the distribution of 1000 bootstrap means of significant D'_{ij} (+).

TABLE 5

Interallelic GD depending on allele frequency

Haplotype data set		No. haplotypes	Interallelic GD		
			Significant ($P \leq 0.05$)	Nonsignificant ($P > 0.05$)	% significant
RR	[Set 1]	9	8 (7)	1 (2)	89 (78)
RC	[Set 2]	149	33 (15)	116 (134)	22 (10)
CC	[Set 3]	171	31 (21)	140 (150)	18 (12)
	Total	329	72 (43)	257 (286)	
		χ^2_1 [Set 1 *vs.* Set 2] = 19.7 (32.5); $P \ll 0.001$ ($\ll$0.001)			
		χ^2_1 [Set 1 *vs.* Set 3] = 25.2 (27.9); $P \ll 0.001$ ($\ll$0.001)			
		χ^2_1 [Set 2 *vs.* Set 3] = 0.80 (0.39); $P = 0.37$ (= 0.53)			

R, rare allele (frequency ≤3%); C, common allele (frequency >3%). Results of analysis by the chi-square test with Yates's correction are shown in parentheses.

tive number of comparisons in significant overall GD (~44%) were *Lsp-1*γ, *Lsp-2*, and *Me*.

The intensity of significant overall GD between pairs of loci extends practically over the possible range of variation of $D'(+)$. In this case, the possible range of $D'(+)$ varies from 0 to 1, because the most common allele at each locus studied had a frequency >0.5 (see Zapata 2000). Observed $D'(+)$ values ranged from 0.017 (*Est-C* × *Xdh*) to 1.0 (*Lsp-1*γ × *Lsp-2* and *Est-6* × *Me*) and averaged 0.31 ± 0.07. Note that overall GD between *Est-C* and *Xdh* was significant although its intensity was very weak. This is readily explained, because the pair *Est-C* × *Xdh* includes some haplotypes (two out of six) carrying rare alleles in strong GD [$D'_{ij}(+) = 0.5$]. These rare alleles show nonrandom significant association but contribute very little to the increase in the intensity of overall GD, since $D'(+)$ is calculated using the values of $D'_{ij}(+)$ weighted by the frequencies of the haplotypes expected at gametic equilibrium.

Overall GD showed similar patterns of dependence on EFR and the functional relationships among loci to those detected for interallelic GD, although statistically this dependence was less consistent (analyses not shown). Naturally, overall GD analyses are less informative, because all the associations within a given locus pair are included in a single intensity measure and significance

TABLE 6

Pairs of protein loci in significant overall GD

Locus pair	cM	$D'(+)$	α'[a]	P value[b]
*Lsp-1*γ × *Lsp-2*	15.3	1.000	0.0125	0.000000
*Lsp-1*γ × *Me*	19.1	0.321	0.0125	0.000006
*Lsp-1*γ × *Po*	20.1	0.223	0.0125	0.006115
*Lsp-1*γ × *Ao-1*	20.1	0.114	0.0083	0.000006
Idh × *Est-6*	4.3	0.105	0.0250	0.007799
To-1 × *Go*	7.5	0.121	0.0500	0.010797
Est-6 × *Me*	7.7	1.000	0.0125	0.000741
Est-6 × *Xdh*	7.8	0.198	0.0083	0.000790
Est-6 × *Po*	10.0	0.044	0.0062	0.000070
Lsp-2 × *Me*	7.2	0.322	0.0500	0.000006
Lsp-2 × *Po*	9.5	0.260	0.0125	0.006115
Lsp-2 × *Ao-1*	9.6	0.137	0.0083	0.000006
Pgm × *Ao-1*	6.7	0.588	0.0250	0.013208
Est-C × *Odh*	0.7	0.129	0.0125	0.000024
Est-C × *Xdh*	2.1	0.017	0.0083	0.000038
Odh × *Xdh*	1.4	0.317	0.0125	0.009198
Po × *Ao-1*	0.1	0.381	0.0083	0.000000
Mean ± SE	8.8 ± 1.6	0.310 ± 0.072		

[a] Level of significance of 0.05, after Bonferroni correction for multiple comparisons.

[b] The smallest probability from $\chi^2_{ij}(+)$ values within each locus pair, which was used to reject the null hypothesis of overall gametic equilibrium.

test, and significant and nonsignificant associations are considered jointly.

DISCUSSION

This study provides the first comprehensive data on the distribution of gametic disequilibrium (GD) between pairs of protein loci along a single chromosome. A first striking result of the study is that GD is frequent between protein loci on the third chromosome of *D. melanogaster*, since 22% of both allele pairs (72/329) and locus pairs (17/79) were in significant GD. In addition, GD often extends over considerable chromosomal distances: 32% of the interallelic associations involved pairs of loci separated by EFRs of 15–20 cM. This incidence of GD between pairs of protein loci is substantially larger than that previously reported for the same chromosome in other natural populations: for example, 10% of 20 pairs in a population in the United Kingdom (Charlesworth and Charlesworth 1973), 10% of 20 pairs in a population in North Carolina (Langley *et al.* 1977), 20% of 15 pairs in a population in Japan, and none of 15 pairs in a population in Texas (Langley *et al.* 1974), with an average EFR for significant pairwise comparisons generally <4 cM. A number of factors may have contributed to the higher frequency of significant GD detected in this study, such as the larger number of haplotypes sampled. However, the use of the sign-based GD estimation method seems to have been the most important factor, since the percentage of significant interallelic associations decreased in our study from 22% to only 8% when both positive and negative interallelic associations were considered. In addition, our results are similar to those obtained by Zapata and Alvarez (1992) in a metaanalysis of a large number of GD studies between pairs of protein loci on the second chromosome of *D. melanogaster* and the O chromosome of *D. subobscura* (see Introduction). Taken together, these observations argue against the long-standing paradigm that GD between protein loci is a rare phenomenon, even between very closely linked loci.

The relatively large number of protein loci considered in this study has allowed us to identify the specific factors that primarily determine the distribution of GD along the third chromosome of *D. melanogaster*. Both interallelic and overall GD were distributed all along the chromosome, involving closely and loosely linked loci on the same arm and on different arms. In addition, the amount (frequency and intensity) of interallelic and overall GD varies markedly among pairs of alleles and pairs of loci. These observations suggest that the factors governing the distribution of GD cannot be described in terms of simple rules. It should be noted that the distribution of GD between rather loosely linked loci along a chromosome always depends on the relative magnitude of two antagonistic forces. Specifically, the GD generated between two loci by some evolutionary force (*i.e.*, selection, genetic drift or bottlenecks, admixture or stratification, and recent mutations) is broken down each generation by recombination during gametogenesis in doubly heterozygous individuals, at a rate determined by the recombination rate. Therefore, GD should tend to decay in proportion to the between-locus recombination rate if the processes generating GD are acting uniformly along the chromosome. Nevertheless, in this study, we did not find any indication that the strength of GD decreases monotonously with increasing between-locus recombination rate. This observation indicates that the evolutionary factor(s) causing the observed GD operate nonuniformly along the chromosome and with sufficient intensity to impose the high recombination frequencies existing between very distant loci. A common problem in GD studies is that the observed GD can be accounted for by a number of causes besides natural selection, and alternative explanations cannot be totally excluded. Nevertheless, our data set strongly supports the view that epistatic interactions in fitness are of primary importance in determining the amount and distribution patterns of gametic disequilibrium along the third chromosome of *D. melanogaster*. We found that GD was significantly higher between protein loci related by the functions of hormonal control, molybdenum control, antioxidant defense system, and reproduction than between loci without any known functional relationship. This evidence is strongly suggestive of epistatic interactions between functionally related protein loci. An interpretation of our results based on selection operating on other linked fitness genes—hitchhiking (Thomson 1977), background selection (Charlesworth *et al.* 1993), or associative overdominance (Ohta and Kimura 1970)—and not on the protein loci themselves can scarcely explain why the related loci are more prone to be in GD: it would be unreasonable to assume a scenario in which only functionally related protein loci have linked fitness genes. Recent migration events could potentially explain GD between very distant loci with different allelic frequencies among populations, even if genotype frequencies conform to HW proportions (Zapata *et al.* 2001). However, an interpretation of the observed GD based on migration would also require us to assume that functionally related loci are more differentiated among populations and this brings us back to epistasis, because genetic differentiation among populations for functionally related loci is very unlikely without epistatic interactions within each population.

It should be noted that in the past there have been many studies that have found no evidence of epistatic interactions generating GD between loci that are functionally related (*e.g.*, Charlesworth and Charlesworth 1973; Langley *et al.* 1974, 1977; Charlesworth *et al.* 1979; Loukas *et al.* 1979; Ward and McAndrew 1985). GD between related protein loci has been reported only occasionally for particular locus pairs in some natural population of species such as *Mytilus edulis* (Mitton

and Koehn 1973), *D. mojavensis* (Zouros and Johnson 1976), and *D. subobscura* (Zouros and Krimbas 1973; Fontdevila *et al.* 1983; Zapata *et al.* 2000). In contrast, our results provide the first evidence showing that extensive GD occurs between functionally related loci along a chromosome. Two-thirds of the alleles in significant GD corresponded to pairs of loci between which functional relationships exist. It is also interesting to note that the pair *Lsp-1*γ × *Lsp-2* is one of the two locus pairs showing the greatest amount of GD in our study, despite being separated by as much as 15.3 cM. In accordance with this result, the *Lsp-1*γ and *Lsp-2* loci code for the most closely related proteins. The functional relationships between *Lsp-1*γ and *Lsp-2* are well documented in *D. melanogaster*. At the end of the third larval instar, the hormone 20-hydroxy-ecdysone triggers the incorporation of larval serum proteins 1 and 2 into the fat body. These larval serum proteins are delivered to storage granules and serve as an energy and amino acid pool used during metamorphosis (Antoniewski *et al.* 1995; Burmester *et al.* 1999). The two proteins serve as nutrient reserves to support metamorphosis and reproduction (Massey *et al.* 1997) and have similar biochemical properties (Akam *et al.* 1978). It has been suggested that the two loci were generated through duplication of a single ancestral gene after separation of the Diptera from other insect orders, with subsequent independent evolution (Roberts and Evans-Roberts 1979; Mousseron-Grall *et al.* 1997). In addition, the fact that GD was detected between related protein loci separated by very considerable recombination distances (EFR of 9 cM on average) suggests that epistatic interactions along chromosome 3 of *D. melanogaster* must be very strong. These observations are of considerable relevance to the long-running neutralist-selectionist controversy (Nei 1987; Mitton 1997). Certainly, epistatic interactions involving a large number of protein loci are difficult to reconcile with the neutral theory. The between-locus GD expected as a result of genetic drift depends only on effective population size and recombination frequency (Hill and Robertson 1968); thus GD generated by genetic drift should tend to decrease with increasing recombination. Therefore, genetic drift cannot explain satisfactorily why GD is not mainly dependent on recombination frequency, but rather on the functional relationships among loci. Widespread GD among functionally related protein loci points to the importance of multilocus genetics for understanding the evolution of protein variation. Unfortunately, little is yet known about the behavior of multilocus systems under epistatic selection. The general equations describing the interaction between selection and linkage do not have explicit solutions at equilibrium even for the simplest two-locus, two-allele viability model, which imposes heuristic approaches by examining particular selection cases. Analytic and simulated selection studies on two-locus and multilocus systems were traditionally based on viability models for the two-allele case, assuming constant selection, sex-independent selection, and particular fitness parameter sets (often symmetric selection models), which considerably simplifies the algebra but results in clearly unrealistic models (Bodmer and Felsenstein 1967; Li 1967; Franklin and Lewontin 1970; Hartl and Clark 1997). Multilocus models incorporating sex-dependent variable selection, fertility, and total fitness components, as well as loci with multiple alleles, need to be explored for more realistic interpretations. Clearly, though, the development and implementation of realistic multilocus models poses mathematical and computational problems that at present are effectively intractable.

One intriguing finding of our study is that a small but important proportion (~30%) of the alleles in significant GD are at pairs of loci without any known functional relationship. This suggests that some other factor in addition to epistatic selection is generating GD in the natural population of Santa Cruz de Rivadulla. Part of this GD could be explained under the observation that two-locus haplotypes bearing rarer alleles are more frequently in GD than the remaining haplotypes. Recently, an allele frequency-dependent GD pattern was reported between microsatellite loci located on the human chromosome 11 (Zapata *et al.* 2001). This pattern was attributed to the recent origin of the rarer alleles due to the high rates of mutation at human microsatellites ($\sim 10^{-3}$) and because time is required for new variants to increase their frequency in the population. However, protein loci show rates of mutation ($\sim 10^{-5}$, Powell 1997) about two orders of magnitude lower than human microsatellites, so that other factors may also contribute. Genetic drift cannot be ruled out, taking into account that haplotypes carrying low-frequency alleles are subject to greater stochastic fluctuations due to their relatively smaller effective size. In any case, it should be pointed out that GD between rarer alleles represents only a small fraction (11%) of the total of significant cases. Interestingly, we have found that background selection (Charlesworth *et al.* 1993) can explain most of these instances of GD between unrelated loci (C. Núñez, T. Velasco and C. Zapata, unpublished results).

The results of our study suggest that epistatic interactions between functionally related protein loci have a considerable impact on the multilocus architecture of the functional genome. Hence, elucidation of proteomics presents us with the challenge of assessing the number and location of related protein loci along chromosomes and the strength of their functional relationships. Genome-wide scans for the presence of GD between protein loci may be a useful approach for unraveling proteomics. One of the best ways to assess the utility of GD for detecting relationships among uncharacterized proteins is to determine whether GD is more likely to occur among protein loci whose functional relationships are known. Our findings suggest that GD is indeed

markedly more frequent between functionally related loci than between unrelated loci. GD analyses may thus be useful not only for tentative assignment of uncharacterized polymorphic proteins to groups of functionally related proteins, but also for untangling the complex among-protein interactions that surely govern the behavior of the cell. Nevertheless, the use of GD in proteomics has two main intrinsic limitations. First, only polymorphic proteins can be analyzed. Second, evidence of GD among protein loci is not necessarily a consequence of their functional relationships. It follows that GD analyses can offer only preliminary assessments of functions and interactions, and these would then have to be confirmed by alternative methods. Certainly, further investigation is required to systematically evaluate the potential utility of GD analysis in proteomics.

We thank J. Gómez-Márquez for discussion and suggestions concerning the functional relationships among protein loci. We thank R. Montero for his technical assistance. We are grateful to an anonymous reviewer for insightful suggestions. This research was supported by grant PB96-0948 (to C.Z.) from the Ministerio de Educación y Cultura (Spain) and by fellowships from the Consellería de Educación e Ordenación Universitaria of the Xunta de Galicia (Spain) to C.N. and T.V.

LITERATURE CITED

Akam, M. E., D. B. Roberts and J. Wolfe, 1978 Drosophila hemolymph proteins: purification, characterization, and genetic mapping of larval serum protein 2 in *D. melanogaster*. Biochem. Genet. **16:** 101–119.

Antoniewski, C., M. S. O'Grady, R. G. Edmonson, S. M. Lassieur and H. Benes, 1995 Characterization of an EcR/USP heterodimer target site that mediates ecdysone responsiveness of the Drosophila *Lsp-2* gene. Mol. Gen. Genet. **249:** 545–556.

Avery, P. J., and W. G. Hill, 1979 Distribution of linkage disequilibrium with selection and finite population size. Genet. Res. **33:** 29–48.

Ayala, F. J., J. R. Powell, M. L. Tracey, C. A. Mourao and S. Pérez-Salas, 1972 Enzyme variability in the *Drosophila willistoni* group. IV. Genic variation in natural populations of *Drosophila willistoni*. Genetics **70:** 113–130.

Barker, J. S. F., 1979 Inter-locus interactions: a review of experimental evidence. Theor. Popul. Biol. **16:** 323–346.

Bodmer, W. F., and J. Felsenstein, 1967 Linkage and selection: theoretical analysis of the deterministic two locus random mating model. Genetics **57:** 237–265.

Burmester, T., C. Antoniewski and J.-A. Lepesant, 1999 Ecdysone-regulation of synthesis and processing of fat body protein 1, the larval serum protein receptor of *Drosophila melanogaster*. Eur. J. Biochem. **262:** 49–55.

Cabrera, V. M., A. M. González, J. M. Larruga and A. Gullón, 1982 Electrophoretic variability in natural populations of *Drosophila melanogaster* and *Drosophila simulans*. Genetica **59:** 191–201.

Cavener, D. R., 1980 Genetics of male-specific glucose oxidase and the identification of other unusual hexose enzymes in *Drosophila melanogaster*. Biochem. Genet. **18:** 929–937.

Cavener, D. R., and R. J. MacIntyre, 1983 Biphasic expression and function of glucose dehydrogenase in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **80:** 6286–6288.

Chakraborty, R., and K. M. Weiss, 1988 Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc. Natl. Acad. Sci. USA **85:** 9119–9123.

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

Charlesworth, B., D. Charlesworth and M. Loukas, 1979 A study of linkage disequilibrium in British populations of *Drosophila subobscura*. Genetics **92:** 983–994.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Cox-Foster, D. L., C. P. Schonbaum, M. T. Murtha and D. R. Cavener, 1990 Developmental expression of the glucose dehydrogenase gene in *Drosophila melanogaster*. Genetics **124:** 873–880.

Curie-Cohen, M., 1982 Estimates of inbreeding in a natural population: a comparison of sampling properties. Genetics **100:** 339–358.

Cypher, J. J., J. L. Tedesco, J. B. Courtright and A. K. Kumaran, 1982 Tissue-specific and substrate-specific detection of aldehyde and pyridoxal oxidase in larval and imaginal tissues of *Drosophila melanogaster*. Biochem. Genet. **20:** 315–331.

David, J. R., 1982 Latitudinal variability of *Drosophila melanogaster*: allozyme frequencies divergence between European and Afrotropical populations. Biochem. Genet. **20:** 747–761.

Dove, A., 1999 Proteomics: translating genomics into products? Nat. Biotechnol. **17:** 233–236.

Eisenberg, D., E. M. Marcotte, I. Xenarios and T. O. Yeates, 2000 Protein function in the post-genomic era. Nature **405:** 823–836.

Farkas, R., and J. Knopp, 1998 Genetic and hormonal control of cytosolic malate dehydrogenase activity in *Drosophila melanogaster*. Gen. Physiol. Biophys. **17:** 37–50.

Fontdevila, A., C. Zapata, G. Alvarez, L. Sánchez, J. Mendez *et al.*, 1983 Genetic coadaptation in the chromosomal polymorphism of *Drosophila subobscura*. I. Seasonal changes of gametic disequilibrium in a natural population. Genetics **105:** 935–955.

Franklin, I., and R. C. Lewontin, 1970 Is the gene the unit of selection? Genetics **65:** 707–734.

Girard, P., L. Palabost and C. Petit, 1977 Enzymatic variation at seven loci in nine natural populations of *Drosophila melanogaster*. Biochem. Genet. **15:** 589–599.

Goldstein, D. B., 2001 Islands of linkage disequilibrium. Nat. Genet. **29:** 109–111.

Gomes, I., A. Collins, C. Lonjou, N. S. Thomas, J. Wilkinson *et al.*, 1999 Hardy-Weinberg quality control. Ann. Hum. Genet. **63:** 535–538.

Hanly, E. W., 1980 Preliminary characterization and physical properties of pyridoxal oxidase activity from *Drosophila melanogaster*. Mol. Gen. Genet. **180:** 455–462.

Hartl, D. L., and A. G. Clark, 1997 *Principles of Population Genetics*. Sinauer Associates, Sunderland, MA.

Hedrick, P. W., 2000 *Genetics of Populations*. Jones & Bartlett, Boston.

Hedrick, P., and G. Thomson, 1986 A two-locus neutrality test: applications to humans, *E. coli* and lodgepole pine. Genetics **112:** 135–156.

Hedrick, P. W., S. Jain and L. Holden, 1978 Multilocus systems in evolution. Evol. Biol. **11:** 101–184.

Hewitt, N. E., 1974 Aldehyde oxidase of Drosophila: coordinate function with octanol dehydrogenase. Genetics **77:** s30–s31.

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

Hilliker, A. J., B. Duyf, D. Evans and J. P. Phillips, 1992 Urate-null rosy mutants of *Drosophila melanogaster* are hypersensitive to oxygen stress. Proc. Natl. Acad. Sci. USA **89:** 4343–4347.

Humphreys, J. M., B. Duyf, M.-L. A. Joiner, J. P. Phillips and A. J. Hilliker, 1996 Genetic analysis of oxygen defense mechanisms in *Drosophila melanogaster* and identification of a novel behavioural mutant with a *Shaker* phenotype. Genome **39:** 749–757.

Karotam, J., and J. G. Oakeshott, 1993 Regulatory aspects of esterase 6 activity variation in sibling *Drosophila* species. Heredity **71:** 41–50.

Klitz, W., and G. Thomson, 1987 Disequilibrium pattern analysis. II. Application to Danish HLA *A* and *B* locus data. Genetics **116:** 633–643.

Kosambi, D. D., 1944 The estimation of map distances from recombination values. Ann. Eugen. **12:** 172–175.

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

Langley, C. H., K. Ito and R. A. Voelker, 1977 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Seasonal variation. Genetics **86:** 447–454.

Lefevre, G., 1976 A photographic representation and interpretation of the polytene chromosomes of *Drosophila melanogaster* salivary glands, pp. 31–66 in *The Genetics and Biology of Drosophila*, Vol. 1a, edited by M. Ashburner and E. Novitski. Academic Press, London.

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

Lewontin, R. C., 1985 Population genetics. Annu. Rev. Genet. **19:** 81–102.

Li, C. C., 1967 Genetic equilibrium under selection. Biometrics **23:** 397–484.

Lindsley, D. L., and E. H. Grell, 1968 *Genetic Variations of Drosophila melanogaster*. Pub. 627, Carnegie Institute, Washington, DC.

Lindsley, D. L., and G. G. Zimm, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.

Loukas, M., and C. B. Krimbas, 1979 The genetics of *Drosophila subobscura* populations. X. A study of dispersal. Genetica **50:** 127–134.

Loukas, M., and C. B. Krimbas, 1980 Isozyme techniques in *D. subobscura*. Dros. Inf. Serv. **55:** 157–158.

Loukas, M., S. Tsakas, C. B. Krimbas, E. Zouros, E. Diamantopoulou-Panopoulou *et al.*, 1974 Appendix to a paper by E. Zouros *et al.* Genetics **78:** 1240–1244.

Loukas, M., C. B. Krimbas and Y. Vergini, 1979 The genetics of *Drosophila subobscura* populations. IX. Studies on linkage disequilibrium in four natural populations. Genetics **93:** 497–523.

Mantel, N., 1967 The detection of disease clustering and a generalized regression approach. Cancer Res. **27:** 209–220.

Massey, H. C. Jr., J. Kejzlarová-Lepesant, R. L. Willis, A. B. Castleberry and H. Benes, 1997 The Drosohila *Lsp-1*β gene. A structural and phylogenetic analysis. Eur. J. Biochem. **245:** 199–207.

Metzler, D. E., 2001 *Biochemistry*. Academic Press, Boston.

Mitton, J. B., 1997 *Selection in Natural Populations*. Oxford University Press, Oxford.

Mitton, J. B., and R. K. Koehn, 1973 Population genetics of marine pelecypods. III. Epistasis between functionally related isoenzymes of *Mytilus edulis*. Genetics **73:** 487–496.

Mousseron-Grall, S., J. Kejzlarová-Lepesant, T. Burmester, C. Chihara, M. Barray *et al.*, 1997 Sequence, structure and evolution of the ecdysone-inducible *Lsp-2* gene of *Drosophila melanogaster*. Eur. J. Biochem. **245:** 191–198.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M., and W.-H. Li, 1973 Linkage disequilibrium in subdivided populations. Genetics **75:** 213–219.

O'Brien, S. J., and R. J. MacIntyre, 1978 Genetics and biochemistry of enzymes and specific proteins of Drosophila, pp. 395–551 in *The Genetics and Biology of Drosophila*, Vol. 2a, edited by M. Ashburner and T. R. F. Wright. Academic Press, London.

Ohta, T., and M. Kimura, 1970 Development of associative overdominance through linkage disequilibrium in finite populations. Genet. Res. **16:** 165–177.

Pandey, A., and M. Mann, 2000 Proteomics to study genes and genomes. Nature **405:** 837–846.

Peterson, A. C., A. Di Rienzo, A.-E. Lehesjoki, A. De la Chapelle, M. Slatkin *et al.*, 1995 The distribution of linkage disequilibrium over anonymous genome regions. Hum. Mol. Genet. **4:** 887–894.

Phillips, J. P., S. D. Campbell, D. Michaud, D. Charbonneau and A. J. Hilliker, 1989 Null mutation of copper/zinc superoxide dismutase in Drosophila confers hypersensitivity to paraquat and reduced longevity. Proc. Natl. Acad. Sci. USA **86:** 2761–2765.

Powell, J. R., 1997 *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York.

Richmond, R. C., K. M. Nielsen, J. P. Brady and E. M. Snella, 1990 Physiology, biochemistry and molecular biology of the *Est-6* locus in *Drosophila melanogaster*, pp. 273–292 in *Ecological and Evolutionary Genetics of Drosophila*, edited by J. S. F. Barker, W. T. Starmer and R. J. MacIntyre. Plenum Press, New York.

Roberts, D. B., and S. Evans-Roberts, 1979 The genetic and cytogenetic localization of the three structural genes coding for the major protein of Drosophila larval serum. Genetics **93:** 663–679.

Robertson, A., and W. G. Hill, 1984 Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. Genetics **107:** 703–718.

Sham, P. C., and D. Curtis, 1995 Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Ann. Hum. Genet. **59:** 97–105.

Singh, R. S., D. A. Hickey and J. David, 1982 Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. Genetics **101:** 235–256.

Slatkin, M., 1972 On treating the chromosome as the unit of selection. Genetics **72:** 157–168.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry*. W. H. Freeman, New York.

Taillon-Miller, P., I. Bauer-Sardiña, N. L. Saccone, J. Putzel, T. Laitinen *et al.*, 2000 Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat. Genet. **25:** 324–328.

Thomson, G., 1977 The effect of a selected locus on linked neutral loci. Genetics **85:** 753–788.

Ward, R. D., and B. J. McAndrew, 1985 A search for gametic disequilibrium in the plaice, *Pleuronectes platessa*. Heredity **55:** 105–110.

Warner, C. K., and V. Finnerty, 1981 Molybdenum hydroxylases in Drosophila. II. Molybdenum cofactor in xanthine dehydrogenase, aldehyde oxidase and pyridoxal oxidase. Mol. Gen. Genet. **184:** 92–96.

Weir, B. S., and C. C. Cockerham, 1978 Testing hypothesis about linkage disequilibrium with multiple alleles. Genetics **88:** 633–642.

Williams, K. L., 1999 Genomes and proteomes: towards a multidimensional view of biology. Electrophoresis **20:** 678–688.

Zapata, C., 2000 The *D′* measure of overall gametic disequilibrium between pairs of multiallelic loci. Evolution **54:** 1809–1812.

Zapata, C., and G. Alvarez, 1992 The detection of gametic disequilibrium between allozyme loci in natural populations of Drosophila. Evolution **46:** 1900–1917.

Zapata, C., and G. Alvarez, 1993 On the detection of nonrandom associations between DNA polymorphisms in natural populations of Drosophila. Mol. Biol. Evol. **10:** 823–841.

Zapata, C., G. Alvarez, F. Rodríguez-Trelles and X. Maside, 2000 A long-term study on seasonal changes of gametic disequilibrium between allozymes and inversions in *Drosophila subobscura*. Evolution **54:** 1673–1679.

Zapata, C., S. Rodríguez, G. Visedo and F. Sacristán, 2001 Spectrum of nonrandom associations between microsatellite loci on human chromosome 11p15. Genetics **158:** 1235–1251.

Zouros, E., and W. Johnson, 1976 Linkage disequilibrium between functionally related enzyme loci of *Drosophila mojavensis*. Can. J. Genet. Cytol. **18:** 245–254.

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

Communicating editor: W. Stephan