RESEARCH ARTICLE

Nucleotide variation at the dopa decarboxylase (*Ddc*) gene in natural populations of *Drosophila melanogaster*

ANDREY TATARENKOV* and FRANCISCO J. AYALA

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525, USA

Abstract

We studied nucleotide sequence variation at the gene coding for dopa decarboxylase (*Ddc*) in seven populations of *Drosophila melanogaster*. Strength and pattern of linkage disequilibrium are somewhat distinct in the extensively sampled Spanish and Raleigh populations. In the Spanish population, a few sites are in strong positive association, whereas a large number of sites in the Raleigh population are associated nonrandomly but the association is not strong. Linkage disequilibrium analysis shows presence of two groups of haplotypes in the populations, each of which is fairly diverged, suggesting epistasis or inversion polymorphism. There is evidence of two forms of natural selection acting on *Ddc*. The McDonald–Kreitman test indicates a deficit of fixed amino acid differences between *D. melanogaster* and *D. simulans*, which may be due to negative selection. An excess of derived alleles at high frequency, significant according to the *H*-test, is consistent with the effect of hitchhiking. The hitchhiking may have been caused by directional selection downstream of the locus studied, as suggested by a gradual decrease of the polymorphism-to-divergence ratio. Altogether, the *Ddc* locus exhibits a complicated pattern of variation apparently due to several evolutionary forces. Such a complex pattern may be a result of an unusually high density of functionally important genes.

[Tatarenkov A. and Ayala F. J. 2007 Nucleotide variation at the dopa decarboxylase (*Ddc*) gene in natural populations of *Drosophila melanogaster. J. Genet.* **86**, 125–137]

Introduction

Drosophila melanogaster is a prominent genetic model organism because of its biological characteristics, and also for historical reasons, which has resulted in considerable knowledge about its population structure and genome organization. Studies of genetic variation in natural populations of *D. melanogaster* have been invaluable for understanding the mechanisms of evolution. The development of technologies for obtaining DNA sequences, coupled with the availability of the complete genome sequence, facilitates obtaining data on genetic variation in *D. melanogaster*, while accumulated knowledge provides the possibility of comprehensive data interpretation, which would be difficult or impossible in a less well-studied species. Several recent advances in our understanding of evolutionary mechanisms come from studies of *D. melanogaster*: correlation between recombination rates and genetic polymorphism (Begun and Aquadro 1992), balancing selection (Kreitman and Hudson 1991), positive selection and selective sweeps (Hudson *et al*. 1994) (review by Aquadro *et al*. 2001). Despite recent progress, however, additional sampling of both genomic and geographic regions is called for in order "to further our understanding of the ways that selection, recombination, demography and other factors have shaped genomic diversity" (Aquadro *et al*. 2001).

The *Ddc* gene codes for dopa decarboxylase (DDC, EC 4.1.1.26), which has several functions in *D. melanogaster*. The gene is located in the middle of a large dense cluster of 18 functionally related genes, on the left arm of the chromosome II (polytene band 37C1–2). The product of the gene catalyzes the decarboxylation of dopa to dopamine. Dopamine is a catecholamine that plays a fundamental role as a neurotransmitter, affecting mating behaviour, fertility, circadian rhythms, endocrine secretion, aggression, learning and memory (De Luca *et al*. 2003). In insects, dopamine also participates in cuticle pigmentation and sclerotization. More than 50 *Ddc* mutations have been isolated, most of which are recessive lethals (Wright 1996), underscoring its functional importance. Genetic variation in a 65-kb region surround-

*For correspondence. E-mail: tatarenk@uci.edu.

Keywords. Linkage disequilibrium; hitchhiking; selection; tests of neutrality.

ing the *Ddc* gene has been studied using six-cutter restriction mapping (Aquadro *et al*. 1992). The distribution of restriction site variation was in accordance with an equilibrium neutral model, but insertions and deletions were rarer than expected under neutrality. Moreover, Aquadro *et al*. (1992) found significant linkage disequilibrium in the region, and significant associations between the level of DDC enzyme activity and restriction map variants.

In a recent study, De Luca *et al*. (2003) presented evidence that polymorphism at the *Ddc* locus affects variation in *Drosophila* longevity. In particular, linkage disequilibrium mapping in a sample of 173 alleles from a population in Raleigh, North Carolina (USA), shows that three common molecular polymorphisms in *Ddc* account for 15.5% of the genetic variance in longevity attributable to chromosome 2. There is extensive linkage disequilibrium along the gene and a trend toward excess of intermediate frequency alleles (although nonsignificant), indicating the possible action of balancing selection.

Here we present an analysis of molecular variation at *Ddc* in several population samples of *D. melanogaster*. We have sequenced 2585 bp of the *Ddc* gene encompassing part of the third intron, exon 4, and the 3′ untranslated region in 40 lines of *D. melanogaster* from six distant geographic areas. For comparative purposes, we have sequenced the corresponding region in *D. simulans*. Our objective is to test whether the pattern of variation observed at *Ddc* in the Raleigh population is common to other populations worldwide, and to also examine how selection may have shaped genetic variation in this gene.

Materials and methods

Drosophila stocks

We investigated 40 *D. melanogaster* strains from six geographically distant regions. DNA for 15 strains from Montblanc, Spain, isogenic for the second chromosome, was generously provided by Dr Montserrat Aguadé (see Aguadé (1998) for details on collection and genetic manipulations). Labels for these strains are the same as used in Aguade's laboratory. Six isofemale lines are from two Caribbean islands, St Lucia and St Vincent (collected by F. J. Ayala). Nineteen isofemale lines of wild type *D. melanogaster*, from four different continents, were obtained from the Mid-America Stock Center (Bowling Green, Ohio): three strains are from Athens, Greece, five from North Kinangop, Kenya, six from Chateau Tahbilk, Australia, and five from Ohio, USA; labels for these strains in figure 1 are their respective stock numbers in the Mid-America Stock Center. We have also included in our analysis *Ddc* sequences of 12 *D. melanogaster* lines from Raleigh and two lines, 2*b* and *Ore*, that were used for mapping QTLs affecting life span (De Luca *et al*. 2003).

DNA extraction, amplification and sequencing

Genomic DNA from a single male of each line, except those from Spain, was extracted using DNAeasy Tissue Kit (QIAGEN). We used a previously published *Ddc* sequence of *D. melanogaster* (Eveleth *et al*. 1986) to design amplification and sequence primers. The amplification primers were 5′-CTTGACCTCAGCATTTTAGTTCG-3′ (forward primer) and 5′-TGTATATCAACACGAAAACTAGTC-3′ (reverse primer). The amplified region was ~ 2.6-kb long and included part of intron 3 (948 bp), exon 4 (1345 bp), and 3′ untranslated region (308 bp). Takara Ex Taq polymerase with Ex Taq buffer (PanVera Corp.) was used for amplification. PCR amplification was conducted under the following conditions: initial denaturation at 95°C for 5 min, followed by 31 cycles of denaturation at 95°C for 30 s, annealing at 57°C for 30 s, and elongation at 72°C for 2 min. Amplification of the corresponding region in *D. simulans* was conducted under the same conditions and with the same primers. The PCR products were purified with Wizard PCR Preps DNA Purification System (Promega) and directly sequenced using Big Dye Terminator Ready Reaction Kit (Perkin Elmer) on an ABI PRISM 377 DNA sequencer. The sequence of both strands was determined for each line. Sequencing primers were spaced at approximately 350 bp from one another. Inspection of the sequences and contigs assembling were performed with ABI PRISM AutoAssembler version 2.0 (Perkin Elmer). Three lines of *D. melanogaster* were heterozygous at some nucleotide positions in the studied region. PCR products from these lines were cloned using the TA Cloning Kit (Invitrogen) in order to resolve the polymorphisms at those positions. Plasmid DNA was purified with QIAprep Plasmid Kit (QIAGEN) and cloned fragments were sequenced as described above. The sequences were aligned using modules Pileup and Lineup of the Wisconsin Package (Version 9.1-UNIX, Genetics Computer Group (GCG), Madison, Wisc.).

Statistical analyses

Estimation of the evolutionary parameters and most of the standard DNA polymorphism analyses were carried out using either DNAsp version 3.53 (Rozas and Rozas 1999) or SITES (Hey and Wakeley 1997), unless mentioned otherwise.

We used the statistic K^*_{ST} (Hudson *et al*. 1992) to evaluate population differentiation. Significance of the observed values of the statistic was assessed by a permutations-based method with 10,000 replications in DNAsp version 3.99.5.

We compared the ratios between amino acid substitutions and synonymous substitutions for fixed differences between species and intraspecies polymorphism (coding region only) using the test of McDonald and Kreitman (1991).

We performed several tests of neutrality based on comparing different estimates of the population mutation rate $\theta = 4N\mu$. Under neutrality, these estimates have the same value, whereas certain departures from neutrality will affect the estimates to different degrees. Tajima's (1989) D statistic and Fu and Li's (1993) D and F statistics, with and without outgroups, were calculated using DNAsp version 3.53 (Rozas and Rozas 1999). θ_H and significance of the

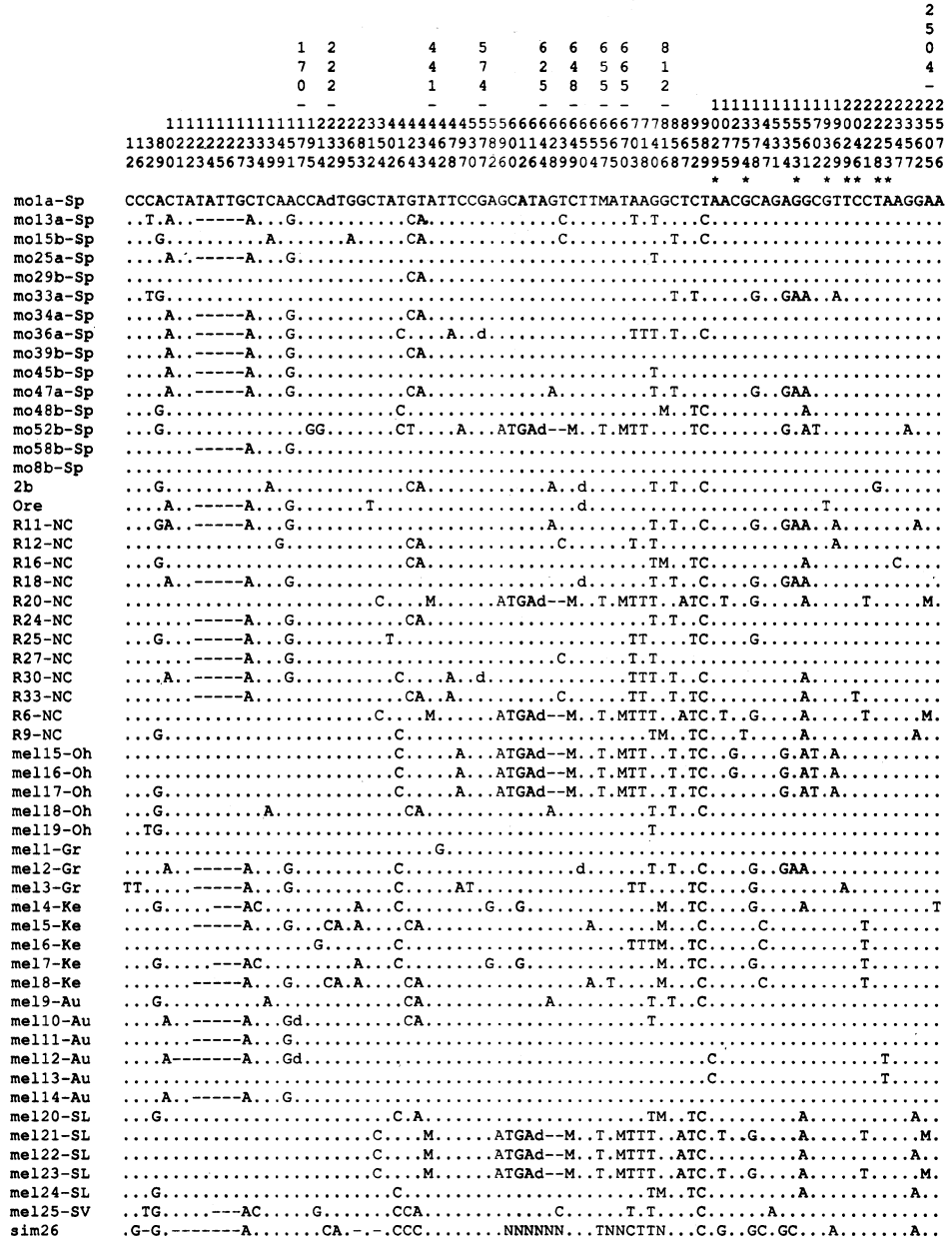

Figure 1. Nucleotide polymorphism at the *Ddc* gene region in 54 strains of *Drosophila melanogaster* from seven natural populations of *D. melanogaster*. Lines are from Spain (Sp), Raleigh (North Carolina, NC), Greece (Gr), Kenya (Ke), Australia (Au), Ohio (Oh), and the Caribbean (St Lucia, SL; and St Vincent, SV), 2b and Ore are lines used by De Luca *et al*. (2003) for mapping QTLs affecting life span. Sim26 is a sequence of *D. simulans*. Dots represent the same nucleotide as in the first sequence. Asterisks(*) show sites with nonsynonymous polymorphism. The alignments between *D. melanogaster* and *D. simulans* for the intervals 599–644, 662–670, and 812–828 are unreliable and are omitted (coded as gaps) in the analyses including interspecific comparisons; these segments are represented by *N* in the *D. simulans* sequence. *M* designates single-step complex mutations and is encoded as single nucleotide polymorphisms (SNP). *d* designates insertions or deletions (indels); *d*− indicates that there is nucleotide polymorphism in the region spanned by a deletion. Polymorphism spanned by deletions (indels) is not considered in the statistical analyses, but the indels themselves are included as SNPs. Indel polymorphism 121–127 is not considered, because due to several separate mutations (events) in uncertain order.

difference between θ_π and θ_H (H-test, Fay and Wu 2000) were calculated using a program from J. Fay (http://crimp.lbl.gov/htest.html), which implements the coalescent with recombination.

Another test of neutrality, HKA (Hudson *et al*. 1987), evaluates the neutral hypothesis by comparing the correlation of polymorphism and divergence at two (or more) loci. We used as a neutral-locus reference the *Adh*-5′ of *D. melanogaster*, even though it was studied in a sample that was a mixture from different populations. We also used as a reference locus the *Idgf3*, which appears to evolve neutrally (Zurovcova and Ayala 2002). This locus was studied in the same sample from Spain, and indeed using many of the same chromosomes, as in the present *Ddc* study. Additionally, we have compared different functional parts of the studied region (intron, exon 4, and 3′ untranslated region), to ascertain whether they are evolving homogeneously.

Heterogeneity in the polymorphism-to-divergence ratio across the DNA region was tested using statistics proposed by McDonald (1998) and implemented in his program DNA Slider. The test was applied with a range of recombination rates in order to account for heterogeneity in polymorphism levels between fragments caused by recombination. Unlike the HKA test, the McDonald tests do not require a priori decision about subdivision of the DNA region into the parts to be compared, and thus may be preferable to the HKA test for single gene studies.

Linkage disequilibrium (r^2, the squared correlation of allele frequencies) was calculated between pairs of polymorphic informative sites segregating for two nucleotides, and significance was determined by Fisher's exact test, using SITES (Hey and Wakeley 1997). Sign tests of Lewontin (1995) were applied to test the overall degree of linkage disequilibrium, as well as the direction of linkage associations. The sign tests make use of polymorphic sites with asymmetrical allele frequencies and thus are informative in situations when two-by-two contingency tables are not applicable. Initially, adjacent pairs of polymorphic sites are classified into classes according to the absolute numbers of copies of rarer alleles, signs of linkage associations (measured by D) are determined for each pair, and numbers of negative and positive pairs are counted for each class. These observed numbers of pairs are compared with the expected values (calculated under the null hypothesis of no disequilibrium) using the likelihood ratio statistic, G, for each class. In one test proposed by Lewontin (1995), the G-values are summed across classes and compared with chi-square distribution with degrees of freedom equal to the number of classes. This summed-G tests overall linkage disequilibrium in the region. In another test, observed and expected numbers of pairs with negative D are summed across classes and compared using goodness-of-fit test with one degree of freedom. This tests for a general bias toward a negative or positive association. Lewontin (1995) showed that classes of pairs of sites with rare alleles (e.g. singletons) have a high power to detect coupling disequilibria but very low power to detect repulsion disequilibria. If no genuine coupling disequilibrium is present in the sample, presence of classes with rare alleles may obscure detection of repulsion disequilibrium. In order to increase power for the detection of negative associations (i.e. repulsions), the summation of G-values or numbers of pairs of certain sign can be conducted for classes of sites with even allele frequency distribution i.e. omitting pairs of sites with a few copies of alleles. In our case, we omitted classes with pairs of sites having singletons (in addition to using all classes).

Results

DNA sequence variation

Figure 1 gives a summary of nucleotide sequence variation in

Table 1. Summary statistics of populatin genetic variation at the *Ddc* gene in *D. melanogaster.*

Population	N	S	θ	θ/bp	π	π/bp	D
Spain	15	37	11.379	0.00440	9.81	0.00379	−0.587
Raleigh	12	42	13.908	0.00538	14.23	0.00550	0.105
Ohio	5	26	12.480	0.00483	14.40	0.00557	1.146
Greece	3	21	14.000	0.00542	14.00	0.00542	n.a.
Kenya	5	24	11.520	0.00446	12.40	0.00480	0.568
Australia	6	13	5.693	0.00220	6.20	0.00240	0.544
Caribbean	6	30	13.139	0.00508	15.40	0.00596	1.090
Total	54	72	15.800	0.00611	13.99	0.00541	−0.398

N, sample size (number of chromosomes); n.a., Tajima's test could not be performed due to low sample size; S, number of segregating sites; θ and θ/bp, Watterson's (1975) estimator of population parameter $4N\mu$, per gene and per base pair, respectively; π and π/bp, the average number of nucleotide differences between two sequences, per gene and per base pair, respectively; D, Tajima's (1989) test statistic, calculated using the total number of segregating sites. The length of the region is 2585 bp. Total includes all populations, and also lines 2b and Ore.

Table 2. Nucleotide polymorphism for separate segments of *Ddc*.

Region	Intron[a] 1–948	Intron[b] 1-948	Coding Region 949–2293		3′ region 2294–2601	Silent[b]	Total[b] 1–2601
			Syn	Nsyn			
Length (bp)	932	811	313.6	1030.4	306	1430.6	2462
Total data set (54 lines)							
π	0.0107	0.0096	0.0098	0.0005	0.0014	0.0079	0.0047
θ	0.0113	0.0103	0.0084	0.0017	0.0036	0.0084	0.0056
K		0.0613	0.1367	0.0011	0.0491	0.0753	0.0044
Spain							
π	0.0079	0.0078	0.0067	0.0002	0.0004	0.0060	0.0036
θ	0.0096	0.0083	0.0069	0.0003	0.0010	0.0066	0.0039
K		0.0616	0.1366	0.0010	0.0492	0.0754	0.0442
Raleigh							
π	0.0105	0.0094	0.0098	0.0006	0.0025	0.0080	0.0049
θ	0.0100	0.0086	0.0084	0.0010	0.0033	0.0074	0.0047
K		0.0610	0.1363	0.0011	0.0493	0.0750	0.0441
Ohio							
π	0.0112	0.0086	0.0127	0	0	0.0077	0.0045
θ	0.0098	0.0077	0.0107	0	0	0.0067	0.0039
K		0.0614	0.1376	0.0010	0.0490	0.0755	0.0443
Greece							
π	0.0100	0.0115	0.0106	0.0013	0	0.0089	0.0057
θ	0.0100	0.0115	0.0106	0.0013	0	0.0089	0.0057
K		0.0629	0.1338	0.0013	0.0490	0.0755	0.0444
Kenya							
π	0.0101	0.0099	0.0083	0	0.0013	0.0077	0.0045
θ	0.0093	0.0089	0.0076	0	0.0016	0.0070	0.0041
K		0.0597	0.1356	0.0010	0.0497	0.0742	0.0435
Australia							
π	0.0052	0.0059	0.0011	0.0010	0	0.0036	0.0025
θ	0.0047	0.0054	0.0014	0.0009	0	0.0034	0.0023
K		0.0621	0.1366	0.0016	0.0490	0.0756	0.0446
Caribbean							
π	0.0125	0.0089	0.0083	0	0.0037	0.0076	0.0044
θ	0.0103	0.0076	0.0084	0	0.0029	0.0067	0.0039
K		0.0604	0.1395	0.0010	0.0485	0.0752	0.0441

π, the average number of nucleotide differences between two sequences, per site; θ, Watterson's (1975) estimator of population parameter $4N\mu$, per site; K is the average number of nucleotide differences per site between *D. melanogaster* and *D. simulans*.
[a]Values of π and θ for the whole length of the intron in *D. melanogaster*, including stretches of nucleotides absent in *D. simulans.*
[b]Values of π and θ for homologous stretches of the intron present in both *D. melanogaster* and *D. simulans.*

the *Ddc* region, sampled from seven populations distributed worldwide. Although our sample sizes are small, except for samples from Spain and Raleigh, consideration of several populations can be helpful in deciding whether the same evolutionary pattern exists over the species range. Summary statistics of genetic variation, for each population and for the worldwide collection, are presented in table 1.

We observed 72 polymorphic nucleotide sites in the total data set (including insertion/deletion variation and complex mutations). Actually, variation was somewhat higher than indicated in table 1, because a few sites segregated within indels, and were excluded from the analysis (e.g. positions 628, 639). Some mutations considered as single-step complex mutations may be several-step mutations. Insertion-length variation spanning positions 121–127 clearly arose as a result of several mutations, but it was not considered in the analyses due to its complexity. Among 72 polymorphic sites, eight were replacement polymorphisms (figure 1).

Table 2 gives estimates of nucleotide variation for the different functional parts of the studied region. The levels of

silent variation in the coding region and in the intron are quite similar, although somewhat higher in the intron. The level of silent variation in the 3′ untranslated region is noticeably lower: in three populations, Greece, Australia, and Ohio, no variation is found; in the well-sampled Spanish population, the region is 15 and 18 times less variable than synonymous sites in the coding region and intron, respectively.

Estimates of nucleotide variation in the Kenya population are comparable to estimates in other populations, with the exception of Australia. The level of nucleotide variability in the Australian population is two times lower than, in the other populations, for the genetic region as a whole, as well as for its functional parts separately.

The *Ddc* gene is located in a region of low to intermediate recombination (recombination rate $r = 0.00112$ for 37C chromosome map position; Kliman and Hey 1993). In the ~ 2.6-kb region studied, the four-gamete test (Hudson and Kaplan 1985) detects 8 recombination events in the Spanish sample, 10 events in Raleigh, and 16 recombination events in the total sample of *D. melanogaster*. Linkage disequilibrium between sites in the populations from Raleigh and Spain are summarized in figure 2 (open and closed symbols, respectively), and disequilibria in the total data set are shown in figure 3. The graphs show significance of the nonrandom associations between pairs of nonunique polymorphic sites. In the Spanish sample, the number of significant cases is more than expected by chance; eight out of 136 are significant with $P < 0.01$. There are two small clusters of significant associations between nearby pairs of sites (positions 89–151 and 1378–1561). The second cluster is probably a case of spurious association, because it results from the correlation of rare alleles, and the association is absent in other populations. Noteworthy cases of significant association are between the relatively distant sites 989 and 416, 989 and 703, 952 and 1561. These sites are highly heterozygous, variable in most populations, and significant association between these pairs of sites is evident in other populations as well (but not in Raleigh). Another noteworthy association is between sites 100 and 740 (see Discussion). The pattern of nonrandom associations is quite different in Raleigh. As mentioned above, some of the sites that show significant disequilibria in Spain are not associated in Raleigh. The overall degree of nonrandom associations appears to be rather high; 99 out of 595 pairwise comparisons are significant with $P < 0.05$ (i.e., every sixth comparison). However, the great majority of the nonrandom associations are caused by the presence

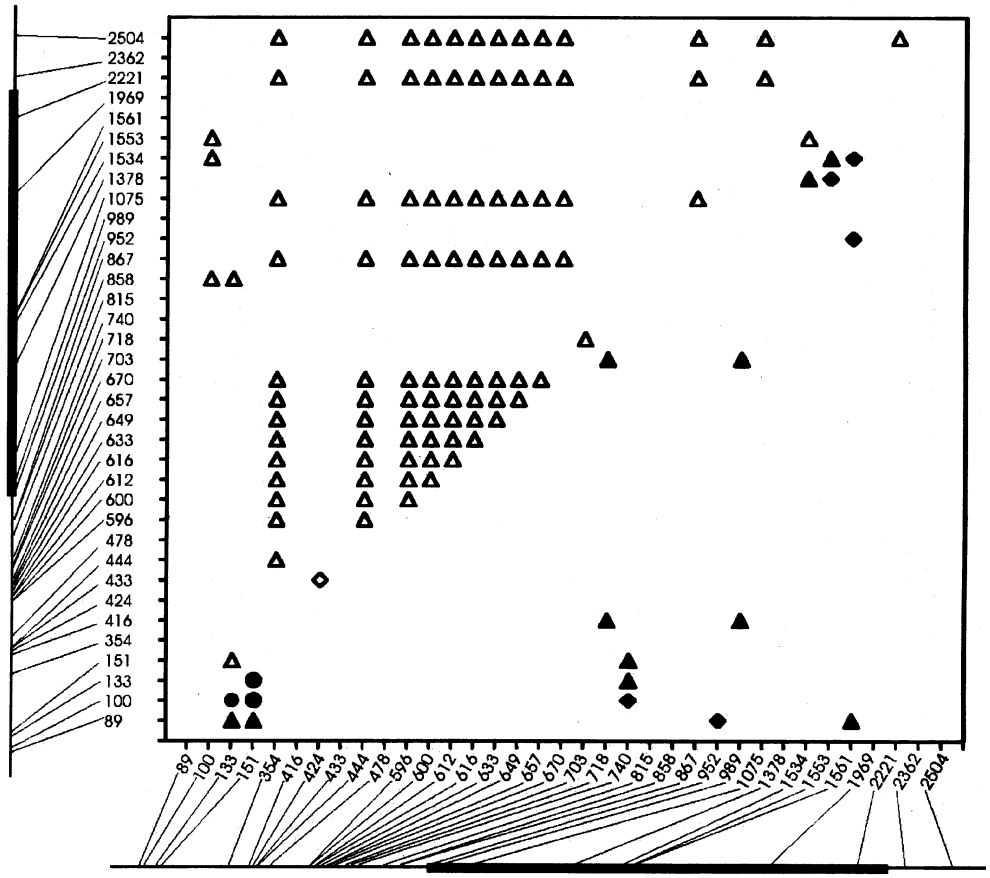

Figure 2. Linkage disequilibrium between pairs of polymorphic informative sites in Raleigh (upper left corner, open symbols) and Spain (lower right corner, closed symbols). Significance determined by Fisher's exact test, triangles, $P < 0.05$; diamonds, $P < 0.01$; circles, $P < 0.001$.

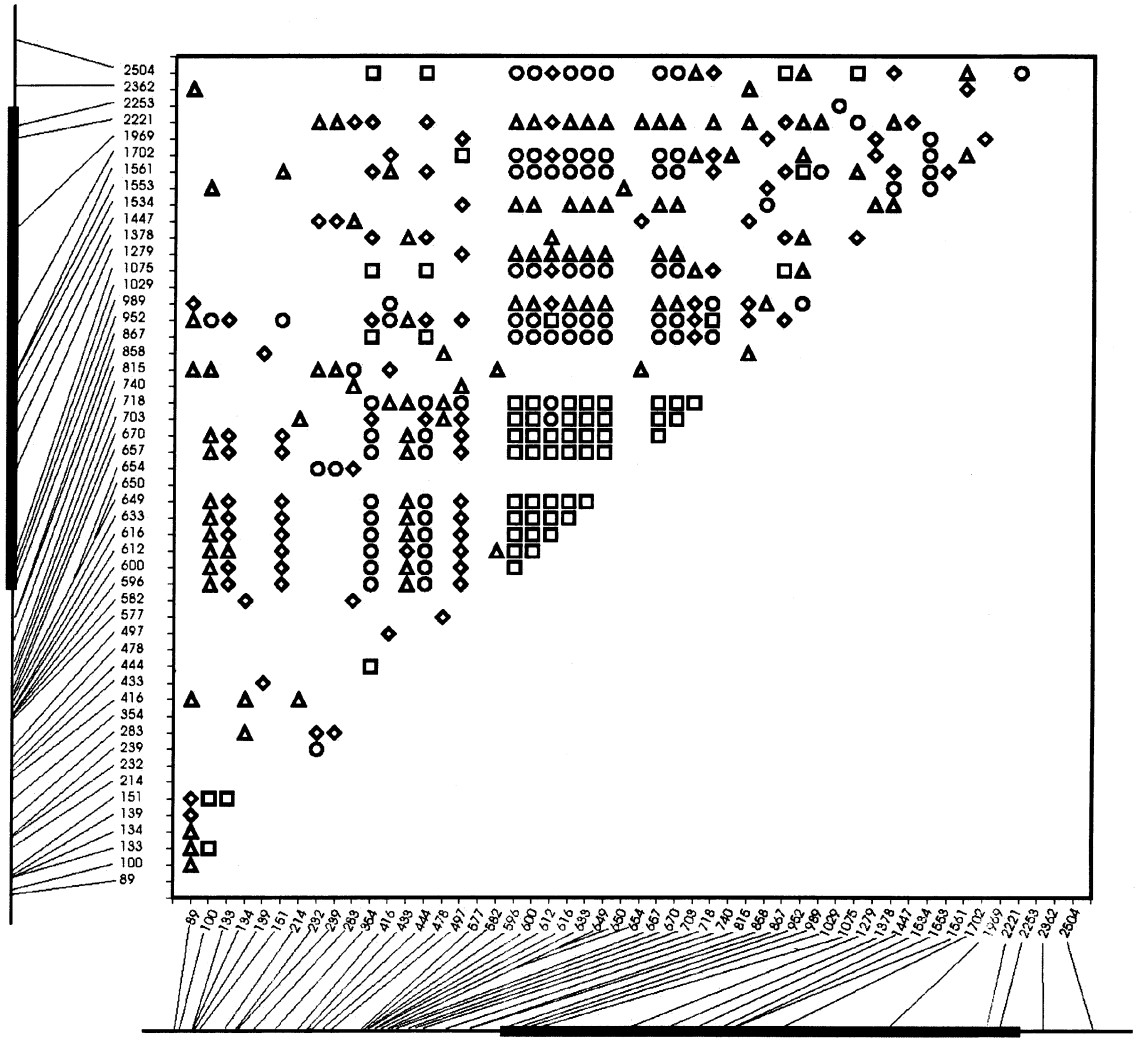

Figure 3. Linkage disequilibrium between pairs of polymorphic informative sites in the total data set. Significance determined by Fisher's exact test, triangles, $P < 0.05$; diamonds, $P < 0.01$; circles, $P < 0.001$; squares, significant after Bonferroni correction.

of two identical sequences (R20 and R6) that are quite differentiated from other sequences. Dropping one of the sequences from the analysis eliminates most of the disequilibria. In the total data set of 54 *D. melanogaster* sequences, there are a considerable number of significant associations (figure 3). Thus, 146 out of 1326 comparisons are significant with $P < 0.001$, whereas only 14 significant cases are expected by chance at such level. Moreover, 60 comparisons are significant after correction for multiple testing.

The Fisher's exact test of linkage disequilibrium is often not applicable because many alleles are present only once or twice in a sample. To address this problem, and also to obtain an overall estimate of linkage disequilibrium and some idea on the direction of the linkages, we performed Lewontin (1995) sign test. The results based on biallelic sites are shown in table 3 for the Spanish and Raleigh populations, and also for the total data set of *D. melanogaster*. In the Spanish population, the total G for goodness-of-fit with the conservative Williams' correction for continuity is 41.62 with 14 degrees of freedom ($P < 0.001$), indicating extensive nonrandom associations. The expected total number of negative associations is 25.01, but only 14 are observed ($G = 14.96$, $P < 0.001$); thus, there is a significant excess of positive associations. Mixture of divergent haplotypes can be one reason for the excess of positive associations, i.e. of rare alleles at different sites in coupling. Allele *mo52b-Sp* is one of the most divergent haplotypes in the Spanish sample. However, there is still an excess of positive associations when it is excluded (the observed number of positive D values is 16 and the expected number is 8.2, $G = 11.46$, $P = 0.0007$). Neither of the two Lewontin (1995) sign tests rejects the null hypothesis of no true linkage disequilibrium in Raleigh or in the total data set. Lewontin (1995) has shown that inclusion of pairs of sites with asymmetric allele fre-

quencies provides high power to the sign test for detecting positive disequilibrium but low power for the detection of negative associations. Since the presence of classes with rare alleles may obscure detection of the excess repulsions, we excluded the class of pairs of sites having singletons. Both sign tests were in perfect accord with the null hypothesis of no true linkage disequilibrium in the Spanish or Raleigh population, or in the total data set of *Ddc* sequences.

Genetic differentiation between populations

If all seven samples are considered simultaneously, differentiation is high and statistically significant (K^*_{ST} = 0.092, $P(K^*_{ST})$ = 0.000; table 4). However, not all samples appear to be equally diverged. The population from Greece, which is represented by only three sequences, does not show statistically significant differences in any pairwise comparison and so it will not be discussed further. The Spanish and the Australian populations do not differ significantly from each other as measured by K^*_{ST} (Hudson *et al*. 1992), but they differ significantly from the Kenya and the two American populations in all pairwise comparisons (K^*_{ST} values range from 0.023 to 0.147 with the average being 0.090). Differentiation is also pronounced between the Kenya and the two American populations (average K^*_{ST} = 0.069; all pairwise comparisons are statistically significant). The populations from Raleigh and the Caribbean are the closest (non-significant difference), while the Ohio population is more distinct (K^*_{ST} = 0.051, P = 0.02 between Ohio and Raleigh; K^*_{ST} = 0.085, P = 0.052 between Ohio and Caribbean). Altogether, with samples of only 5 to 15 alleles from each locality, there is evidence of differentiation between populations of *D. melanogaster* from different continents, but also between samples from the same continent (e.g. Ohio versus Raleigh).

Strong population subdivision is also evident in the presence of sites with fixed differences. The complex mutation spanning positions 812–816 observed in all five sequences from Kenya is absent in samples from Greece, Australia, and Ohio. This mutation is observed in low frequencies in the Spanish (1/15) and Raleigh (2/12) populations, but has a higher frequency in the Caribbean sample (2/6). All pairwise comparisons between Kenya and the other samples are statistically significant with regard to this polymorphism (Fisher's exact test: $P < 0.05$ for Greece and Caribbean; $P < 0.01$ for Raleigh, Ohio and Australia; $P < 0.001$ for Spain).

Nucleotide polymorphism, divergence and tests of neutrality

In the coding region, differences are fixed between *D. melanogaster* and *D. simulans* at 38 synonymous sites; 12 synonymous sites are polymorphic in *D. melanogaster* (including one site, 1075, which does not have a common nucleotide with *D. simulans*). There is only one fixed nonsynonymous difference at site 1553. This site is polymorphic

Table 3. Sign tests (Lewontin 1995) of linkage disequilibrium in the *Ddc* gene.

Sample	Test based on summing G # classes	G^c	Test based on summing negative pairs Pairs with negative D Obs	Exp	$G(d.f. = 1)$
Spain[a]	14	4162*	14	25.01	14.96*
Spain[b]	8	11.20	4	6.01	1.37
Raleigh[a]	10	12.76	26	23.88	0.47
Raleigh[b]	7	11.29	14	13.05	0.14
Total[a]	41	41.81	51	50.93	0.00
Total[b]	30	31.14	21	21.26	0.01

*$P < 0.001$.
[a]using all polymorphisms;
[b]excluding singletons;
[c]summed G with William's (1976) correction.

Table 4. Population subdivision (K^*_{ST}) among six populations of *D. melanogaster.*

	Spain	Raleigh	Ohio	Greece	Kenya	Australia
Raleigh	0.023*					
Ohio	0.081**	0.051*				
Greece	0.016	−0.006	0.090			
Kenya	0.074***	0.045**	0.137*	0.059		
Australia	−0.013	0.032*	0.147*	0.066	0.128**	
Caribbean	0.091***	0.008	0.085	0.069	0.093*	0.146**

*$P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

in *D. melanogaster* for nucleotides A and G (amino acids Met and Val), whereas *D. simulans* has C (Leu). Eight nonsynonymous sites (including 1553) are polymorphic in the whole sample of *D. melanogaster*. Thus the polymorphism-to-divergence ratios are different for synonymous (12:38) and nonsynonymous sites (4:1). The McDonald–Kreitman (MK) test is significant ($P < 0.001$, Fisher's exact test). If applied separately to each population, the MK test yields significant values in the samples from Raleigh, Greece, and Australia.

Table 2 shows the levels of interspecific nucleotide divergence (K) at silent and synonymous sites. As in the case of intraspecific polymorphism, the lowest level of divergence is in the 3′ untranslated region. However, the discrepancy between regions is less pronounced than in the case of polymorphism: only 1.2 times higher in the intron than in the coding region, and 2.8 higher in synonymous than in nonsynonymous sites of the coding region. Although the levels of intraspecific polymorphism are approximately similar in the intron and synonymous sites, interspecific divergence is 2.2 times lower in the intron. The HKA test (which compares polymorphism and divergence; Hudson *et al*. 1987) for the intron, coding, and 3′ untranslated region is not significant for any separate region or for all combined, either separately for each population or for the total data set, suggesting that intraspecific polymorphism and interspecific divergence are correlated in these regions. The HKA test for the region as a whole was compared against the presumably neutrally evolving genes *Adh*-5′ and *Idgf3*, which are both on the second chromosome. Fifteen of the 20 homozygous lines from Spain studied for *Idgf3* are the same as for *Ddc*. None of the tests were statistically significant, either for the whole *Ddc* data set of *D. melanogaster* or separately for the Spanish and Raleigh populations.

We have conducted the McDonald (1996, 1998) tests for heterogeneity in the ratio of polymorphism-to-divergence across the studied region, separately for the total set of 54 *D. melanogaster* sequences, for the 15 sequences from the Spanish population, and for the 12 sequences from Raleigh. When the total sample is considered, the heterogeneity is significant using Kolmogorov–Smirnov statistic D_{KS} for silent polymorphic sites ($D_{KS} = 0.070$, $P = 0.022$, 10,000 coalescent simulation runs). Another statistic, the number of runs of contiguous polymorphisms and contiguous fixed differences, points towards heterogeneity ($P < 0.1$). Similar results are obtained for the Spanish sample: the Kolmogorov–Smirnov statistic is significant ($D_{KS} = 0.064$, $P = 0.03$, 10,000 coalescent simulation runs), and the number of runs is close to significant ($P = 0.07$). In Raleigh, however, none of McDonald's tests is significant, although the runs of contiguous polymorphisms and contiguous fixed differences tend to be longer than expected under the neutral model ($P < 0.1$). Our results are consistent with McDonald (1998) conclusion that the Kolmogorov–Smirnov statistic is suitable for detecting single changes across a gene from higher to lower polymorphism; for our data the polymorphism-to-divergence ratio decreases from highest in the intron to lowest in the 3′ untranslated region (figure 4). The slope is less pronounced in the Raleigh sample, for which D_{KS} is not significant.

Neither the Tajima (1989) nor the Fu and Li (1993) tests detect any departure from neutral expectations in any population studied, whether in the complete region or only in the coding part. Even though these two tests support the hypothesis of neutrality, there is an indication that *Ddc* has been

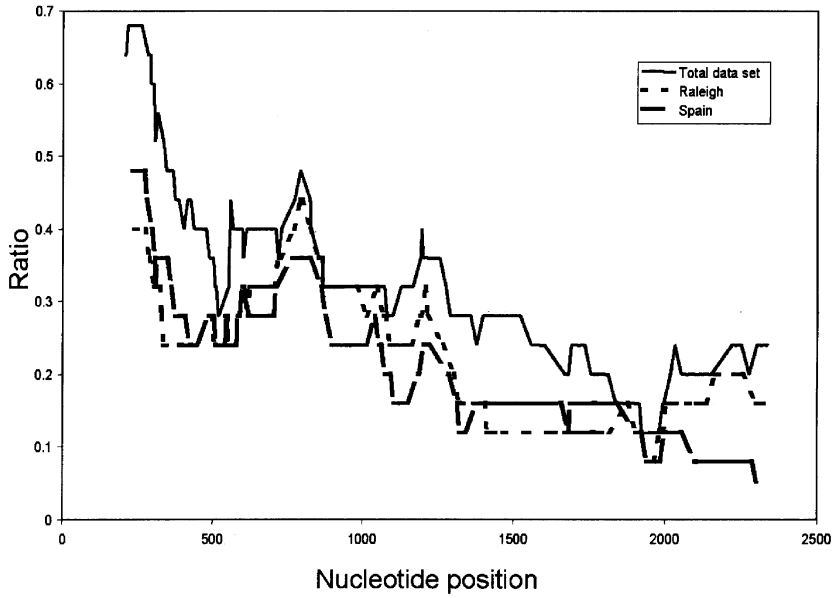

Figure 4. Sliding-window plot of polymorphism-to-divergence ratio for *Ddc* region. Window length is 25 silent variable sites. *D. simulans* is used as outgroup to *D. melanogaster*.

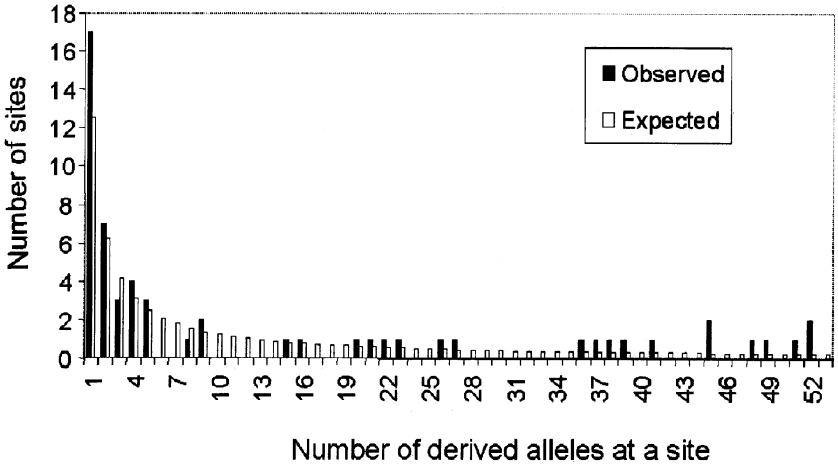

Figure 5. Observed and expected frequency spectra of the newly derived nucleotides in *Ddc* of *D. melanogaster*.

influenced by natural selection in the recent past. Figure 5 shows the observed and expected frequency spectra of newly derived nucleotides for the total *Ddc* data set. There is an excess of sites with derived nucleotides at high frequency, which is advocated as a unique feature of a hitchhiking (Fay and Wu 2000). We applied the H-test of Fay and Wu (2000) to examine the fit between the observed and expected frequency spectrum under neutrality. Assuming no recombination, the probability of the observed frequency distribution is low but not significant ($H = -9.3$, $P = 0.085$). However, this assumption is very conservative. As discussed above, at least 16 recombination events are evident for the whole data set, which can be taken into consideration when estimating the significance of H. The estimate of the recombination parameter R (Hudson 1987), based on the variance of the average number of nucleotide differences between pairs of sequences, is 22.8 for the studied region, or 0.0088 per bp. Another estimate of the recombination parameter R, the gamma estimate of Hey and Wakeley (1997), also based on polymorphism data, is 92.3 per region, or 0.0357 per bp. An independent estimate of the recombination parameter can be calculated using an estimate of the recombination rate r, based on the relationship between the physical and genetic maps of *D. melanogaster*. Using as estimate of the recombination rate $r = 0.00112$ centimorgans per 1000 bp per generation for the band 37C where *Ddc* is localized (Kliman and Hey 1993), and assuming $N_e = 10^6$ (Przeworski *et al.* 2001), and no recombination in males, we obtain $R = 57.9$ per the studied region, or 0.0224 per bp. Using the more conservative estimate of $R = 22.8$ *sensu* Hudson (1987), the observed excess of sites with derived nucleotides becomes significant (Fay and Wu's H-test, $P < 0.05$).

Discussion

Analysis of nucleotide variation at *Ddc* reveals differences both within and among populations. The Australian population is only about half as variable as the others, perhaps as a consequence of a founder or bottleneck effect. Despite small sample sizes, there is significant genetic differentiation among populations from different continents, except between the Australian and Spanish populations, which might indicate that the Australian population is derived from an European population. Strong population subdivision between samples from different continents with similar values of F_{ST} have been reported for many genes (Begun and Aquadro 1993, 1995), although small population differentiation has been reported for some genomic regions (Andolfatto and Kreitman 2000).

There is significant linkage disequilibrium in *Ddc*. In the total data set, 11% of pairwise comparisons are significant with $P < 0.001$, and 60 comparisons remain significant after the Bonferroni correction. Many significant cases may have resulted from pooling genetically differentiated populations. However, there are several cases that are not a result of the mixing of samples: a block of sites between positions 596 and 718 exhibits very strong linkage disequilibrium; there is also strong association between this block of sites and sites at positions 354, 444, 867, 952, 1075, 1561, 1702, and 2504. All these sites are segregating in more than one population, which may be an indication that the variation is not recent.

Is this stretch of strongly associated sites maintained by selection, or could recent migration alone account for it? Inspection of the polymorphisms (figure 1) shows that the cases of strong linkage disequilibrium are due to the presence of five haplotypes represented by nine sequences. There is limited recombination between these haplotypes and the rest of the sequences. The R20 and R6 sequences from Raleigh are identical to two sequences from St Lucia (Caribbean) and, therefore, at least in one of these populations linkage disequilibrium may have arisen owing to recent immigration of the diverged sequences. But recent admixture does not seem the only explanation for the linkage disequilibrium. The five haplotypes must be quite old because they are fairly divergent among themselves and have wide distribution, indicat-

ing that they must have persisted in their local populations for long time, given that they display a geographical pattern of differentiation. Epistasis or inversion polymorphisms thus appear as more likely explanations for the observed nonrandom associations. One particularly interesting association that deserves separate mentioning is between sites 100 and 740 in the Spanish population. Both sites are highly variable (gene diversity is about 0.5), alleles are embedded in diverse haplotypes, and yet they are positively associated. Site 740 is one of three sites in the *Ddc* locus that have been implicated in regulating life span De Luca *et al*. (2003), but there is no association between sites 100 and 740 in any other than the Spanish population.

Lewontin (1995) sign test, developed specifically for the analysis of nucleotide variation, confirms strong positive associations in the Spanish sample, but does not show evidence of linkage disequilibrium in Raleigh or in the total data set. The results of the sign test seem to contradict the conventional analysis of linkage disequilibrium. In the total data set we found extensive nonrandom associations encompassing the whole studied region (figure 3). A possible explanation for the discrepancy might be that, in the sign tests, D is calculated between adjacent sites, which are expected to be in the greatest disequilibrium. However, this expectation does not always hold. Figures 2 and 3 show that most significant nonrandom associations are between distant sites separated by stretches with polymorphic sites in linkage equilibrium. We see little clustering of significant scores near the diagonal (i.e. between neighboring sites). Sign tests summarizing associations among adjacent sites may not be sensitive when distant sites are in statistically significant linkage disequilibrium but numerous sites between them are in linkage equilibrium.

Despite significant associations between distant sites, in general linkage disequilibrium decays rapidly with distance. In the total sample, the average value of r^2 between sites less than 500-bp apart is 0.139, but 0.061 for sites of 500–1000-bp apart. This is in accordance with the prediction of linkage disequilibrium decay based on recombination rates in the region: the estimate of the recombination parameter R (Hudson 1987) is 0.0088 per bp, so that linkage disequilibrium should quickly decay between sites separated more than 114 bp (= 1/0.0088). What is unexpected is that average value of r^2 between even more distant sites does not decrease but rather becomes higher, 0.085 for sites separated by distances of 1000–1500 bp, and 0.068 for sites more than 1500 bp apart. This points once more to the presence of long distance nonrandom associations in the *Ddc* locus.

The McDonald–Kreitman test indicates a deficit of fixed nonsynonymous differences in the coding region of *Ddc* in the total data set and in Raleigh, Greece, and Australia. We suggest that this deficit may be due to negative selection, which eliminates alleles harbouring nonsynonymous mutations (potentially deleterious). This suggestion is supported by the observation that, at seven out of eight nonsynonymous sites, new variants are in low frequencies. Five substitutions (at positions 1354, 1932, 2029, 2046 and 2228) are observed as singletons, and substitutions at positions 1029 and 2253 are observed in only two similar sequences from Australia (interestingly, both mutations are present in the same sequences). Tajima (1989) test applied specifically to the replacement polymorphisms significantly deviates from zero ($D = -1.92$, $P < 0.01$), which is consistent with the hypothesis that they are slightly deleterious. It is noteworthy that out of the seven replacement polymorphisms at low frequency, six substitutions are nonconservative replacements (which change the polarity of amino acids and thus have a strong effect on physicochemical properties). Amino acid exchanges at site 1553, which is the only fixed replacement between *D. melanogaster* and *D. simulans*, and which is also the only highly polymorphic replacement site in *D. melanogaster*, are conservative.

Excess of replacement polymorphisms relative to divergence might also be due to recent relaxation of selective constraints on the protein. However, the low frequencies of the derived replacement variants argue against such an explanation. If replacement substitutions behaved neutrally, more of them would be expected to reach intermediate or high frequencies.

Using a conservative estimate of the recombination rate, Fay and Wu (2000) H-test detects an excess of sites with derived nucleotides in high frequency. Population structure (Przeworski 2002) and certain demographic scenarios (Lazzaro and Clark 2003) can cause values of H to depart from neutral expectation. The pooling of data from genetically differentiated populations does not account for the results of the H-test in our data. First, with our sampling scheme in which particular populations make between 6 and 28% of the total data set, such pooling should increase intermediate frequency and low-frequency variation and thus make the test more conservative. Second, excess of sites with derived nucleotides at high frequency is observed in the separate populations as well. In the Spanish population, for example, the H-test is significant ($H = -7.98$, $P < 0.05$), assuming recombination as described above. We also note that H-values are negative and probabilities are quite low in all other populations, with the exception of Greece (in which only three sequences are studied). If the probabilities from separate tests for each of six populations (except Greece) are combined using Fisher's method, the H-test is significant even under the conservative assumption of no recombination ($\chi^2 = 27.13$, d.f. = 12, $P = 0.012$).

The availability of samples from different regions of the world also allows us to discard bottlenecks as the explanation for the H-values (Lazzaro and Clark 2003). If derived alleles increased in frequency solely due to genetic drift, it would be unlikely that the same alleles would change in the same direction in independent populations. Therefore, the effect of bottlenecks should disappear when pooling samples, which is not the case.

More generally, demography is expected to have a genome-wide effect. We performed H-tests for five genes from the second chromosome (*Acp26Aa, Acp26Ab, Acp29Ab, Idgf1*, and *Idgf3*) sampled from the same Spanish population as ours. The H-values are negative for *Acp26Aa* and *Acp29Ab* (i.e., genes previously shown to be influenced by positive selection), but the H-values are positive for *Acp26Ab, Idgf1*, and nonsignificantly negative for *Idgf3*. Clearly, at least for the Spanish sample, demographical factors are not the only evolutionary force shaping variation and overriding the consequences of natural selection.

We favour hitchhiking as the explanation for the excess of high-frequency derived *Ddc* alleles. If so, we need to account for the low level of variability in the 3′ flanking region. The HKA test does not detect heterogeneity in different regions (intron, exon, and 3′ flanking), but the McDonald tests do. The K/π ratio at silent sites ranges considerably, from 6.4 to 14.0 to 35.6, pointing to the possibility of directional selection at the 3′ flanking end, or further downstream. It seems plausible that this region has been affected by directional selection and that nearby closely linked alleles hitchhiked to high frequency (or even became fixed), as indicated by the H-test.

Adaptive changes may have noticeable effects at the DNA level on linked genes, as a consequence of selective sweeps which reduce nucleotide variation and, in particular, haplotype diversity (Hudson *et al*. 1994, 1997; Depaulis *et al*. 1999; Nurminsky *et al*. 2001; Parsch *et al*. 2001; Sáez *et al*. 2003). Another distinctive feature of selective sweeps is the skew of the frequency spectrum of polymorphic sites toward rare variants (Tsaur *et al*. 1998; Fay and Wu 2000). Such distortion has been observed by Tsaur *et al*. (1998) in *Acp26Aa* in the absence of noticeable reduction of nucleotide variation. Theoretical work by Kim and Stephan (2000) has shown that the interaction of positive and background selection slows down the reduction of variation due to hitchhiking, compared to the situation when hitchhiking occurs in the absence of background selection. Additional sampling of *Ddc* and the neighbour downstream genomic regions, in particular from Africa (see Andolfatto and Przeworski 2001), might help to sort out the forms of natural selection that have shaped nucleotide variation in the *Ddc* region.

The pattern of genetic variation at the *Ddc* locus suggests that it was shaped by hitchhiking due to directional selection acting on a site downstream of the locus: there is an excess of high-frequency-derived alleles, a distinctive feature of hitchhiking; there is a reduction of genetic variation in the 3′ untranslated region, which is not accompanied by the corresponding reduction of interspecific divergence, as would be expected from directional selection on a site within the region or closely downstream; significant nonrandom association between pairs of polymorphic sites indirectly suggests past selection. Absence of prominent haplotype structure and the failure of several neutrality tests to discern selection, indicate that most consequences of selection have been obliterated since its occurrence. The McDonald–Kreitman test indicates that negative selection has also played a role in shaping variation at *Ddc*. Further evidence of negative selection is the low K_a/K_s ratio for the gene, 0.008. Kim and Stephan (2000) have shown that the reduction of variation is smaller when hitchhiking is combined with background selection than when there is no background selection. The limited reduction of variation, relative to the level of interspecific divergence in the 3′-flanking region could be a result of such interaction. Investigation of DNA polymorphisms downstream from *Ddc*, including the closely linked *Cs* and *amd* genes, might pinpoint the site of directional selection.

Acknowledgements

We are grateful to Dr Montserrat Aguadé for DNA samples, and the Mid-America Stock Center in Bowling Green for *D. melanogaster* strains. Dr Martina Žurovcová kindly provided the data for *Idgf3*. Jan Chochola modified the source code of Lewontin's program so that output is appended to a file instead of directly printed. M. Žurovcová and A. G. Sáez commented on a draft of the manuscript.

References

Aguadé M. 1998 Different forces drive the evolution of the *Acp26Aa* and *Acp26Ab* accessory gland genes in the *D. melanogaster* species complex. *Genetics* **150**, 1079–1089.

Andolfatto P. and Kreitman M. 2000 Molecular variation at the *In(2L)t* proximal breakpoint site in natural populations of *D. melanogaster* and *D. simulans*. *Genetics* **154**, 1681–1691.

Andolfatto P. and Przeworski M. 2001 Regions of lower crossing over harbor more rare variants in African populations of *D. melanogaster*. *Genetics* **158**, 657–865.

Aquadro C. F., Jennings R. M. Jr, Bland M. M., Laurie C. C. and Langley C. H. 1992 Patterns of naturally occurring restriction map variation, dopa decarboxylase activity variation and linkage disequilibrium in the *Ddc* gene region of *Drosophila melanogaster*. *Genetics* **132**, 443–452.

Aquadro C. F., DuMont V. B. and Reed F. A. 2001 Genome-wide variation in the human and fruitfly: a comparison. *Curr. Opin. Genet. Dev*. **11**, 627–634.

Begun D. J. and Aquadro C. F. 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster. Nature* **356**, 519–520.

Begun D. J. and Aquadro C. F. 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550.

Begun D. J. and Aquadro C. F. 1995 Molecular variation at the *vermilion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. *Genetics* **140**, 1019–1032.

De Luca M., Roshina N. V., Geiger-Thornsberry G. L., Lyman R. F., Pasyukova E. G. and Mackay T. F. C. 2003 Dopa decarboxylase (*Ddc*) affects variation in *Drosophila* longevity. *Nature Genet*. **34**, 429–433.

Depaulis F., Brazier L. and Veuille M. 1999 Selective sweep at the *Drosophila melanogaster Suppressor of Hairless* locus and its association with the *In(2L)t* inversion polymorphism. *Genetics* **152**, 1017–1024.

Eveleth D. D., Gietz R. D., Spencer C. A., Nargang F. E., Hodgetts R. B. and Marsh J. L. 1986 Sequence and structure of the dopa decarboxylase gene of *Drosophila*: evidence for novel RNA splicing variants. *EMBO J*. **5**, 2663–2672.

Fay J. C. and Wu C-I. 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413.

Fu Y-X. and Li W-H. 1993 Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

Hey J. and Wakeley J. 1997 A coalescent estimator of the population recombination rate. *Genetics* **145**, 833–846.

Hudson R. R. 1987 Estimating the recombination parameter of a finite population model without selection. *Genet. Res*. **50**, 245–250.

Hudson R. R. and Kaplan N. L. 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.

Hudson R. R., Kreitman M. and Aguadé M. 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Hudson R. R., Boos D. D. and Kaplan N. L. 1992 A statistical test for detecting geographic subdivision. *Mol. Biol. Evol*. **9**, 138–151.

Hudson R. R., Bailey K., Skarecky D., Kwiatowski J. and Ayala F. J. 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* **136**, 1329–1340.

Hudson R. R., Sáez A. G. and Ayala F. J. 1997 DNA variation at the *Sod* locus of *Drosophila melanogaster*: an unfolding story of natural selection. *Proc. Natl. Acad. Sci. USA* **94**, 7725–7729.

Kim Y. and Stephan W. 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics* **155**, 1415–1427.

Kliman R. M. and Hey J. 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster. Mol. Biol. Evol*. **10**, 1239–1258.

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

Lazzaro B. P. and Clark A. G. 2003 Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster. Mol. Biol. Evol*. **20**, 914–923.

Lewontin R. C. 1995 The detection of linkage disequilibrium in molecular sequence data. *Genetics* **140**, 377–388.

McDonald J. H. 1996 Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol*. **13**, 253–260.

McDonald J. H. 1998 Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol*. **15**, 377–384.

McDonald J. H. and Kreitman M. 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila. Nature* **351**, 652–654.

Nurminsky D., Aguiar D. D., Bustamante C. D. and Hartl D. L. 2001 Chromosomal effects of rapid gene evolution in *Drosophila melanogaster*. Science **291**, 128–130.

Parsch J., Meiklejohn C. A. and Hartl D. L. 2001 Patterns of DNA sequence variation suggest the recent action of positive selection in the *janus-ocnus* region of *Drosophila simulans*. *Genetics* **159**, 647–657.

Przeworski M. 2002 The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189.

Przeworski M., Wall J. D. and Andolfatto P. 2001 Recombination and the frequency spectrum in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol*. **18**, 291–298.

Rozas J. and Rozas R. 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.

Sáez A. G., Tatarenkov A., Barrio E., Becerra N. H. and Ayala F. J. 2003 Patterns of DNA sequence polymorphism at *Sod* vicinities in *Drosophila melanogaster*: unraveling the footprint of a recent selective sweep. *Proc. Natl. Acad. Sci. USA* **100**, 1793–1798.

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

Tsaur S-C., Ting C-T. and Wu C-I. 1998 Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*: II. Divergence versus polymorphism. *Mol. Biol. Evol*. **15**, 1040–1046.

Watterson G. A. 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol*. **10**, 256–276.

Williams D. A. 1976 Improved likelihood ratio tests for complete contingency tables. *Biometrika* **63**, 33–37.

Wright T. R. F. 1996 Phenotypic analysis of the dopa decarboxylase gene cluster mutants in *Drosophila melanogaster*. *J. Hered*. **87**, 175–190.

Zurovcova M. and Ayala F. J. 2002 Polymorphism patterns in two tightly linked developmental genes, *Idgf1* and *Idgf3*, of *Drosophila melanogaster. Genetics* **162**, 177–188.

Received 13 October 2006