Title
Nucleotide Variation in the tinman and bagpipe homeobox genes of Drosophila melanogaster

Journal
Genetics, 166(4)

ISSN
0016-6731

Authors
Balakirev, E S
Ayala, F J

Publication Date
2004-04-01

Peer reviewed

Nucleotide Variation in the *tinman* and *bagpipe* Homeobox Genes of *Drosophila melanogaster*

Evgeniy S. Balakirev*,†,‡ and Francisco J. Ayala*,[1]

*Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525, †Institute of Marine Biology, Vladivostok 690041, Russia and ‡Academy of Ecology, Marine Biology, and Biotechnology, Far Eastern State University, Vladivostok 690600, Russia

Manuscript received August 15, 2003
Accepted for publication January 9, 2004

ABSTRACT

The *tinman* (*tin*) and *bagpipe* (*bap*) genes are members of the NK homeobox gene family of Drosophila, so that *tin* occupies a higher position than *bap* in the regulatory hierarchy. Little is known about the level and pattern of genetic polymorphism in homeobox genes. We have analyzed nucleotide polymorphism in 27 strains of *Drosophila melanogaster* and one each of *D. simulans* and *D. sechellia*, within two closely linked regions encompassing a partial sequence of *tin* and the complete sequence of *bap*. The two genes exhibit different levels and patterns of nucleotide diversity. Two sets of sharply divergent sequence types are detected for *tin*. The haplotype structure of *bap* is more complex: about half of the sequences are identical (or virtually so), while the rest are fairly heterogeneous. The level of silent nucleotide variability is 0.0063 for *tin* but significantly higher, 0.0141, for *bap*, a level of polymorphism comparable to the most polymorphic structural genes of *D. melanogaster*. Recombination rate and gene conversion are also higher for *bap* than for *tin*. There is strong linkage disequilibrium, with the highest values in the introns of both genes and exon II of *bap*. The patterns of polymorphism in *tin* and *bap* are not compatible with an equilibrium model of selective neutrality. We suggest that negative selection and demographic history are the major factors shaping the pattern of nucleotide polymorphism in the *tin* and *bap* genes; moreover, there are clear indications of positive selection in the *bap* gene.

THE population polymorphism of homeobox genes in Drosophila has received scant attention. Begun *et al.* (1994) investigated restriction-map polymorphism in the *cut* locus region of *D. melanogaster* and *D. simulans*. The amount of nucleotide variation was in both species about one-fourth of the average amount in comparable restriction-map studies of gene regions with normal recombination (Aquadro 1992), even though the gene is located in a genomic region with "normal" recombination rate. Begun *et al.* (1994) suggested that selective sweeps associated with some closely linked gene(s) might account for the decreased variability of *cut*. Baines *et al.* (2002) investigated nucleotide polymorphism in the *D. melanogaster bicoid* region. The level of silent polymorphism for the noncoding region was lower than typical values observed in *D. melanogaster* (Moriyama and Powell 1996); silent diversity in the coding region was also low ($\pi = 0.0002$). An interesting feature of the *bicoid* coding region variability was that 6 of 7 polymorphic sites involved replacement polymorphisms, which could indicate a relaxation of selective constraints in this region (Baines *et al.* 2002). A significant excess of intraspecific replacement polymorphism (16 of 21) was also observed in the *Arabidopsis thaliana CAULIFLOWER* homeotic gene, which was in this case attributed to positive selection (Purugganan and Suddith 1998). Nucleotide polymorphism has been investigated also for the *even-skipped* (Ludwig and Kreitman 1995), *Ultrabithorax* (Gibson and Hogness 1996), and *OdysseusH* (Ting *et al.* 1998) Drosophila homeobox genes. The data obtained for these genes, however, involved only interspecific comparisons or only very few sequences, so that intraspecific variability could not be estimated. On the average, Drosophila regulatory loci display reduced levels of diversity (less than half) compared to structural genes (Moriyama and Powell 1996). In contrast, the level of polymorphism is about the same in some plant regulatory genes as in structural loci (Purugganan and Suddith 1999; Purugganan 2000). However, recent data indicate that even adjacent regulatory genes can show a wide range in the level and patterns of sequence variation, which suggests that different members of a regulatory gene cluster may be subject to distinct evolutionary forces (Lawton-Rauh *et al.* 2003; Shepard and Purugganan 2003; see also Purugganan and Suddith 1999).

Previously, we have investigated nucleotide variability in the β-*esterase* gene cluster located on the left arm of chromosome 3 of *Drosophila melanogaster*, at 68F7–69A1 in the cytogenetic map (Balakirev and Ayala 1996, 2003a,b, 2004; Balakirev *et al.* 1999, 2002, 2003; Ayala

[1]*Corresponding author:* Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697-2525. E-mail: fjayala@uci.edu

et al. 2002). The cluster comprises two tandemly duplicated genes, *Est-6* and ψ*Est-6* (originally named *Est-P* by Collet *et al.* 1990). We detected complex haplotype structures in both genes. We now present a population genetic analysis of the *tinman* (*tin*) and *bagpipe* (*bap*) homeobox genes in 27 strains of *D. melanogaster*.

The *tin* and *bap* genes are members of the NK homeobox gene family, which consists of several closely linked interacting regulatory genes, located on the right arm of *D. melanogaster* chromosome 3, at 93DE in the cytogenetic map (Kim and Nirenberg 1989; review in Jagla *et al.* 2001). The common cosmopolitan inversion *In(3R)* (89C2.3;96A1.19; Lemeunier and Aulard 1992) encloses the *tin* and *bap* genes, but no third-chromosome inversions have been found segregating in the El Rio population studied in the present investigation (Smit-McBride *et al.* 1988 and unpublished data from our laboratory). There are two exons (444 and 705 bp) in *bap*, but three (186, 609, and 453 bp) in *tin*. Correspondingly, there is one intron in *bap* (121 bp) and two (905 and 430 bp) in *tin*. The two genes are closely linked. The transcription start site of *bap* is 7.1 kb downstream of the 3′-flanking end of the *tin* gene. No open reading frame has been recorded in the region between the two genes (Adams *et al.* 2000). Both genes are transcribed in the same direction. There is a hierarchical relationship between the two genes since *bap* activation in the dorsal mesoderm depends on *tin*, while *tin* does not require *bap* (Azpiazu and Frasch 1993; Bodmer 1993). We have sequenced 3818 bp, which include the 5′-flanking region of *tin* (767 bp), the partial *tin* gene (exon I, 186 bp; intron I, 905 bp; and partial exon II, 570 bp), the 5′-flanking region of *bap* (64 bp), the complete *bap* gene (exon I, 444 bp; intron, 121 bp; and exon II, 705 bp), and the 3′-flanking region of *bap* (56 bp).

MATERIALS AND METHODS

Drosophila strains: The 27 *D. melanogaster* strains derive from a random sample of wild flies collected by F. J. Ayala (October 1991) in El Rio Vineyard, Acampo, California. The strains were made fully homozygous for the third chromosome by crosses with balancer stocks, as described by Seager and Ayala (1982). The 27 strains were previously investigated for β-*esterase* gene cluster (see above). Strains 1–27 successively correspond to the following strains in Balakirev *et al.* (2003): S-114S, F-517S, F-517F, F-531F, F-1461S, S-2588S, S-581F, S-255S, F-357F, F-611F, S-174F, S-501F, S-501S, S-1224F, S-377F, F-274F, S-512F, F-96S, US-255F, S-968F, S-521S, S-549S, F-775F, S-438S, S-26F, S-510S, and S-565F.

DNA extraction, amplification, and sequencing: Total genomic DNA was extracted using the tissue protocol of the QIAamp tissue kit (QIAGEN, Valencia, CA). The published Celera genome sequence of *D. melanogaster* was used for designing PCR and sequencing primers (Adams *et al.* 2000). The primers used for the *tin* PCR amplification reactions were 5′-acaaacg taaatacatggggactat-3′ (forward primer) and 5′-attgacacttaacg caaggaaacag-3′ (reverse primer). The primers used for the *bap* PCR amplification reactions were 5′-atcaatagaaatccgtagt gaac-3′ (forward primer) and 5′-ttacacatagaggctaaaacac-3′ (reverse primer). The PCR reactions were carried out in final volumes of 50 μl, using TaKaRa Ex Taq in accordance with the manufacturer's description (Takara Biotechnology, Berkeley, CA). The reaction mixtures were overlaid with mineral oil; placed in a DNA thermal cycler (Perkin-Elmer-Cetus, Norwalk, CT); incubated 5 min at 95°; and subjected to 30 cycles of denaturation, annealing, and extension: 95° for 30 sec, 60° for 30 sec, and 72° for 2.0 min (for the first cycle and progressively adding 3 sec at 72° for every subsequent cycle), with a final 7-min extension period at 72°. The PCR reactions were purified with the Wizard PCR preps DNA purification system (Promega, Madison, WI), directly sequenced by the dideoxy chain-termination technique using Dye Terminator chemistry, and separated with the ABI PRISM 377 automated DNA sequencer (Perkin-Elmer). For each strain, both strands were sequenced using six overlapping internal primers spaced, on average, 350 nucleotides apart. At least two independent PCR amplifications were sequenced for each polymorphic site in all *D. melanogaster* strains to prevent possible PCR or sequencing errors. The GenBank accession numbers for the sequences are AY368077–AY368109 and AY369088–AY369115.

DNA sequence analysis: Multiple alignment was carried out using the program CLUSTAL W (Thompson *et al.* 1994). Linkage disequilibrium between polymorphic sites was evaluated using Fisher's exact test of independence. The computer programs DnaSP, version 3.4 (Rozas and Rozas 1999) and PROSEQ, version 2.4 (Filatov and Charlesworth 1999) were used to analyze the data by means of the "sliding-window" method (Hudson and Kaplan 1988) and for most intraspecific analyses. Departures from neutral expectations were investigated using the tests of Hudson *et al.* [1987; Hudson-Kreitman-Aguadé (HKA) test], Hudson *et al.* (1994; haplotype test), McDonald (1996, 1998), Kelly (1997), and Wall (1999). The permutation approach of Hudson *et al.* (1992) was used to estimate the significance of sequence differences between haplotype families. The coalescent simulations (Hudson 1983, 1990) were performed with the PROSEQ program to estimate the probabilities of the observed values of Kelly's Z_{nS} and Wall's B and Q statistics and confidence intervals for the nucleotide diversity values. The method of Sawyer (1989, 1999) was used to analyze intra- and intergenic conversion events.

RESULTS

Nucleotide polymorphism and recombination: Figure 1 shows a total of 69 polymorphic sites in a sample of 27 sequences of the *tin* and *bap* genes. There are five length polymorphisms: four in intron I of *tin* and one in exon II of *bap*. Some relevant statistics are given in Table 1. The π value for the full sequence is 0.0056, which is within the range of values observed in highly recombining gene regions in *D. melanogaster* (Moriyama and Powell 1996; Powell 1997). The silent π value is 0.0063 for *tin* but significantly higher (by coalescent simulation), 0.0140, for *bap*. The level of synonymous variation is 0.0052 in the *tin* coding region but three times higher, 0.0167, in the *bap* coding region (the difference is highly significant by coalescent simulation even without recombination, $P = 0.01$). The difference is less pronounced for nonsynonymous variation, which is 0.0011 in the *tin* gene and two times higher, 0.0021, in the *bap* gene ($P > 0.05$). Silent variability is three times higher in exon II ($\pi = 0.0224$) than in exon I ($\pi =$

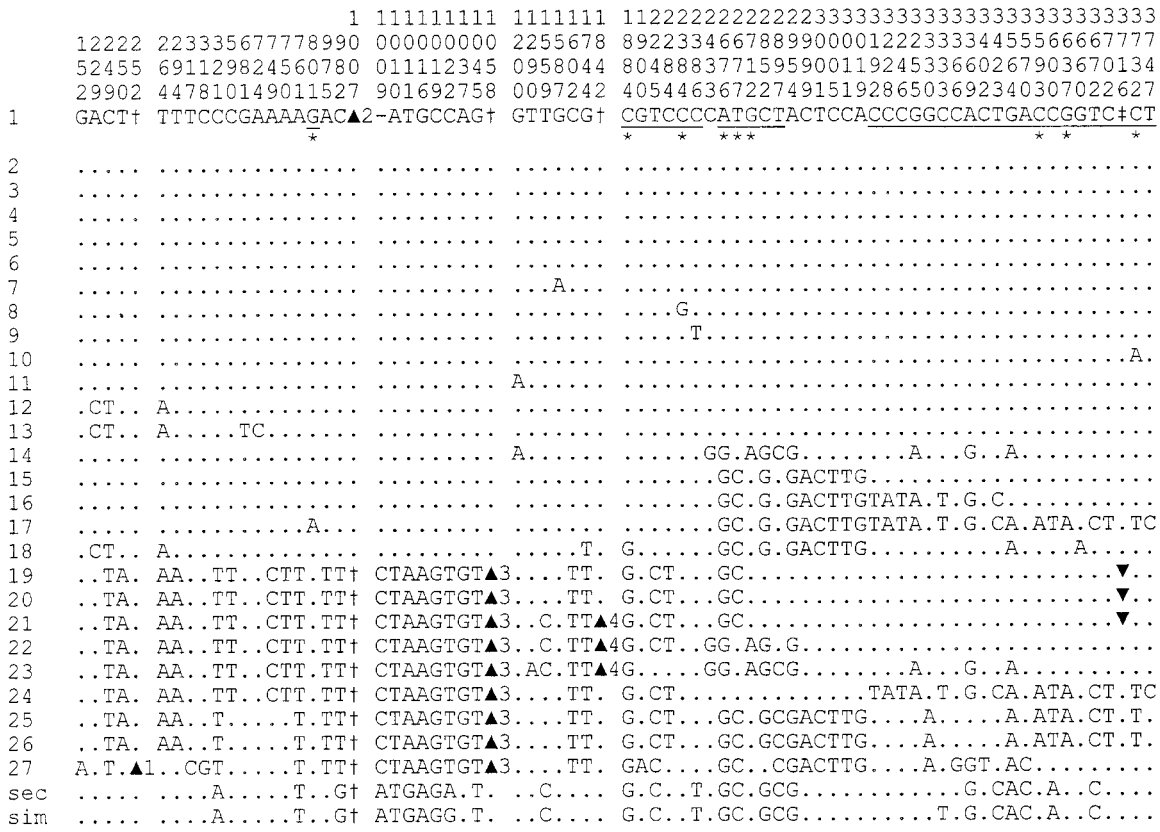

FIGURE 1.—DNA polymorphism in the *tin* and *bap* genes of 27 strains of *Drosophila melanogaster*. The lines are numbered consecutively according to their genetic similarity. The numbers above the top sequence represent the position of segregating sites and the start of a deletion or insertion. Nucleotides are numbered from the beginning of our sequence, position 155034 for the *tin* and 165505 for the *bap* gene in ADAMS *et al.* (2000). The coding regions of the genes are underlined below the top reference sequence. Amino acid replacement polymorphisms are marked with asterisks. Dots indicate the same nucleotide as the reference sequence. A hyphen represents deleted nucleotides. ▲, a deletion; †,the absence of a deletion. ▲1, a 2-bp deletion of TT (position 252–253); ▲2, a 3-bp deletion of TCC (position 1007–1009); ▲3, a 101-bp deletion of TACCAAAATCTACTAAAT CGAATCAGTGTACATATATGATATAAAACTAATTATTATACCGGTTAACTATCTAAATTGCAATGATTTAAGAAGGTATCTGT (position 1058–1158); ▲4, a 9-bp deletion of TCCTGCTTT (position 1842–1850). ▼, a 12-bp insertion of GAGCGGAGAGCG (position 3716–3727); ‡, the absence of an insertion. The coordinates for the functional regions of the genes are: 1–767 (*tin*, 5′-flanking region), 768–953 (*tin*, exon I), 954–1858 (*tin*, intron I), 1859–2428 (*tin*, exon II), 2429–2492 (*bap*, 5′-flanking region), 2493–2936 (*bap*, exon I), 2937–3057 (*bap*, intron), 3058–3774 (*bap*, exon II), and 3775–3830 (*bap*, 3′flanking region). sec, *D. sechellia*; sim, *D. simulans*.

0.0073) of *bap*; the difference is statistically significant by coalescent simulation, even without recombination ($P = 0.01$). The elevated level of nucleotide variability within the *bap* exon II cannot be accounted for by relaxation of functional constraints. This is because the *bap* exon II contains a functionally important homeodomain (183 bp at position 3136–3318, our coordinates), which is conserved within a wide phylogenetic scale (DUBOULE 1994). There are four polymorphic sites within this *bap* homeodomain region (Figure 1, positions 3192, 3228, 3246, and 3255), with $\pi = 0.0051$, which is slightly lower than that for the whole *bap* exon II ($\pi = 0.0063$).

The silent variability of *bap* ($\pi = 0.0140$) is close to that of one of the most polymorphic genes of *D. melanogaster*, *Est-6* ($\pi = 0.0156$), in the same North American population (BALAKIREV *et al.* 2002; BALAKIREV and AYALA 2003a). The silent variability of *bap* exon II is very close to that of ψ*Est-6* ($\pi = 0.0253$), which is almost twice as variable as *Est-6* (BALAKIREV *et al.* 2003; BALAKIREV and AYALA 2004). The nucleotide variability detected in the *tin* gene is in the range of values observed in many other regulatory genes evolving under negative selection (MORIYAMA and POWELL 1996; PURUGGANAN 2000; RILEY *et al.* 2003). The differences in the level of variability between *tin* and *bap* could be due to positive selection reflecting different degrees of functional constraint (see below).

The level of divergence, *K*, between *D. melanogaster* and *D. simulans* is similar for *tin* and *bap* (Table 1). The K/π ratio is, however, substantially higher for *tin* in both the coding and noncoding regions: K_{cod}/π_{cod} is 10.67 for *tin*, but 5.29 for *bap* and K_{ncod}/π_{ncod} is 9.28 for *tin*, but 3.85 for *bap*. These differences reflect that the pressure of purifying selection is higher for *tin* than for *bap*. Silent divergence is very similar in exon I ($K =$

TABLE 1

Nucleotide diversity and divergence in the *tin* and *bap* genes of *D. melanogaster*

	Syn	Nsyn	Cod	Ncod	Silent	Total
			tin			
N	187	569	756	1557	1744	2313
S	4	3	7	30	34	37
π	0.0052	0.0011	0.0021	0.0065	0.0063	0.0050
θ	0.0056	0.0014	0.0024	0.0050	0.0051	0.0042
$K_{\text{mel-sim}}$	0.0668	0.0083	0.0224	0.0603	0.0610	0.0477
$K_{\text{mel-sec}}$	0.0614	0.0119	0.0238	0.0640	0.0637	0.0506
			bap			
N	275	868	1146	244	519	1390
S	19	6	25	7	26	32
π	0.0167	0.0021	0.0056	0.0110	0.0140	0.0065
θ	0.0189	0.0018	0.0059	0.0074	0.0135	0.0060
$K_{\text{mel-sim}}$	0.1081	0.0065	0.0296	0.0427	0.0767	0.0319
$K_{\text{mel-sec}}$	0.1217	0.0111	0.0362	0.0467	0.0855	0.0459
			tin and *bap*			
N	462	1437	1902	1801	2263	3703
S	23	9	32	37	60	69
π	0.0121	0.0017	0.0042	0.0071	0.0081	0.0056
θ	0.0135	0.0016	0.0045	0.0053	0.0070	0.0048
$K_{\text{mel-sim}}$	0.0911	0.0072	0.0267	0.0579	0.0645	0.0417
$K_{\text{mel-sec}}$	0.0967	0.0114	0.0313	0.0616	0.0686	0.0459

N, number of sites (indels are excluded); S, number of polymorphic sites; π, average number of nucleotide differences per site among all pairs of sequences (Nei 1987, p. 256); θ, average number of segregating nucleotide sites among all sequences based on the expected distribution of neutral variants in a panmictic population at equilibrium (Watterson 1975); $K_{\text{mel-sim}}$ and $K_{\text{mel-sec}}$ are the proportion of nucleotide differences between *D. melanogaster* and *D. simulans,* and between *D. melanogaster* and *D. sechellia,* respectively, corrected according to Jukes and Cantor (1969); Syn, synonymous sites; Nsyn, nonsynonymous sites; Cod, coding (exon) regions; Ncod, noncoding (intronic and flanking) regions; Silent, silent sites (synonymous and noncoding sites).

0.0988) and exon II ($K = 0.1012$) of *bap*, despite the fact that the level of silent variability is significantly different in these coding regions, as noted above. The level of divergence between *D. melanogaster* and *D. sechellia* is very close to the level of divergence between *D. melanogaster* and *D. simulans* (Table 1); the only difference concerns nonsynonymous divergence, which is 1.4 and 1.7 times higher between *D. melanogaster* and *D. sechellia* than between *D. melanogaster* and *D. simulans* for the *tin* and *bap* genes, respectively.

The method of Hudson and Kaplan (1985) reveals a minimum of nine recombination events in the whole region analyzed. The minimum number of recombination events is six for *bap* but two for *tin*. There is a large difference (33 times) in the recombination estimator (ρ, McVean *et al.* 2002) for *tin* ($\rho = 0.0008$) and *bap* ($\rho = 0.0264$) (Table 2). For the chromosomal region 93DE the laboratory estimate of recombination rate (C_{lab}; Comeron *et al.* 1999) based on the physical and genetic maps of *D. melanogaster* is 0.0744. A substantial difference between the laboratory-based and sequence-based estimates of recombination has been explained by a recent bottleneck (for review, see Wall 2001). The large difference between the sequence-based estimates of recombination rate for the closely linked *tin* and *bap* genes (7.1 kb apart) is inconsistent with the demographic explanation of the difference between the laboratory-based and sequence-based estimates of recombination rate (Andolfatto and Przeworski 2000; Wall 2001).

The method of Sawyer (1989, 1999) detects gene conversion events within both *tin* ($P = 0.0097$) and *bap* ($P = 0.0029$). The number of significant fragments is 7 for *tin* (fragment length varies from 1149 to 1581 bp, average 1396 bp) but 49 for *bap* (fragment length varies from 323 to 1135 bp, average 665 bp). Intragenic conversion involves two *tin* regions (coordinates 751–1899 and 319–1899) in 6 sequences; for *bap*, intragenic conversion involves eight regions (coordinates 1–763, 1–1135, 250–763, 592–1135, 592–1303, 592–1402, 996–1318, and 996–1402) in 23 sequences. Thus, intragenic gene conversion is more pronounced within *bap* than within *tin*. Intergenic gene conversion is not detected, which is likely due to the low nucleotide similarity between *tin* and *bap* (42.7%), which is insufficient to satisfy the homology requirements for efficient intergenic conversion. The recombination machinery is sensitive even to a single-nucleotide mismatch; individual nucleotide substitutions have been shown to affect recombination in many organisms (for review, see Balakirev and Ayala 2003c).

Haplotype structure: We have previously described

TABLE 2

Recombination estimates

Full sequence			*tin*			*bap*		
Per gene	Per site	ρ/θ	Per gene	Per site	ρ/θ	Per gene	Per site	ρ/θ
27.655	0.0072	1.5	2.004	0.0008	0.1905	37.074	0.0264	4.2581

ρ estimates the population recombination rate, $4N_e r$ (N_e is the effective population size and r is the recombination rate per nucleotide site per generation), obtained by the method of McVean *et al.* (2002). All site types are included in the recombination analysis.

the presence of two sets of haplotypes for the β-*esterase* gene cluster of *D. melanogaster*, localized on the left arm of the third chromosome (Balakirev *et al.* 1999, 2002, 2003). There is also strong haplotype structure for the *tin* and *bap* genes (Figures 1–3). For the *tin* gene, there are two sets of haplotypes, each completely associated with one of two deletions, 3 bp and 101 bp, within intron I (Figure 1, ▲2 and ▲3) and almost completely associated with the replacement polymorphism at position 1884 (Figure 1). The ▲2 deletion exists also in *D. simulans* and *D. sechellia* (Figure 1), which suggests that it is the ancestral condition, whereas the ▲3 deletion has appeared after the split of *D. melanogaster* from the other two species. Strong haplotype structure is also observed for the *bap* gene (Figures 1 and 3). There is a set of 13 sequences, 12 of them identical to each other plus one (no. 10) that differs by a nonsynonymous substitution, and a second set of fairly heterogeneous sequences (nos. 14–27). The difference between the two *tin* haplotype sets (1–18 *vs.* 19–27) is highly significant ($K_{st}^* = 0.4692$; $K_{st}^*{}_{0.999} = 0.1580$, $P < 0.001$ by the permutation test of Hudson *et al.* 1992). However, the level of variability is not significantly different between the two *tin* haplotype sets ($\pi = 0.009$ *vs.* $\pi = 0.0019$).

The homogeneous haplotype set of *bap* (strains 1–13, Figures 1 and 3) has scant variability ($\pi = 0.0001$).

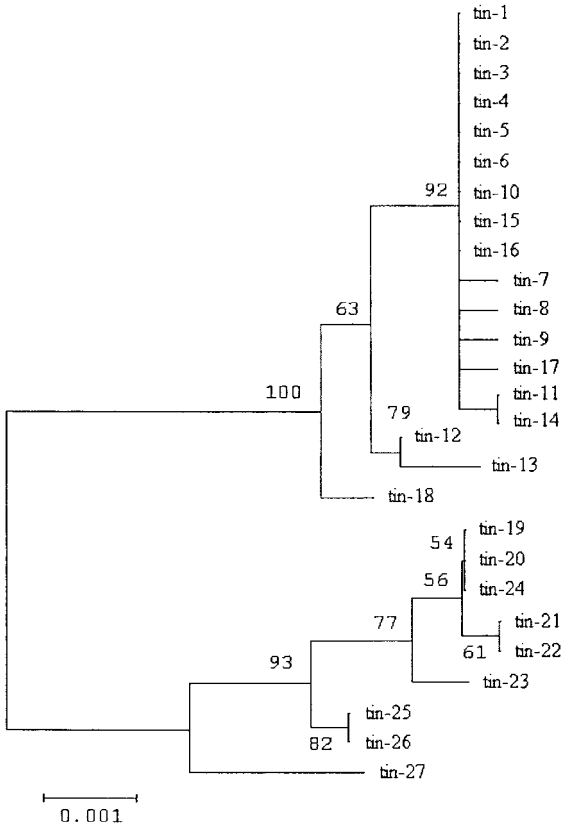

Figure 2.—Neighbor-joining tree of the *tin* haplotypes of *D. melanogaster*, based on Kimura's two-parameter distance. The numbers at the nodes are bootstrap percentage of probability values based on 10,000 replications.

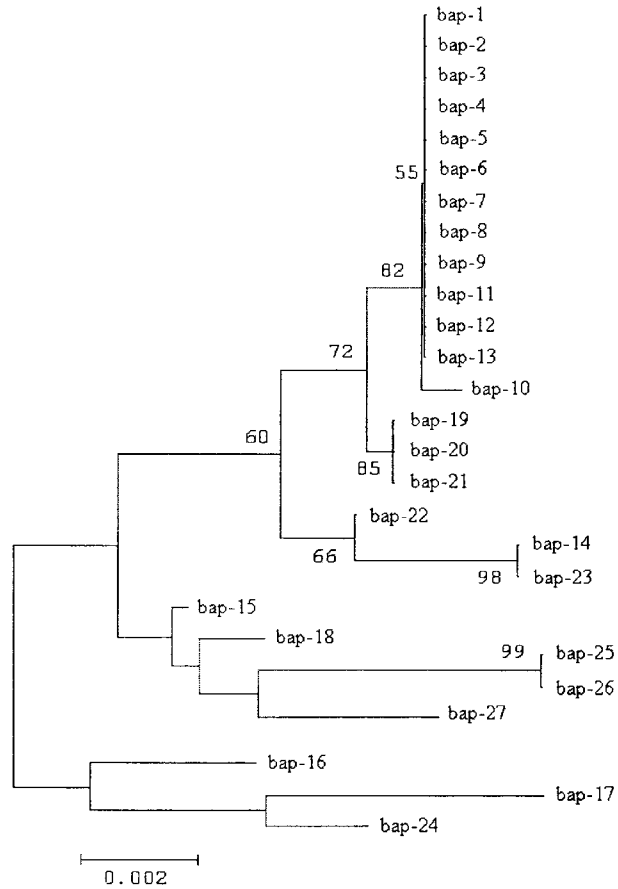

Figure 3.—Neighbor-joining tree of the *bap* haplotypes of *D. melanogaster* based on Kimura's two-parameter distance. The numbers at the nodes are bootstrap percentage of probability values based on 10,000 replications.

The heterogeneous set (strains 14–27) is significantly different from the first set ($K_{st}^* = 0.2852$; $K_{st}^*{}_{0.999} = 0.1002$, $P < 0.001$ by the permutation test) and has significantly greater variability ($\pi = 0.0089$; $P < 0.0001$). The distribution of mutations in the *bap* gene is highly asymmetrical in the two haplotype sets: the heterogeneous set is 74 times more variable than the homogeneous haplotype set.

Sliding-window analysis: Figure 4 shows sliding-window plots of the distribution of nucleotide polymorphism in *D. melanogaster* (Figure 4A), divergence between this species and *D. simulans* (Figure 4B), polymorphism-to-divergence ratio (Figure 4C), and linkage disequilibrium (Figure 4D). There are two distinct peaks of nucleotide variability in *tin* (Figure 4A), in the 5′-flanking region (midpoint coordinates 271–280) and intron (midpoint coordinates 992–1027). These peaks coincide with the peaks of divergence (Figure 4B) and linkage disequilibrium (Figure 4D). The highest peak of the polymorphism-to-divergence ratio within the *tin* gene, however, does not coincide with the peaks of variability and divergence (it is centered on the *tin* exon I, Figure 4C). There is also a peak of variability in the *bap* gene (Figure 4A) in the intron region (midpoint coordinates 2969–3003). This peak is not accompanied by a peak of divergence (Figure 4B); rather, there is an obvious decrease of divergence in the intron region of *bap*, which produces a peak in the silent divergence-to-polymorphism ratio (Figure 4C, midpoint coordinates 2700–2900). The peak in the *bap* intron coincides with a peak of linkage disequilibrium (Figure 4D). The highest peak of the silent divergence-to-polymorphism ratio is centered on the *bap* exon II (Figure 4C, midpoint coordinate 3200). The minimal values of polymorphism-to-divergence ratio within the *bap* gene are at the beginning of exon I and centered on the two replacement polymorphic sites (Figures 4C and 5, positions 2676 and 2677). The lowest and highest polymorphism-to-divergence ratios are accompanied by the largest maximum and average sliding *G* values of the McDonald's (1996, 1998) tests (Figure 6).

The variants within the intron sequences of both genes segregate as single-haplotype blocks (Figure 1). Correspondingly there are obvious peaks of nucleotide variability and linkage disequilibrium in the introns (Figure 4, A and D). For the *tin* gene the intron peak of nucleotide variability also corresponds to the high peak of divergence but for the *bap* gene there is an opposite tendency: the peak of variability corresponds to the lowest level of the divergence (Figure 4, A and B). A similar tendency is observed for the *bap* exon II: the level of variability and divergence is close in this region (Figure 4B). Thus for the *tin* gene the amount of divergence between species is in accordance with the amount of intraspecific polymorphism but for the *bap*

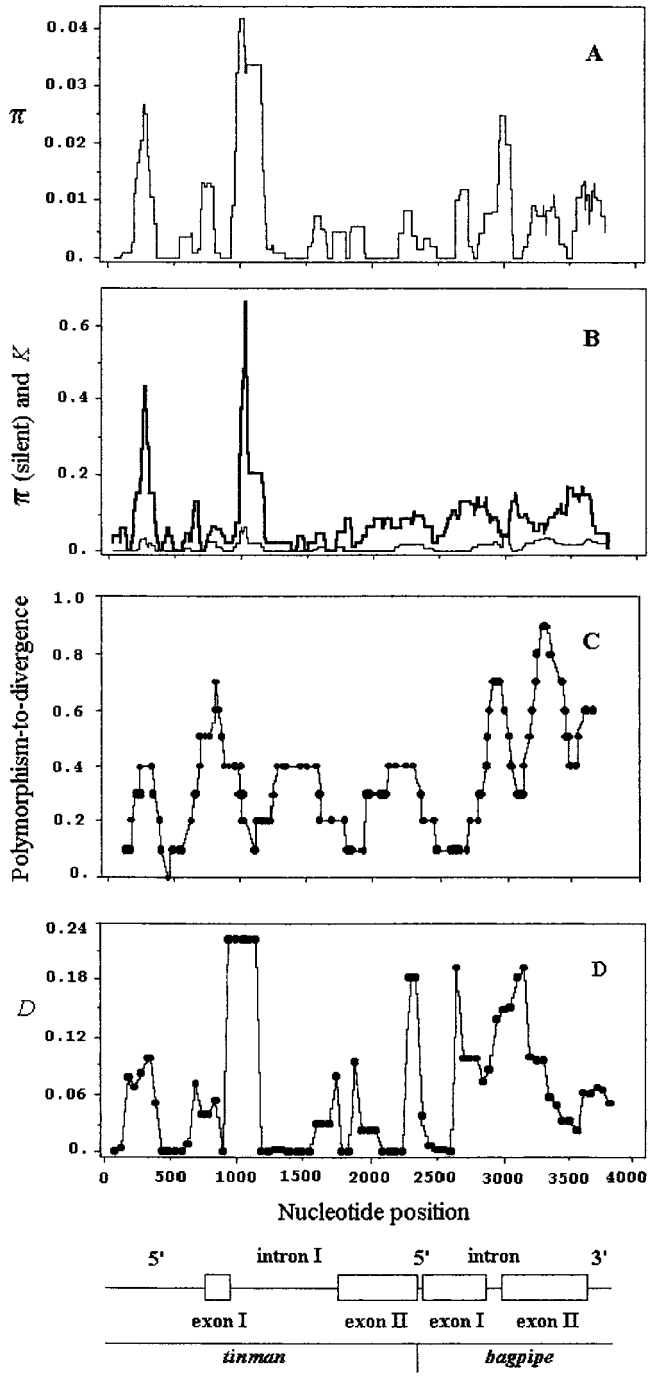

FIGURE 4.—Sliding-window plots of (A) nucleotide diversity (measured by π), (B) silent nucleotide diversity (thin line) and divergence (*K*, thick line), (C) polymorphism-to-divergence ratio, and (D) linkage disequilibrium (measured by *D*) along the *tin* and *bap* genes of *D. melanogaster*. *K* is the average number of nucleotide substitutions per site between *D. melanogaster* and *D. simulans*. Window sizes are 100 nucleotides with 1-nucleotide increments for A, 50 nucleotides with 1-nucleotide increments for B, and 200 nucleotides with 50-nucleotide increments for D. Window size is 10 variable substitutions for the sliding-window plot of polymorphism-to-divergence ratio (C). A schematic of the region investigated is displayed at bottom. Exons are indicated by boxes; the introns and the 5′- and 3′-flanking regions are shown by thin lines.

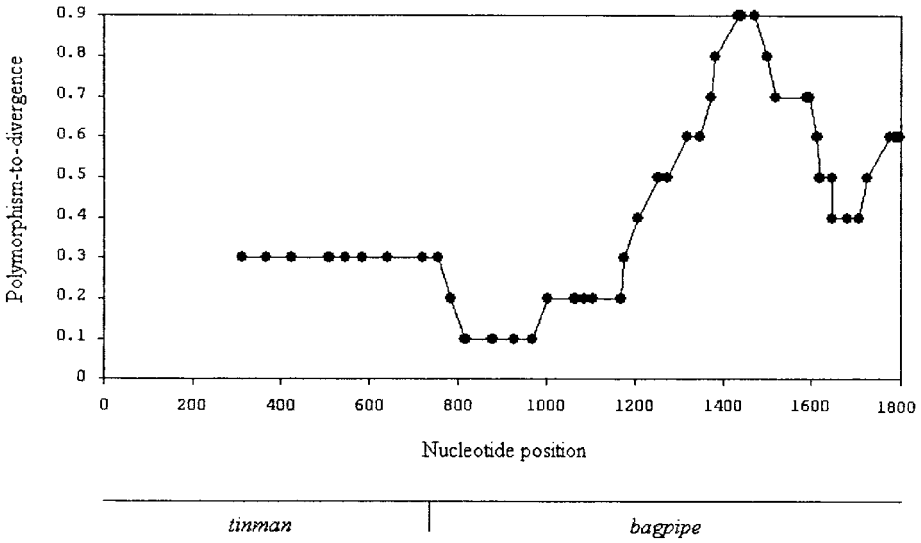

FIGURE 5.—Sliding-window plot of synonymous polymorphism-to-divergence ratio in the *tin* and *bap* genes of *D. melanogaster*. A vertical line indicates the separation between *tin* and *bap*. *D. simulans* is used as an outgroup species. Window size is 10 variable substitutions.

gene the pattern is different: a decrease of the divergence and a parallel increase of polymorphism in the intron and exon II regions that could be explained by the influence of positive selection (see below).

We have measured heterogeneity in the distribution of polymorphic sites along the sequence and discordance between the level of within-*melanogaster* polymorphism and the *melanogaster-simulans* (or *melanogaster-sechellia*) divergence by means of GOSS and LEWONTIN's (1996) and MCDONALD's (1996, 1998) statistics and have assessed their significance by Monte Carlo simulations of the coalescent model incorporating recombination (MCDONALD 1996, 1998; Table 3). On the basis of 10,000 simulations, with the recombination parameters varying from 1 to 64, the tests are not significant for the *tin* gene. However, the tests are significant for the

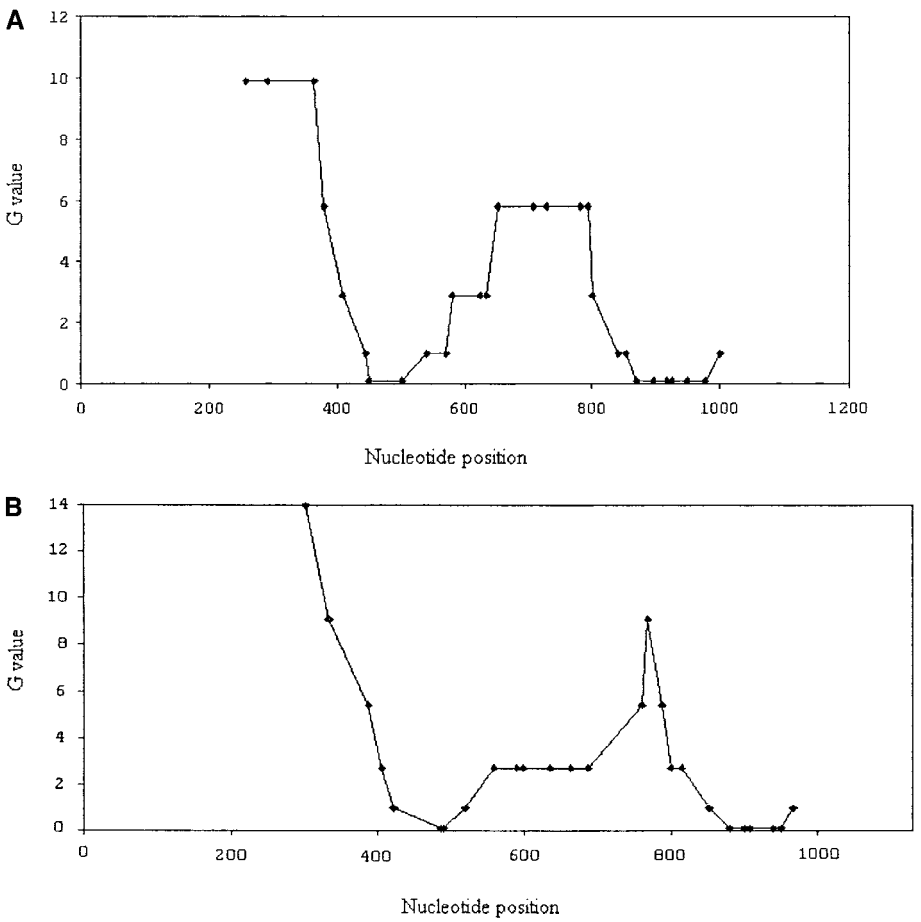

FIGURE 6.—Sliding-window plot of the largest average sliding *G* value (A) and the largest maximum sliding *G* value (B) for the *bap* gene (synonymous variability). Window size is 13 variable substitutions for A and 15 for B.

TABLE 3

Test statistics for the *tin* and *bap* genes of *D. melanogaster*

	G_{max}		Runs		K.-S.		Var		Var_{mod}		G_{avg}	
	Sim	Sec	Sim	Sec	Sim	Sec	Sim	Sec	Sim	Sec	Sim	Sec
tin	10.043	12.467	51	51	0.0436	0.0398	0.0005	0.0006	0.0017	0.0018	1.212	1.504
All sites	(0.328)	(0.219)	(0.187)	(0.148)	(0.235)	(0.302)	(0.269)	(0.254)	(0.328)	(0.328)	(0.694)	(0.607)
tin	9.260	10.738	47	47	0.0406	0.0387	0.0006	0.0006	0.0021	0.0022	1.105	1.411
Silent sites	(0.396)	(0.268)	(0.234)	(0.163)	(0.321)	(0.322)	(0.402)	(0.301)	(0.397)	(0.268)	(0.764)	(0.601)
tin	NA	NA	7	7	0.0918	0.0714	0.0204	0.0062	0.0921	0.0793	NA	NA
Syn sites	–	–	(0.721)	(0.712)	(0.523)	(0.770)	(0.602)	(0.967)	(0.613)	(0.907)	–	–
bap	11.977	12.063	26	32	0.0888	0.0831	0.0010	0.0009	0.0028	0.0026	2.540	2.583
All sites	(0.159)	(0.160)	(0.064)	(0.164)	(0.061)	(0.064)	(0.024)	(0.029)	(0.028)	(0.081)	(0.245)	(0.237)
bap	11.631	18.467	20	20	0.0863	0.0942	0.0015	0.0021	0.0039	0.0046	3.061	3.206
Silent sites	(0.149)	(0.020)	(0.025)	(0.007)	(0.105)	(0.051)	(0.014)	(0.003)	(0.040)	(0.006)	(0.142)	(0.125)
bap	14.024	13.946	14	16	0.1375	0.1302	0.0020	0.0023	0.0064	0.0069	3.081	3.228
Syn sites	(0.040)	(0.048)	(0.028)	(0.038)	(0.010)	(0.008)	(0.073)	(0.053)	(0.121)	(0.057)	(0.108)	(0.093)

G_{max}, Runs, K.-S., Var, Var_{mod}, and G_{avg} are test statistics (see Goss and Lewontin 1996 and McDonald 1996, 1998). *P* values are in parentheses. Significant and marginally significant *P* values are underlined.

bap gene (Table 3). There are two areas within the *bap* gene with the largest average and maximum sliding *G* values (Figure 6, A and B). The first (and most pronounced) area is located at the beginning of *bap* exon I and coincides with a region of low polymorphism-to-divergence ratio (Figure 5). The second area is located in *bap* exon II and coincides with a region of high polymorphism-to-divergence ratio. The region of low polymorphism-to-divergence ratio is centered on the two replacement substitutions (positions 2676 and 2677). The region of high polymorphism-to-divergence ratio is localized within exon II but is not centered on the replacement polymorphism (Figures 1 and 4C). An area of low polymorphism could result from a selective sweep whereas high polymorphism could result from balancing selection (McDonald 1996, 1998). Previously, we have shown that both types of selection are involved in the evolution of the *Est-6* gene of *D. melanogaster* (Balakirev *et al.* 2002). We suggest that both types of selection are involved within the *bap* gene. To examine this suggestion one would analyze separately the polymorphism-to-divergence ratio in the regions with low and high polymorphism (that roughly correspond to exon I and exon II of *bap*), using the McDonald and Kreitman (1991) and/or the HKA test (Hudson *et al.* 1987). However, both tests are hardly applicable in this case because there are only two silent polymorphic sites within exon I of *bap*. Thus, we cannot contrast the pattern of polymorphism-to-divergence in both exons. However, for *bap* exon II the HKA test reveals a higher polymorphism-to-divergence ratio in comparison with the complete *tin* gene ($\chi^2 = 4.484$, $P = 0.034$ if *D. simulans* is used as an outgroup and $\chi^2 = 4.813$, $P = 0.028$ if *D. sechellia* is used as an outgroup), which is in accordance with the possible action of positive selection on this region.

Linkage disequilibrium: For the whole region there are 1378 pairwise comparisons and 514 (37.30%) of them are significant (Figure 7). With the Bonferroni correction, 156 (11.32%) remain significant. There is strong linkage disequilibrium within the *tin* gene: 83.69% (272 out of 325) pairwise comparisons are significant (55.69%, 181 with the Bonferroni correction). The significant associations are due mostly to polymorphic sites located within the *tin* noncoding regions (5′-flanking region and intron); only three exonic polymorphic sites (positions 1884, 2245, and 2284) are involved in significant associations (Figure 7). Within the *bap* gene 48.43% pairwise comparisons (170 out of 351) show statistically significant linkage disequilibrium (9.69%, 34 with the Bonferroni correction). The distribution of significant associations within *bap* is not homogeneous: more than half of the sites involved in significant associations are located within exon II. Between *tin* and *bap* genes only 5.22% (72 out of 1378) of tests are significant (Figure 7, shaded areas); none is significant with the Bonferroni correction.

Within both the *tin* and *bap* genes there are strong associations between intronic sites (Figure 7). Clusters of significant linkage disequilibrium that occur predominantly in introns have been repeatedly observed in Drosophila (Miyashita and Langley 1988; Miyashita *et al.* 1993; Schaeffer and Miller 1993). Kirby *et al.* (1995) showed that the linkage disequilibria clustered within the introns of the *Adh* locus of *D. pseudoobscura* are caused by epistatic selection maintaining the secondary structures of precursor mRNA. The mechanism underlying the action of epistatic selection is based on a model of compensatory fitness interactions (Kimura 1985), which suggests that mutations occurring in RNA helices are individually deleterious but become neutral in appropriate combinations. Stephan (1996) has shown

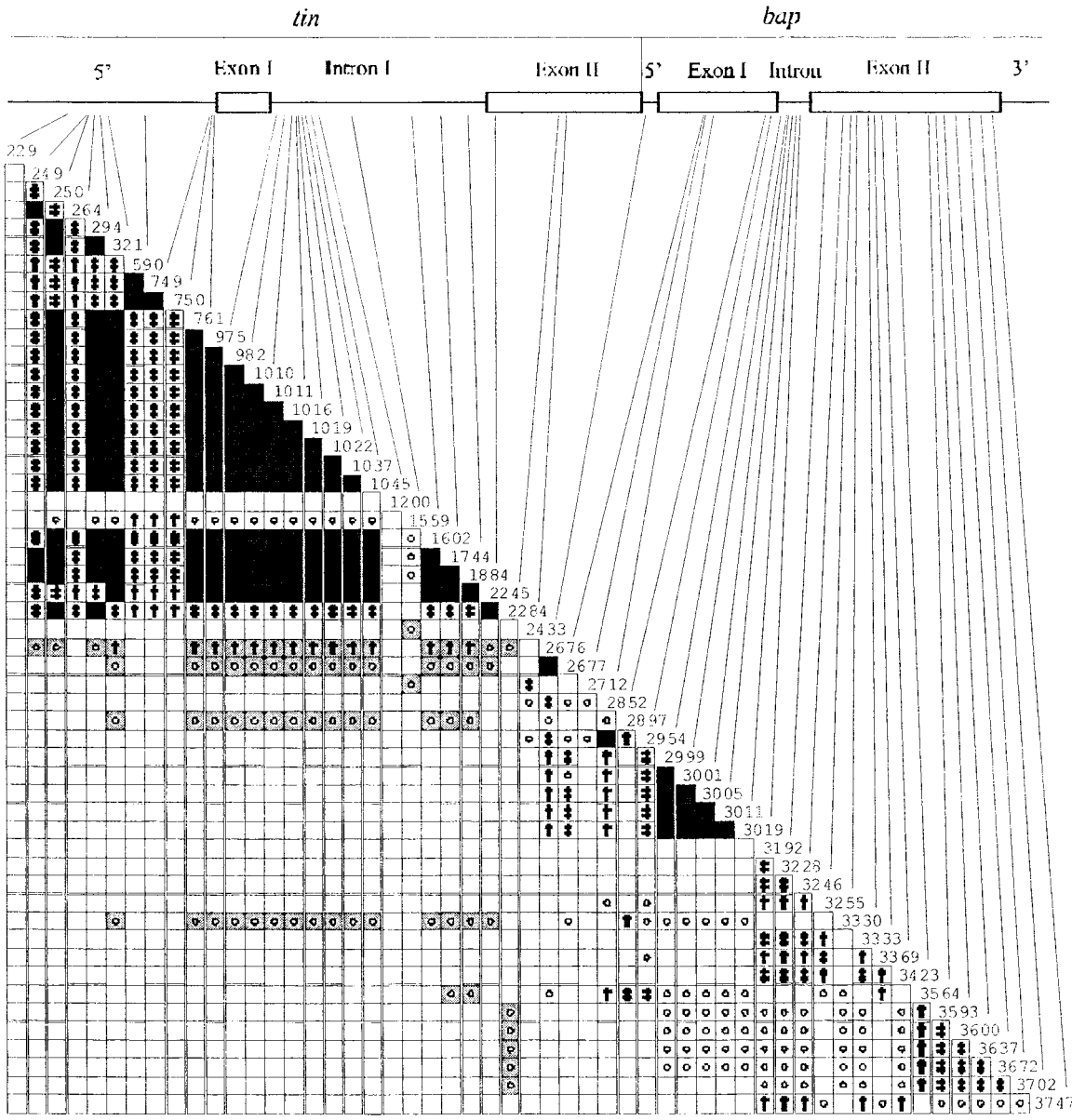

FIGURE 7.—Fisher's exact test of nonrandom associations between pairs of *tin* and *bap* polymorphisms. Singleton mutations are excluded from the analysis. Each box in the matrix represents the comparison of two polymorphic sites. The location of the segregating sites on the *tin* and *bap* genes is shown on the diagonal, which indicates the position of the 5′-flanking, coding, and 3′-flanking regions. The intergenic associations are shaded boxes. The associations that remain significant after Bonferroni correction are solid boxes. ‡, $P < 0.001$; †, $0.001 < P < 0.01$; ○, $0.01 < P < 0.05$.

that the rate of compensatory evolution substantially decays over a distance of 100 nucleotides, which is in agreement with our results. Indeed, the distance between the intronic sites is <100 nucleotides for both *tin* (positions 975, 982, 1010, 1011, 1016, 1019, 1022, 1037, and 1045) and *bap* (positions 2954, 2999, 3001, 3005, 3011, and 3019). Thus the evolution of the intronic sequences of the *tin* and *bap* genes may be subjected to secondary structure constraints.

We have analyzed the relationship between linkage disequilibrium (LD; measured as D') and physical distance between sites by the method of McVEAN *et al.* (2002), with the significance of Pearson's correlation coefficient estimated by 10,000 permutations. There is a significant decline in LD with increasing distance for both *tin* (Pearson's correlation coefficient is −0.1432; $P = 0.0070$) and *bap* (−0.2189; $P = 0.0009$).

Tests of neutrality: KELLY's (1997) Z_{nS} test (Table 4; based on linkage disequilibrium between segregating sites) detects significant deviations from neutrality for the entire region with recombination 0.0072 (the value of recombination obtained by the method of McVEAN *et al.* 2002, Table 2). WALL's (1999) B and Q tests are significant even without recombination. The areas of significant values of Kelly's and Wall's statistics coincide with peaks of linkage disequilibrium within both *tin* and

TABLE 4

Tests of neutrality for the *tin* and *bap* genes of *D. melanogaster*

	tin			*bap*			Entire region		
		0.05	0.01		0.05	0.01		0.05	0.01
Z_{nS}	0.3286	0.005	0.005	0.2118	0.010	0.020	0.1704	0.010	0.010
B	0.3333	0	0	0.3125	0	0	0.3188	0	0
Q	0.4595	0	0	0.3939	0	0	0.4284	0	0

The values refer to Kelly's (1997) Z_{nS} and Wall's (1999) *B* and *Q* values and to the recombination rates at which a test is significant at the 5% (column 0.05) or 1% (column 0.01) level of significance. Indels are ignored.

bap (not shown). For *tin* the tests are significant with a lower level of recombination than for *bap*. For instance, the Z_{nS} statistic obtained for *tin* is significant ($P = 0.01$) with recombination rate $C = 0.0072$, while *bap* requires $C = 0.020$ (Table 4). Overall the tests are significant for the entire region as well as for each gene separately, with a recombination rate much lower than the laboratory estimate ($C_{lab} = 0.0744$) based on the physical and genetic maps of *D. melanogaster* (J. M. Comeron, personal communication; Comeron *et al.* 1999). We suggest that the significance of the tests could reflect the action of selection combined with the demographic history of *D. melanogaster*, which originated from Africa and migrated in relatively recent times to the rest of the world. The higher test statistics values for *tin* may reflect the specific character (rare recombination) of the evolution of this gene, as well as the demographic history of *D. melanogaster*.

There are two sets of divergent haplotypes for the *tin* and *bap* genes (Figure 1). It is appropriate to use the haplotype test (Hudson *et al.* 1994) in this case to see whether directional selection has increased the frequency of some haplotypes. For *tin*, there are a total of 37 polymorphic sites and a subset of 18 sequences with 13 sites (Figure 1, strains 1–18). The haplotype test is not significant ($P = 0.103$) even with a laboratory estimate of recombination equal to 0.0744 (see above). A total of 32 polymorphic sites are in the sample of 27 *bap* sequences, but the set of homogeneous strains (1–13) includes just one polymorphic site (Figures 1 and 3). The probability of this configuration, obtained by the haplotype test, is 0.002, even without recombination. The region of amino acid substitution between species (Figure 1) at the beginning of *bap* exon I may be a likely candidate for a selective sweep.

We have also used the neutrality tests of Depaulis and Veuille (1998) to analyze the haplotype distribution. The tests are not significant for the *tin* gene (as in the Hudson *et al.* 1994 and McDonald 1996, 1998 tests). The tests are significant for the *bap* gene when applied to the homogeneous and heterogeneous sets of sequences separately (the haplotype groups are the same as for the Hudson *et al.* 1994 test). Particularly, for the homogeneous set of sequences, the observed haplotype diversity is 0.167, while the expected haplotype diversity is significantly higher (0.640, $P < 0.05$), which is compatible with the hypothesis of directional selection. For the heterogeneous set of sequences, this test reveals a significant excess of variability: the observed haplotype diversity is 0.952, but the expected haplotype diversity is 0.850 ($P < 0.05$), which is compatible with the hypothesis of diversifying selection. We have also applied the Depaulis and Veuille (1998) tests for different functional parts of the *tin* and *bap* genes. There is no deviation from neutral expectation for any partition (5′-flanking region, intron, and exon II) of the *tin* gene (exon I is excluded from these separate analyses because it is only 186 bp). For the *bap* gene, the tests are significant for exon I and exon II separately, but this significance is due to opposite patterns. For exon I, the test reveals a significant excess of different haplotypes: the observed number is 7, but only 4.9 haplotypes are expected ($P < 0.05$). For exon II, the test reveals a significant deficit of haplotypes: the observed diversity is 0.604, while the expected haplotype diversity is significantly higher (0.820, $P < 0.05$). These additional tests corroborate our suggestion that the *bap* gene is under the complex influence of positive selection (see above).

DISCUSSION

There is a significant difference in the level and pattern of nucleotide variability in *tin* and *bap*, two closely linked homeobox genes of *D. melanogaster*. The level of *tin* variability is within the range observed in other regulatory genes of Drosophila (Moriyama and Powell 1996; Powell 1997; Baines *et al.* 2002; Riley *et al.* 2003) and some other organisms (Purugganan 2000). The silent variability of *bap* is, however, significantly higher and close to the values observed for the most polymorphic Drosophila genes, such as *Est-6* and ψ*Est-6* (Balakirev and Ayala 1996, 2003a,b, 2004; Balakirev *et al.* 1999, 2002, 2003; Ayala *et al.* 2002). The pattern of nucleotide variability in *tin* and *bap* is not compatible with an equilibrium model of selective neutrality. We suggest that

the colonizing and demographic history of *D. melanogaster* together with negative (purifying) selection may be the main factors shaping the observed patterns of nucleotide variability. The *bap* data suggest that positive selection may also contribute to the observed patterns: diversifying selection would have increased the level of nucleotide variation, while directional selection would account for the excess of nearly identical sequences. Positive selection in the *bap* gene is supported by significant HKA (Hudson *et al.* 1987), McDonald (1996, 1998), and haplotype (Hudson *et al.* 1994; Depaulis and Veuille 1998) tests. We have previously suggested that the pattern of nucleotide variability of the *Est-6* coding region is shaped by the influence of both directional and balancing selection (Balakirev *et al.* 1999, 2002; Ayala *et al.* 2002). A similar account has been proposed for the *Adh* locus of *A. thaliana* (Hanfstingl *et al.* 1994), the *Acp29AB* gene of *D. melanogaster* (Aguadé 1999), and regulatory gene *TFL1* of *A. thaliana* (Olsen *et al.* 2002). The results of the neutrality tests must, nevertheless, be cautiously interpreted, given the modest-sized sample of sequences from a single population (Simonsen *et al.* 1995). Moreover, there are nonselective factors that could partly account for the patterns of the *tin* and *bap* polymorphisms. Possible explanatory processes include bottlenecks and founding effects and/or population admixture, as well as varying recombination rates in different genomic regions. One way of distinguishing between selective and demographic processes could be to perform similar investigations in other populations of *D. melanogaster*.

The homeobox *tin* and *bap* genes are involved in recruiting cardioblasts and visceral muscle primordia from the mesodermal mass (Azpiazu and Frasch 1993; Bodmer 1993). We have detected significant differences in the level and pattern of nucleotide variability between two closely linked genes that occupy different hierarchical positions in the interacting gene network. *tin* functions at the top of a genetic hierarchy, specifying the heart and the visceral mesoderm; *tin* is first expressed in all mesoderm and *tin* mutations abolish *bap* expression but not vice versa (Azpiazu and Frasch 1993). The level of variability is significantly lower and the ratio of divergence-to-polymorphism is higher in *tin* than in *bap*. Also, silent divergence between *D. melanogaster* and *D. simulans* is higher in the introns than in the exons of *tin*, suggesting that selective constraints reduce the level of variation in the *tin* coding regions. The difference between synonymous and nonsynonymous divergence is less than half for the *tin* gene compared to that for *bap*, suggesting that negative selection is stronger in *tin*. But there is evidence that positive selection is involved in the molecular evolution of *bap*. Overall, it appears that the higher hierarchical position of *tin* is associated with lower genetic variability and negative selection, whereas the functionally dependent component of this interacting complex, *bap*, exhibits evidence of adaptive evolution. A similar relationship was observed between genes of the *Ras*-mediated signal transduction pathway of Drosophila (Riley *et al.* 2003).

We are grateful to J. H. McDonald, G. McVean, D. A. Filatov, J. K. Kelly, J. D. Wall, J. M. Comeron, F. Depaulis, and J. Rozas for useful advice on analyses and for providing computer programs. We thank Elena Balakireva for encouragement and help. This work is supported by National Institutes of Health grant GM42397 to F. J. Ayala.

LITERATURE CITED

Adams, M. D., S. E. Celniker, R. A Holt, C. A. Evans, J. D. Gocayne *et al.*, 2000 The genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195.

Aguadé, M., 1999 Positive selection drives the evolution of the Acp 29AB accessory gland protein in Drosophila. Genetics **152:** 543–551.

Andolfatto, P., and M. Przeworski, 2000 A genome-wide departure from the standard neutral model in natural populations of Drosophila. Genetics **156:** 257–268.

Aquadro, C. F., 1992 Why is the genome variable? Insights from *Drosophila*. Trends Genet. **8:** 355–362.

Ayala, F. J., E. S. Balakirev and A. G. Sáez, 2002 Genetic polymorphism at two linked loci, *Sod* and *Est-6*, in *Drosophila melanogaster*. Gene **300:** 19–29.

Azpiazu, N., and M. Frasch, 1993 *tinman* and *bagpipe*: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. Genes Dev. **7:** 1325–1340.

Baines, J. F., Y. Chen, A. Das and W. Stephan, 2002 DNA sequence variation at a duplicated gene: excess of replacement polymorphism and extensive haplotype structure in the *Drosophila melanogaster bicoid* region. Mol. Biol. Evol. **19:** 989–998.

Balakirev, E. S., and F. J. Ayala, 1996 Is esterase-P encoded by a cryptic pseudogene in *Drosophila melanogaster*? Genetics **144:** 1511–1518.

Balakirev, E. S., and F. J. Ayala, 2003a Nucleotide variation of the *Est-6* gene region in natural populations of *Drosophila melanogaster*. Genetics **165:** 1901–1914.

Balakirev, E. S., and F. J. Ayala, 2003b Molecular population genetics of the β-esterase gene cluster of *Drosophila melanogaster*. J. Genet. **82:** 101–117.

Balakirev, E. S., and F. J. Ayala, 2003c Pseudogenes: Are they "junk" or functional DNA? Annu. Rev. Genet. **37:** 123–151.

Balakirev, E. S., and F. J. Ayala, 2004 The β-esterase gene cluster of *Drosophila melanogaster*: Is ψ*Est-6* a pseudogene, a functional gene, or a potogene? Genetica **121** (in press).

Balakirev, E. S., E. I. Balakirev, F. Rodriguez-Trelles and F. J. Ayala, 1999 Molecular evolution of two linked genes, *Est-6* and *Sod*, in *Drosophila melanogaster*. Genetics **153:** 1357–1369.

Balakirev, E. S., E. I. Balakirev and F. J. Ayala, 2002 Molecular evolution of the *Est-6* gene in *Drosophila melanogaster*: contrasting patterns of DNA variability in adjacent functional regions. Gene **288:** 167–177.

Balakirev, E. S., V. R. Chechetkin, V. V. Lobzin and F. J. Ayala, 2003 DNA polymorphism in the β-*esterase* gene cluster of *Drosophila melanogaster*. Genetics **164:** 533–544.

Begun, D. J., S. N. Boyer and C. F. Aquadro, 1994 *cut* locus variation in natural populations of *Drosophila*. Mol. Biol. Evol. **11:** 806–809.

Bodmer, R., 1993 The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila*. Development **118:** 719–729.

Collet, C., K. M. Nielsen, R. J. Russell, M. Karl, J. G. Oakeshott *et al.*, 1990 Molecular analysis of duplicated esterase genes in *Drosophila melanogaster*. Mol. Biol. Evol. **7:** 9–28.

Comeron, J. M., M. Kreitman and M. Aguadé, 1999 Natural selection on synonymous sites is correlated with gene length and recombination in Drosophila. Genetics **151:** 239–249.

Depaulis, F., and M. Veuille, 1998 Neutrality tests based on the distribution of haplotypes under an infinite-site model. Mol. Biol. Evol. **15:** 1788–1790.

Duboule, D. (Editor), 1994 *Guidebook to the Homeobox Genes*. Sam-

brook & Tooze Publications/Oxford University Press, London/New York/Oxford.

Filatov, D. A., and D. Charlesworth, 1999 DNA polymorphism, haplotype structure and balancing selection in the Leavenworthia PgiC locus. Genetics **153:** 1423–1434.

Gibson, G., and D. S. Hogness, 1996 Effect of polymorphisms in the *Drosophila* regulatory gene *Ubx* on homeotic stability. Science **271:** 200–203.

Goss, P. J. E., and R. C. Lewontin, 1996 Detecting heterogeneity of substitution along DNA and protein sequences. Genetics **143:** 589–602.

Hanfstingl, U., A. Berry, E. A. Kellogg, J. T. Costa, III, W. Rüdiger *et al.*, 1994 Haplotype divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: Roles for both balancing and directional selection? Genetics **138:** 811–828.

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

Hudson, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Biol. **7:** 1–44.

Hudson, R. R., and N. Kaplan, 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., and N. Kaplan, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

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

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

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

Jagla, K., M. Bellard and M. Frasch, 2001 A cluster of *Drosophila* homeobox genes involved in mesoderm differentiation program. BioEssays **23:** 125–133.

Jukes, T. H., and C. R. Cantor, 1969 Evolution of protein molecules, pp. 21–120 in *Mammalian Protein Metabolism,* edited by H. M. Munro. Academic Press, New York.

Kelly, J. K., 1997 A test of neutrality based on interlocus associations. Genetics **146:** 1197–1206.

Kim, Y., and M. Nirenberg, 1989 *Drosophila* NK-homeobox genes. Proc. Natl. Acad. Sci. USA **86:** 7716–7720.

Kimura, M., 1985 The role of compensatory neutral mutations in molecular evolution. J. Genet. **64:** 7–19.

Kirby, D. A., S. V. Muse and W. Stephan, 1995 Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Natl. Acad. USA **92:** 9047–9051.

Lawton-Rauh, A., R. H. Robichaux and M. D. Purugganan, 2003 Patterns of nucleotide variation in homeologous regulatory genes in the allotetraploid Hawaiian silversword alliance (Asteraceae). Mol. Ecol. **12:** 1301–1313.

Lemeunier, F., and S. Aulard, 1992 Inversion polymorphism in *Drosophila melanogaster*, pp. 339–405 in *Drosophila Inversion Polymorphism*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Cleveland.

Ludwig, M. Z., and M. Kreitman, 1995 Evolutionary dynamics of the enhancer region of *even-skipped* in *Drosophila.* Mol. Biol. Evol. **12:** 1002–1011.

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

McVean, G., P. Awadalla and P. Fearnhead, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics **160:** 1231–1241.

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

Miyashita, N., M. Aguadé and C. H. Langley, 1993 Linkage disequilibrium in the *white* locus region of *Drosophila melanogaster*. Genet. Res. **62:** 101–109.

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.

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

Olsen, K. M., A. Womack, A. R. Garrett, J. I. Suddith and M. D. Purugganan, 2002 Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. Genetics **160:** 1641–1650.

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

Purugganan, M. D., 2000 The molecular population genetics of regulatory genes. Mol. Ecol. **9:** 1451–1461.

Purugganan, M. D., and J. I. Suddith, 1998 Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proc. Natl. Acad. Sci. USA **95:** 8130–8134.

Purugganan, M. D., and J. I. Suddith, 1999 Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana.* Genetics **151:** 839–848.

Riley, R. M., W. Jin and G. Gibson, 2003 Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in Drosophila. Mol. Ecol. **12:** 1315–1323.

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

Sawyer, S. A., 1989 Statistical tests for detecting gene conversion. Mol. Biol. Evol **6:** 526–538.

Sawyer, S. A., 1999 *GENECONV*: a computer package for the statistical detection of gene conversion. Department of Mathematics, Washington University, St. Louis.

Schaeffer, S. W., and E. L. Miller, 1993 Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura.* Genetics **135:** 541–552.

Seager, R. D., and F. J. Ayala, 1982 Chromosome interactions in *Drosophila melanogaster*. I. Viability studies. Genetics **102:** 467–483.

Shepard, K. A., and M. D. Purugganan, 2003 Molecular population genetics of the Arabidopsis *CLAVATA2* region: the genomic scale of variation and selection in a selfing species. Genetics **163:** 1083–1095.

Simonsen, K. L., G. A. Churchill and C. J. Aquadro, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–429.

Smit-McBride, Z., A. Moya and F. J. Ayala, 1988 Linkage disequilibrium in natural and experimental populations of *Drosophila melanogaster*. Genetics **120:** 1043–1051.

Stephan, W., 1996 The rate of compensatory evolution. Genetics **144:** 419–426.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Ting, C. T., S. C. Tsaur, M. L. Wu and C.-I Wu, 1998 A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282:** 1501–1504.

Wall, J. D., 1999 Recombination and the power of statistical tests of neutrality. Genet. Res. **74:** 65–79.

Wall, J. D., 2001 Insights from linked single nucleotide polymorphisms: what we can learn from linkage disequilibrium. Curr. Opin. Genet. Dev. **11:** 647–651.

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

Communicating editor: L. Harshman