

Transpecific Polymorphisms in an Inversion Linked Esterase Locus in *Drosophila buzzatii*

Gloria A. Gómez and Esteban Hasson

Departamento de Ecología Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

Nucleotide variation was studied in a 1.1 kb section of the coding region of an *Esterase* gene (*Est-A*) that maps in the center of the segments rearranged by polymorphic inversions in the cactophilic *Drosophila buzzatii*. We examine 30 homozygous second-chromosome lines differing in gene arrangement and three *D. koepferae* isofemale lines as outgroups. Our data show that *Est-A* is a highly polymorphic gene at both synonymous and replacement sites. Significant departures from homogeneity in the distribution of the ratio of silent polymorphism to divergence predicted by the neutral theory reveals a local excess of silent polymorphism. This is consistent with the presence of two apparent narrow peaks of elevated silent polymorphism surrounding nonconservative amino acid substitutions. These polymorphisms as well as others at synonymous and nonsynonymous sites are shared with *D. koepferae*. We suggest that the presence of shared nucleotide polymorphisms is probably due to interspecific gene flow and/or balancing selection acting on replacement variants and/or to a decreased probability of loss of ancestral polymorphisms caused by linkage to an adaptive inversion polymorphism. Recurrent mutation and persistence of neutral ancestral polymorphisms cannot, however, be ruled out. The analysis of the distribution of nucleotide variation among the three chromosomal arrangements sampled reveals that derived arrangements (*J* and *JZ*³) are less polymorphic than the ancestral *ST*, and that the widely distributed *ST* and *J* arrangements are genetically differentiated. However, a significant number of polymorphisms are shared between arrangements, suggesting frequent exchange either from gene conversion or from double crossovers in heterokaryotypes. Finally, our present results in combination with data of sequence variation at the breakpoints of inversion *J* suggest that this old gene arrangement has risen in frequency in relatively recent times.

Introduction

The observation of parallel and reciprocating clines along geographic and/or climatic gradients in different continents, long-term trends, temporal cycles, and more sophisticated experimental approaches in natural populations supplied the types of data that give support to the hypothesis of the adaptive role of chromosomal inversion polymorphisms in *Drosophila* (reviewed in Krimbas and Powell 1992; Powell 1997). Nevertheless, understanding the mechanisms involved in their maintenance in natural populations has been an intriguing and recurrent issue. Heterokaryotype advantage has been the most often invoked mechanism (reviewed in Krimbas and Powell 1992), whereas spatial (so-called multiple niche selection) and/or temporal variation in selection coefficients, epistasis or antagonistic pleiotropic effects on different fitness components are alternative explanations for the maintenance of inversion polymorphisms in natural populations (reviewed in Krimbas and Powell 1992; Powell 1997).

The utilization of sequencing techniques in population surveys has caused renewed interest in the study of inversion polymorphisms (Andolfatto, DePaulis, and Navarro 2001). Several issues of their evolutionary history, such as mutational origin, extent to which different recombination environments—near the breakpoints or in the center more recombining parts of the rearranged segment—affect nucleotide variation and the mechanisms involved in the maintenance of inversion polymorphisms, may be revisited within the framework of coalescent theory (Hudson 1990).

Key words: *Drosophila buzzatii*, inversion polymorphism, *Esterase-A*, gene flow, natural selection, nucleotide variation.

E-mail: ehasson@bg.fcen.uba.ar.

Mol. Biol. Evol. 20(3):410–423. 2003

DOI: 10.1093/molbev/msg051

© 2003 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

In recent years a great effort has been devoted to study nucleotide variation in coding and noncoding regions with different degrees of linkage with inversion complexes in well-known *Drosophila* model systems such as *In(3L)Payne* (Wesley and Eanes 1994; Hasson and Eanes 1996; Verrelli and Eanes 2000) and *In(2L)t* (Bénassi et al. 1993, 1999; Andolfatto, Wall, and Kreitman 1999; DePaulis, Brazier, and Veuille 1999; Andolfatto and Kreitman 2000; DePaulis et al. 2000) in *D. melanogaster*; the third chromosome inversion system (Popadic, Popadic, and Anderson 1995) and the sex ratio complex (Babcock and Anderson 1996) in *D. pseudoobscura*; and the *O* (Rozas et al. 1999) and *A* chromosome inversion complexes (Munté, Aguadé, and Segarra 2000) in *D. subobscura*. Three main conclusions can be drawn from these studies. First, patterns of distribution of nucleotide variation within and between arrangements are not compatible with the hypothesis that inversions are old balanced polymorphisms (Andolfatto, DePaulis, and Navarro 2001). Second, inversions have a monophyletic origin, even in those cases, such as *D. melanogaster In(2L)t* and *D. buzzatii* inversion *2J*, in which the involvement of transposable elements in the origin of inversions has been demonstrated, (Andolfatto, Wall, and Kreitman 1999; Cáceres et al. 1999; Cáceres, Puig, and Ruiz 2001; and see also Lyttle and Haymer 1992; Ladeveze et al. 1998). Third, genetic exchange is greatly impeded in the vicinity of inversion breakpoints, while in the middle of inversions genetic exchange by means of double crossovers and/or gene conversion can lead to a complete homogenization of the genetic content between inverted and noninverted chromosomes (Navarro, Barbadiña, and Ruiz 2000; Andolfatto, DePaulis, and Navarro 2001).

In the cactophilic *D. buzzatii*, previous studies have shown that populations are strongly structured for the

inversion polymorphism but less structured for variation at electrophoretic loci (Rodríguez et al. 2000). On one hand, the concordant latitudinal clines in inversion frequencies reported in original South American and recently colonized Australian populations are indicative of strong adaptive forces (Hasson et al. 1995; Rodríguez et al. 2000). On the other hand, clinal variation for allozymes linked to inversions can be mainly accounted for by hitchhiking with inversions, with the exception of the *Esterase-2* locus (Rodríguez et al. 2000). Moreover, linkage disequilibrium between inversions and allozymes (Rodríguez et al. 2001) and patterns of population structure with different genetic markers (Rossi et al. 1996; Rodríguez et al. 2000) suggest that population expansions may have been important in the recent history of *D. buzzatii*.

In this article, we analyze intraspecific nucleotide variation at the *Esterase-A* (*Est-A*) locus in samples of three different second chromosome inversion arrangements segregating in natural populations of *D. buzzatii* and interspecific divergence with the closely related *D. koepferae*. This gene, which has been isolated, sequenced and described (East, Graham, and Whittington 1990) as a putative member of the β -cluster of the carboxyl/cholinesterases family specific for carboxylester substrates (Oakeshott et al. 1999), is located in the center of the segments rearranged by two overlapping inversions, $2J$ and $2Z^3$. The former arose from a *ST* ancestral chromosome and the latter occurred on a $2J$ chromosome and defines the $2JZ^3$ arrangement that differs from $2J$ by a single overlapping inversion.

We address the following questions: (1) Are nucleotide sequence data consistent with the cytogenetic phylogeny accepted for the species? (2) Are closely linked regions like those inside the inversion *J* complex monophyletic? (3) Is nucleotide variation at *Est-A* compatible with neutral expectations? (4) Have derived arrangements reached mutation-drift-flux equilibrium? (5) Are the patterns of distribution of nucleotide variation within and between arrangements compatible with the hypothesis that the inversion polymorphism is an old balanced polymorphism? (6) Does the analysis of nucleotide variation allow us to infer the evolutionary forces that shaped the history of the inversion polymorphism?

Materials and Methods

Drosophila Strains

Thirty *D. buzzatii* strains were isogenized for the second chromosome by means of crosses of individual wild males collected in eight natural populations of Argentina in late summer of 1995 and 1997, with virgin females of the balancer stock *Ant/ Δ^5* (kindly provided by J. S. F. Barker) as described in Rodríguez et al. (2000). Populations sampled were Chumbicha (13 strains), Catamarca (2), Termas de Río Hondo (2), Tilcara (2), Puerto Tirol (3), Berna (2), and Quilmes and Otamendi (4). Detailed descriptions of the populations sampled can be found in Hasson et al. (1995) and Rodríguez et al. (2000). Once homozygous stocks were established the polytene chromosomes of one F_3 third instar larva of each strain were analyzed to ascertain the inversion arrangement

carried by the wild male transmitted to the F_1 male chosen to start the line. Homozygous strains were named using a three-letter code identifying the population of origin, a number and the chromosomal arrangement, s for *ST*, j for *J*, and z for *JZ*³. Individuals of three isofemale lines (fourth generation in the lab) of the symorphic species *D. koepferae* collected in the localities of Cafayate and Famatina (see Hasson et al. 1995 for details of the sampled localities) were employed as outgroups for several purposes.

In Situ Hybridization

Salivary glands of third instar larvae grown in uncrowded cultures were dissected and processed according to the method of Montgomery, Charlesworth, and Langley (1987) to obtain polytene chromosome slides. A polymerase chain reaction (PCR) amplified fragment of 1.1 kb of *Est-A* was used as probe labeled with digoxigenin 11 dUTP (Boehringer Mannheim) using a random priming reaction. Prehybridization solutions and hybridization washes were performed according to the instructions of the manufacturer.

DNA Extraction, Amplification, and Sequencing

Genomic DNA was extracted from individual flies using the kit Puregene (Gentra systems). The published sequence of the *Est-A* gene (East, Graham, and Whittington 1990) was employed to design specific PCR and sequencing primers.

Initially a fragment of 1.1 kb uncovering most of the coding region of the gene was PCR amplified using primers (numbers refer to the nucleotide site in the published sequence) 913: 5'-CTACGTGGACGCGAT-AAAGG-3' (forward primer) and 2014: 5'-CGCTGAAG-TTCCTATAGCCG-3' (reverse primer). All PCR reactions were carried out in final volumes of 50 μ l containing 40 μ M of each dNTP, 2 units of *Taq* DNA polymerase (Promega), and 90 ng of direct and reverse primers, 1.5 mM of $MgCl_2$ and 1 μ l of genomic DNA. Amplifications were performed according to the following PCR profile: 1 min at 94°C and 35 cycles of denaturation (1 min at 94°C), annealing (1 min at 60°C) and extension (2 min at 72°C), and a final extension step of 5 min at 72°C. All PCR products were gel purified from 1% low melting point agarose (GIBCO) gels and used as templates for the amplification of two shorter and overlapping fragments using two primer combinations 913-1615 (5'-CGATGC-TTTGCGACTGTAC-3') and 1391 (5'-TCGGTGGAG-ATCCAGAGAAC-3')-2014. PCR products were excised from low melting point agarose gels and purified using the QIAquick gel extraction kit (QIAGEN Inc.). Purified DNAs were cycle sequenced using fmol sequencing system (Promega). Both strands were sequenced using a set of primers spaced 300 nucleotides. Reactions were run on acrylamide gels and electrophoresed for 3–7 h depending on the length of the fragment.

DNA of *D. koepferae* was also extracted from individual flies and subjected to the procedures as described above, with minor modifications. Two additional primers were needed to complete the sequencing.

Sequences were aligned manually. All sequences reported in this study were deposited in GenBank under accession numbers AY113076–AY113105.

Data Analysis

Nucleotide polymorphism at all sites, synonymous and replacement sites, were estimated as the number of segregating sites (S) and the average number of pairwise differences (k), the average number of pairwise differences per site (π) and the expected heterozygosity per site under the mutation drift-equilibrium given the observed S value (θ_s) (Watterson 1975).

Genetic differentiation between arrangements or between groups of populations was estimated as the average number of pairwise differences per site between arrangements/populations (d_{xy}) and as the net number of nucleotide differences per site (d_a), according to equations 10.20 and 10.21 in Nei (1987), respectively. The extent of genetic differentiation was also analyzed by means of F_{ST} and K_{ST} statistics, two different measures of the proportion of nucleotide diversity attributable to variation between arrangements (or populations), (equations 3 and 9 in Hudson, Slatkin, and Maddison 1992; Hudson, Boos, and Kaplan 1992). Statistical significance of F_{ST} and K_{ST} was tested by means of permutations tests (Hudson, Boos, and Kaplan 1992) implemented in the program Proseq 2.9 (Filatov, D., <http://www.biosciences.bahm.ac.uk/labs/filatov/proseq.html>).

Linkage Disequilibrium and Recombination

The significance of pairwise associations between informative polymorphic sites was determined by means of Fisher Exact tests and adjusted for multiple comparisons by sequential Bonferroni tests (Rice 1989). In addition, the sign test on D (Lewontin 1995) was applied to search for overall evidence of linkage disequilibrium. This test allows analysis of linkage disequilibrium between independent pairs of polymorphic loci. For simplicity we considered adjacent pairs of sites as independent. The data set consisted of all polymorphic sites including singletons. The rationale of this test is to determine whether the number of positive and negative D values between pairs of sites differ from expectations under the null hypothesis of site independence.

The recombination parameter C ($C = 4Nc$, where N is the effective population size and c the recombination rate per generation between the most distant sites) was estimated using the Hudson and Kaplan (1985) method, based on the minimum number of recombination events determined using the four gamete rule. Sites with more than two segregating variants were excluded from the analysis. We also employed the method of Hudson (1987) to obtain estimates of the recombination parameter.

Tests of Neutrality

The tests of Tajima (1989), Fu and Li (1993), and Fu (1997) were applied to determine whether intraspecific patterns of variation detected in the *Est-A* region were in agreement with neutral predictions. Tajima's D test statistic is based on the standardized difference between two estimates of the neutral parameter ($\theta = 4N\mu$), one based on the observed average number of pairwise differences (k) and the other on the observed number of segregating sites in the sample (S). Fu and Li (1993) proposed tests that compare different estimators of the neutral parameter. Statistic D is based on the standardized difference between the total number of mutations (η) in all branches of the genealogy and the number of external mutations, and F is the standardized difference between k and the number of external mutations. External mutations are those occurring in external branches of the genealogy and can be defined as singletons absent in the sequence of an outgroup.

The test devised by Fu (1997), which examines the significance of the probability S' of having no fewer than k_0 haplotypes in a random sample given that $\theta = \pi$ was also applied to the entire sample and to the subsamples of ST and J chromosomes.

DNAsp version 3.53 (Rozas and Rozas 1999) was employed to obtain estimates of nucleotide polymorphism, detection of conversion tracts, linkage disequilibrium and recombination, and to perform most neutrality tests. This program was also used to infer the statistical significance of Tajima, Fu, and Li's statistics by means of simulations of the coalescent process (10000 replications) assuming no recombination.

The direct relationship between polymorphism and divergence predicted by the neutral theory was tested by means of the average sliding G test (McDonald 1998). This test examines the distribution of polymorphic relative to diverged sites along a DNA sequence. The data set consists of an ordered series of variable sites classified as *polymorphic* or *fixed* differences. Values of the test statistics and their significance estimated by means of Monte Carlo simulations of a coalescent model incorporating both sampling and phylogenetic variation were estimated using the program DNASLIDER (McDonald 1998).

Neighbor-Joining trees (Saitou and Nei 1987) were obtained using MEGA version 2.1 (Kumar et al. 2001).

The distribution of observed and expected nucleotide polymorphism along the sequence was investigated using sliding window graphs obtained with DNAsp version 3.53.

Results

Cytological Analysis and In Situ Hybridization

Inversion frequencies in the populations studied have been previously reported (Hasson et al. 1995; Rodríguez

FIG. 1.—Nucleotide polymorphism at the *Est-A* gene region of *D. buzzatii*. Variants found in *D. koepferae* sequence at the polymorphic sites in *D. buzzatii* are indicated in the last rows. Nucleotides identical to the first reference sequence are indicated by a dot. Blocks of nucleotides transferred from one gene arrangement to another are depicted as shaded boxes.

	122233333	3344555677	7888888999	9901344568	11111111	1222222222	2222222222	3333333333	3333344444
	9456701235	7925178824	7013457123	4687217109	9011111123	3345666789	0124556677	8888922245	
Chu109s	TACCTCGCAA	CTCCGCCCTC	TCTAAAGGAG	CCCGATTGAG	GCCAATCACA	ATATCGCTCC	TATTCCTAGT	CTAATGCCAA	
Ota26s	. . . T G G T . A . C C T C T . C G . . G	. . . G	
Trh10s	. . . C G T T T C . C T . G C G . . G	. . . G A G T
Ber1s G . . . T . C	. . . A . G . . . A T T . A C . C T . . . G . . G	. . . A C G
Ber36s	CT G T . C G . . . G A A G T	. . . T . . . C C A C T . . . G . . G C
Cat2s G C G C C T . . . G . . G	. . . T T
Cat8s G . . . C . . C	. . . A . . . G T A . C C . G C T . . . G . . G G
Tir17s T G . . . C . . C	. . . A . . . G . . . A A G T	. . . A . . . C C T G C T . . . G . . G G . . . A . . .
Tir20s	. . . T G C A A A G T A C C C T . G C G . . G
Ota4s	CT G C A . . . G . . . A A G T	. . . A . . . C C A T . . . G . . G G
Chu74s	CT G G T . C A . G . G . A A G T	. . . A . . . C C T T G G . T . G C . T G G T . . C . . T
Chu41s	CT G C A . . . G . . . A G T	. . . A . . . C C T . G C G . . G G
Chu53s G . . . C . . C	. . . A . . . C A A G T G C C . . . T C . C T . G C G . . G G T
Chu43s	. . . C G . . . T . C C . . . A . C . C A A G T G C C T . . . A C . C T C G C G . . G G T .
Tir7J	CT . . . T . . . T	. . . G . . . C C . . . A . . . G A A G T A C C T . . . G . . G G
Chu6J	C . . . T C G G G G C C T . C G . . G T G T . . A . .
Chu7J	C . . . T C G G C G C C C T . C G . . G G . . . A . . .
Chu45J	CT G . . . C . . C	. . . G C G A G G A A G T A C T A T T . . . C G T . . . G . . G	. . . G G T
Chu118J	C G G C C T . G C G . . G G
Chu120J	CT . . . T . . . T	. . . G . . . C . . C	. . . A . . . G A A G T A C C T . G C G . . G G
Till1J	C . . . T G G T . . . G C C C T . . . G C . G G T . . A . .
Ota73J G G C C A C . C T . . . G . T G G T . . A . .
Qui7J	C G G C C T . . . G . . G	. . . T . A G
Qui19J	C G . . . C C C T . . . G . . G	. . . C A T
Trh6J	C G G C C T . . . G . . G T G T . . A . .
Ota7J	C G G G C C T . . . G . . G G
Chu2Z	C G G C C T . . . G . . G T G T . . A . .
Til3Z G . . . C . . C	. . . A . . . G A A G T A T T . . . C . C T . G C G . . G T G T . . . T
Chu20Z T G C . . . A . C A A G T G C C C G T . . . G . . G T G T . . A . .
Chu38Z G C A . C . . . A A G T	. . . G . . . C C C T T . . . G . T G T G T . . A . .
Caf1a	. . . T G . . . C . . C	. . . A . C . C . A T A . C . T G C . C T . . . G . T G	. . . G C . . . T . C . . .
Caf1b	. . . T G . . . C . . C	. . . A . G . C . A T A . C . T G C . C T . . . G . T G	. . . G C . . . T . C . . .
Caf2	C G . . . C . . C C . A . . . A T A C . T G C . C T . . . G . T G	. . . G
Fam1	CT G . . . T C C . . . A . C . C . A A T A . C . T G C . C T . . . G . T G	. . . G C . . . T . C . . T
Fam1b	CT G . . . T C C . . . A . C . C . A A T A . C . T G C . C T . . . G . T G	. . . G C . . . T . C . . T

	4444444455	5555556666	6666667777	8888999999	900000
	5555689901	1135671223	3466783478	2288001124	503445
	1269162540	5942795596	9324455980	3507695876	469070
Chu109s	GAATGCACCA	TCCGCAAGTC	CAGACGTAGC	CTAAATGGCA	GTCATG
Ota26s A T . C C G
Trh10s T . . . T T C G
Ber1s A T . T T C G
Ber36s C T C . T T . G T . A T A T . C . . G . .
Cat2s T C G
Cat8s C T C . T T T C
Tir17s T C . G
Tir20s C T C . T A T C
Ota4s C . T A T C
Chu74s C T C . T . . T T C C
Chu41s C . T A T C C
Chu53s A T A T T G
Chu43s A T C . . . T C C
Tir7 T C
Chu6J T C . A C . A
Chu7J T A C . G
Chu45 C T C . T G . . . G T C
Chu118J T A . C T
Chu120J T A T
Till1J T A T
Ota73J T C . A
Qui7J T T C
Qui19J A T A T . . . C . . T . . .
Trh6J T A . C T
Ota72J T A T
Chu2Z T A T
Til3Z C T C . T T C
Chu20Z A T C
Chu38Z A T C
Caf1a C T C . T T T C
Caf1b C T C . T T T C
Caf2 T T C
Fam1 C T C . T T T C
Fam1b C T C . T T T C

et al. 2000). *Est-A* was localized by means of in situ hybridization in polytene bands D4f-h of *D. buzzatii* second chromosome, in a central position relative to the breakpoints of inversions *J* and *JZ*³ (not shown).

Nucleotide Variation

We sequenced a fragment of 1.1 kb in 30 *D. buzzatii* strains, 14 *ST*, 12 *J*, and 4 *JZ*³. These numbers are proportional to the average inversion frequencies observed in South American populations: *ST* = 0.45, *J* = 0.45 and *JZ*³ = 0.10, though the latter is slightly overrepresented in our sample. However, strains were randomly chosen within each chromosomal class with respect to the gene studied.

A total of 126 polymorphic sites were observed in the sample of 30 alleles (fig. 1); 8 sites were segregating for 3 nucleotide variants which implies that the total number of mutations (η) in our sample was 134. No indel mutations were detected. Nucleotide variation in the *D. buzzatii* *Est-A* gene region is summarized in table 1, in terms of the usual measures of nucleotide polymorphism θ_s and π . Sixty-four (69 mutations) synonymous sites and 62 (65 mutations) replacement sites were polymorphic, and 50% of those were unique mutations (fig. 1; table 1).

Thirty polymorphic sites, 15 at both synonymous and replacement sites, were detected in the three *D. koepferae* isofemale lines sequenced. In this species we also scored sites that were in heterozygous condition because individuals analyzed were not isogenic for the second chromosome (fig. 1). *Caf1* and *Fam3* were heterozygous for more than one site, whereas *Caf2* was fully homozygous. It should be noted, however, that estimates of nucleotide polymorphism in the two species are not directly comparable because *D. koepferae* flies were fourth-generation progeny of wild females used as founders of isofemale lines. Thus, complete nucleotide homozygosity in *Caf2* can easily be explained as the result of inbreeding. In figure 1 we show only the nucleotide present in *D. koepferae* lines at the sites polymorphic in *D. buzzatii*.

Nucleotide Variation Within and Between Chromosomal Arrangements

Levels of variation were estimated for the complete data set and for each chromosomal class (table 2). Comparisons among arrangements revealed that nucleotide variation at synonymous sites was 40% to 45% higher in *ST* chromosomes than in *J* and *JZ*³, and 20% to 34% higher at replacement sites.

Estimates of genetic differentiation among arrangements were low but significant for total nucleotide variation ($F_{ST} = 0.088$, $P = 0.027$; $K_{ST} = 0.026$, $P = 0.003$), synonymous ($F_{ST} = 0.118$, $P = 0.008$; $K_{ST} = 0.041$, $P = 0.000$) and replacement polymorphism ($F_{ST} = 0.058$, $P = 0.125$; $K_{ST} = 0.022$, $P = 0.028$). Similar results were obtained for a pool of neighboring populations including Chumbicha, Catamarca, and Termas de Río Hondo, for which only the comparison between *ST* and *J* was significant ($F_{ST} = 0.091$, $P = 0.03$; $K_{ST} = 0.021$, $P = 0.03$). These estimates are similar to those reported among

Table 1
Summary Statistics of Nucleotide Polymorphism at the *Est-A* Locus in *Drosophila buzzatii* and Divergence Between *D. buzzatii* and the Outgroup *D. koepferae*

	Entire Sample (N=30)		
	Total ^a	Synonymous ^b	Replacement ^c
<i>S</i> (η)	126 (134)	64 (69)	62 (65)
Singletons	69	35	34
θ_s	0.030 (0.0095)	0.06470 (0.0219)	0.0196 (0.0064)
π	0.0225 (0.0010)	0.0493 (0.0220)	0.0148 (0.0067)
<i>K Dbu-Dk</i>	0.0715	0.153	0.0458

NOTE.—Numbers in parentheses are standard deviations of θ_s (without recombination) and π (stochastic plus sampling standard error). *S*: number of segregating sites; η : total number of mutations; π : average number of pairwise differences per site.

^a Total number of sites 1,061.

^b Number of synonymous sites 248.1.

^c Number of nonsynonymous sites 811.9.

gene orders of the *O*₃₊₄ phylad in *D. subobscura* (Rozas et al. 1999), in which the gene region studied is also located in the middle of rearranged segments.

Low levels of nucleotide differentiation among gene arrangements are consistent with the virtual absence of fixed differences (table 3). Moreover, 30% of the sites segregating in two arrangements were polymorphic for the same variants in both gene orders (shared polymorphisms). If we assume that inversions are unique events, i.e., are monophyletic, which in the case of inversion *J* seems to be valid (see Cáceres et al. 1999), there are only two possible explanations for the presence of shared polymorphism: (1) same mutations arising independently in both arrangements or (2) exchange of genetic information between gene arrangements by means of double crossovers and/or gene conversion. There were 24 (11 synonymous, 13 nonsynonymous) shared polymorphisms among the three chromosomal classes. We employed the hypergeometric distribution method to evaluate the probability of a certain number of shared polymorphic sites arising independently by parallel mutation in each chromosomal class. According to the number of synonymous polymorphisms detected in *ST* (48), *J* (33), and *JZ*³ (19), no more than two shared polymorphic sites ($P = 0.068$) are expected by chance. Similarly, given the observed number of nonsynonymous polymorphic sites in each chromosomal arrangement (42 in *ST*, 36 in *2J*, and 18 in *JZ*³), no shared polymorphic sites are expected.

Because in all cases the hypothesis that shared polymorphisms are due to recurrent mutation can be rejected, and the assumption of monophyly of inversions is well supported, we investigated the occurrence of gene conversion tracts between arrangements according to the method of Betrán et al. (1997). Estimates of the statistic ψ which measures the probability of detecting a converted site are given in table 2. The higher the value of ψ , the more accurate the estimate of the number and length of conversion tracts. A total of nine conversion tracts (6 in *J* and 3 in *JZ*³) were identified (fig. 1). It should be noted, however, that the number of gene conversion events might be lower because some lines presented the same tracts.

Table 2
Summary of Nucleotide Polymorphism in Different Gene Arrangements in *D. buzzatii*

	<i>ST</i> (<i>N</i> = 14)			<i>J</i> (<i>N</i> = 12)			<i>JZ</i> ³ (<i>N</i> = 4)		
	Total ^a	Syn ^b	Rep ^c	Total ^a	Syn ^b	Rep ^c	Total ^a	Syn ^b	Rep ^c
<i>S</i> (η)	92 (99)	48 (53)	42 (46)	67 (68)	33 (34)	34 (34)	37 (40)	19 (21)	18 (19)
Singletons	43	24	19	20	9	11	4	1	3
θ_s	0.0273 (0.0028)	0.0636 (0.0092)	0.0172 (0.0026)	0.0218 (0.0026)	0.0460 (0.0080)	0.0149 (0.0024)	0.0198 (0.0124)	0.0442 (0.0270)	0.0128 (0.0082)
π	0.0244 (0.0112)	0.0531 (0.0243)	0.0162 (0.0077)	0.0176 (0.0081)	0.0355 (0.0169)	0.0124 (0.0058)	0.0190 (0.011)	0.0413 (0.0235)	0.0126 (0.0072)

NOTE.—Numbers in parentheses are standard deviations of θ_s (without recombination) and π (stochastic plus sampling standard error). *S*: number of segregating sites; η : total number of mutations; π : average number of pairwise differences per site.

^a Total number of sites 1,061.

^b Number of synonymous sites 248.1.

^c Number of nonsynonymous sites 811.9.

Analysis of linkage disequilibrium between pairs of informative sites using Fisher's Exact test for the full sample showed that 205 (14%) of 1,431 comparisons for 54 informative sites were significant. Similar analysis for each arrangement showed that for 37 informative sites in the sample of *ST* chromosomes, 63 of 666 comparisons (9%) were significant, whereas for *J*, 29 sites were informative and 80 of 406 comparisons (20%) yielded significant results. In both *ST* and *J* the proportion of significant comparisons was near the nominal rejection probability of 5%. However, in the case of arrangement *J*, the higher percentage of pairs of informative sites with significant linkage disequilibrium might be a result of the detected conversion tracts. Pairwise linkage disequilibrium between arrangements and each one of the informative sites in a subsample including only *ST* and *J* chromosomes yielded 14 of 55 significant comparisons, none according to Bonferroni criteria. However, as pointed out by Lewontin (1995), some tests of association cannot give significant results even with extreme values of disequilibrium, particularly when a very large fraction are represented in the data by singletons or only a very few copies of the rare variant. Therefore, we also analyzed linkage disequilibrium using the sign test on *D* (Lewontin 1995). In *ST* we performed 76 independent comparisons and, after pooling cells with low expected values in two classes, the

goodness of fit test revealed a certain excess of negative *D* values ($\chi^2 = 4.9$, $df = 1$, $P = 0.02$), suggesting an excess of repulsion linkage in this arrangement. In contrast, in the 73 independent comparisons performed in *J* we detected an excess of positive *D* values (i.e., of coupling gametes) after pooling cells with low expected values in a single class ($\chi^2 = 5.8$, $df = 1$, $P = 0.01$).

Est-A lies in a potentially high recombining region, and in the middle of the segments rearranged by inversions *J* and *Z*³. The minimum number of recombination events (*MNRE*) for the full sample was 25, and the estimate of the recombination parameter per site according to the method of Hudson (1997) was 0.072 (the method of Hey and Wakeley 1997 gave very similar results). However, estimates of the recombination parameter for each arrangement revealed important differences. On the one hand, *C* per site was high in *ST* (0.15) with a minimum number of 19 recombination events, whereas in *J* it was 7–8 times smaller (0.019), with a minimum number of recombination events of 11.

In figure 2 we present a Neighbor-Joining (NJ) tree based on synonymous variation including the 30 *D. buzzatii* alleles plus five *D. koepferae Est-A* sequences used as outgroups. Although recombination, which as shown above is high, may obscure the genealogical relationships among alleles, the aim of the inclusion of

Table 3
Genetic Differentiation Among Arrangements

	<i>ST–J</i>			<i>ST–JZ</i> ³			<i>J–JZ</i> ³		
	Total ^a	Syn ^b	Rep ^c	Total ^a	Syn ^b	Rep ^c	Total ^a	Syn ^b	Rep ^c
η	130	83	71	107	70	57	80	48	46
Shared	39	24	23	33	23	18	30	21	15
	(60, 31)	(41, 18)	(31, 17)	(66, 7)	(42, 4)	(36, 3)	(40, 10)	(21, 6)	(25, 6)
D_{xy}	0.0238	0.012	0.0129	0.0226	0.012	0.011	0.021	0.010	0.0112
D_a	0.0028	0.002	0.0008	0.0028	0.0008	–0.0002	0.0026	0.0012	0.0014
K_{ST}	0.026**	0.04*	0.023*	0.0067	0.0148*	0	0.175	0.023*	0.019
F_{ST}	0.119*	0.168	0.071*	0.023*	0.065	0	0.122	0.12*	0.125*
ψ		0.00302			0.00236			0.00377	

NOTE.— η : total number of mutations; ψ : parameter that measures the probability of detecting a converted site. Numbers in parentheses represent the number of polymorphisms exclusive to each arrangement.

^a Total number of sites 1,061.

^b Number of synonymous sites 248.1.

^c Number of nonsynonymous sites 811.9.

#: $0.05 < P < 0.10$; * $P < 0.05$, ** $P < 0.01$.

a NJ analysis is simply an exploratory method. There are two features in the tree that deserve attention. One is the absence of geographic structuring. Alleles of North Western (Chu, Til, Cat, Trh, and Qui), North Eastern (Ber and Tir), and Southern (Ota) populations of Argentina, which were shown to be highly differentiated for the inversion polymorphism and some electrophoretic loci (Rodríguez et al. 2000), are not grouped according to geographic origin. Indeed, genetic differentiation among populations was not significant either for synonymous ($F_{st} = 0.0037$, $P = 0.61$; $K_{st} = 0.0059$, $P = 0.40$) or nonsynonymous variation ($F_{st} = 0.0132$, $P = 0.35$; $K_{st} = 0.0032$, $P = 0.36$). In addition, differentiation among populations within arrangement classes was also not significant (not shown). The second aspect is that, although *ST* and derived arrangements do not form clearly defined clusters, nine *J* plus one *JZ*³ form a cluster (with low bootstrap support) that is also present in parsimony trees (not shown). The remaining *J* and *JZ*³ alleles, specifically those involved in events of genetic exchange, are associated with different *ST* alleles (fig. 2).

Neutrality Tests Based on the Polymorphism Frequency Spectrum

Tajima (1989) and Fu and Li (1993) tests were applied to investigate whether the observed frequency distributions of synonymous and replacement polymorphism differed from neutral expectations. Fu and Li tests were negative, indicating an overall excess of external mutations and only significant for all sites ($D = -2.06$, $P = 0.04$ and $F = -2.13$, $P = 0.04$) and replacement sites ($D = -2.21$, $P = 0.02$ and $F = -2.17$, $P = 0.04$) and marginally significant for synonymous sites ($D = -1.72$, $P = 0.07$ and $F = -1.9$, $P = 0.06$) for the entire sample. These tests were not significant when applied to the subsamples of *ST* and *J* chromosomes (not shown). Tajima's test statistic was also negative but not significant (not shown), pointing again to an overall excess of low-frequency variants for total, synonymous, and replacement variation. Though consistent in sign, Fu and Li's and Tajima's tests gave nonsignificant results when applied to a more limited data set consisting of the same group of neighboring populations mentioned above.

Fu's (1997) F_S statistic was negative and significant for the complete dataset for both synonymous ($F_S = -18.48$, $P < 0.0001$) and replacement substitutions ($F_S = -17.36$, $P < 0.0001$) and for the subsamples of *ST* ($F_S = -4.38$, $P = 0.012$ and $F_S = -4.92$, $P = 0.007$ for silent and replacement substitutions, respectively) and *J* ($F_S = -4.39$, $P = 0.007$ and $F_S = -4.68$, $P = 0.009$ for silent and replacement substitutions, respectively). Significant negative values of F_S indicate an excess of haplotypes relative to those expected under neutrality.

Neutrality Tests Based on the Correlation Between Polymorphism and Divergence

The traditional HKA test (Hudson, Kreitman, and Aguadé 1987) designed to test the fit of nucleotide

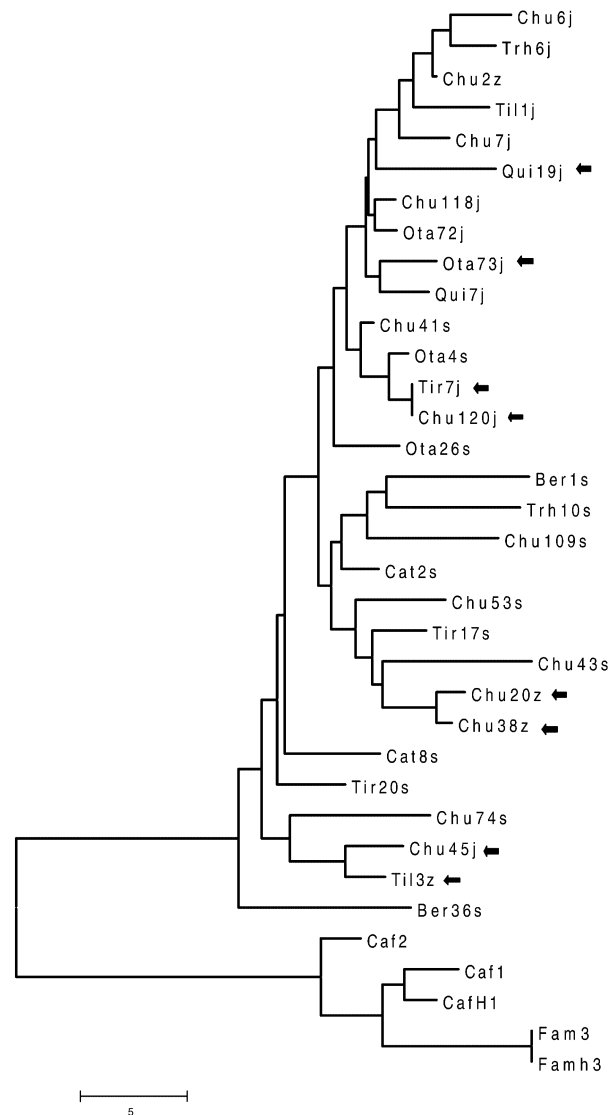


FIG. 2.—Gene genealogy obtained by the neighbor-joining method based on synonymous variation at the *Est-A* locus including 30 *D. buzzatii* sequences and two *D. koepferae* sequences as outgroups. Arrows point to sequences putatively involved in gene conversion events. Bar represents 5 synonymous differences.

polymorphism and interspecific divergence to the neutral model of evolution was not applied, because in *D. buzzatii* we do not have a locus that would serve as a neutral standard. Several tests have recently been reported aimed to analyze the distribution of polymorphic sites relative to diverged sites along a DNA sequence (McDonald 1998). One such test, the average sliding *G* test, which is among the most powerful of those analyzed in McDonald (1998), yielded significant results ($G = 4.15$) when applied to synonymous variation in the subsample of *ST* chromosomes, assuming no recombination ($P = 0.02$), high recombination ($C = 100$), which is close to the value estimated from the data ($P = 0.03$), and an intermediate level of recombination ($C = 10$), which is the value that maximized the probability ($P = 0.05$). Arrangement *J* was not considered, because nucleotide variation does not

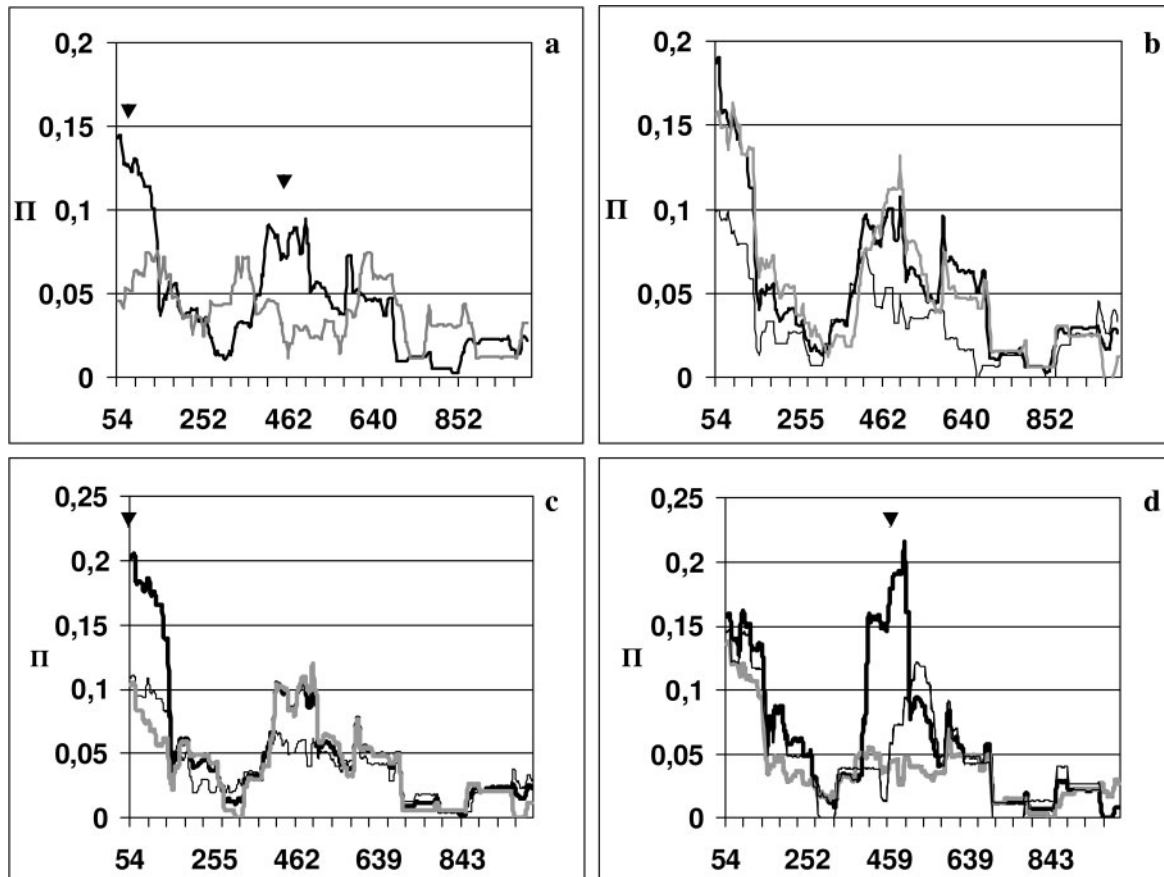


FIG. 3.—Sliding window plots of synonymous variation along the sequenced region of the *Est-A* gene of *D. buzzatii*. (a) Observed (black line) and expected (gray line) nucleotide diversity (π). Expected distribution of nucleotide heterozygosity was obtained after scaling nucleotide synonymous divergence between *D. buzzatii* and the outgroup *D. koepferae*, in terms of the divergence time (see text for further explanation). (b) Nucleotide diversity between (thick black line) and within arrangements *ST* (thick gray line) and *J* (thin black line). Nucleotide diversity between (thick black line) and within allele families classified according to the amino acid variant carried at residue 32 (c), histidine -H- (thick gray line) and asparagine (thin black line), and at residue 151 (d), Alanine (thin black line) and aspartic acid (thick gray line).

appear to be in mutation-drift-flux equilibrium (*sensu* Navarro, Barbadilla, and Ruiz 2000).

A complementary approach to test for heterogeneity in the distribution of silent polymorphism is to calculate a goodness-of-fit test (Kreitman and Hudson 1991), comparing levels of variation among a priori defined regions with those expected under a hypothesis of random distribution of segregating sites. The 1,061 nucleotides were then divided into four regions of equal size, and for each subregion the expected number of segregating sites was calculated on the basis of the number of sites segregating in the sample and the size of each region. Goodness-of-fit tests showed that the distributions of both synonymous ($\chi^2 = 14.8$ $df = 3$, $P = 0.002$) and replacement ($\chi^2 = 8.1$ $df = 3$, $P = 0.044$) variation were significantly heterogeneous, mainly because of an excess of variation in the first half of the sequenced region. In contrast, similar tests for interspecific divergence showed that synonymous ($\chi^2 = 0.24$, $df = 3$, $P = 0.97$) and replacement ($\chi^2 = 1.0$, $df = 3$, $P = 0.80$) diverged sites are distributed homogeneously along the studied gene region.

Graphical Representation of the Patterns of Nucleotide Distribution

The significant results of the average sliding *G* test suggest the presence of regions of elevated polymorphism relative to neutral expectations. Thus, we explored the pattern of distribution of synonymous polymorphism along *Est-A* by means of a sliding window approach using a window of size of 25 base pairs and steps of 1 bp (essentially identical results were obtained varying window size).

The first graph shows the distribution of observed nucleotide heterozygosity at synonymous sites compared to neutral expectations (fig. 3a). Expected values were obtained by dividing the number of synonymous differences between *D. buzzatii* and *D. koepferae* by $T + 1$ (Hudson 1990; Kreitman and Hudson 1991). $T + 1$ (~ 5) was calculated according to the formulae given in Hudson, Kreitman, and Aguadé (1987), after dividing *Est-A* sequence in two regions of equivalent numbers of silent sites and using one *D. koepferae* as the outgroup. T is the time since divergence of the two species measured in units

of $2N$ generations (Hudson 1990). $T + 1$ can be thought of as a scaling factor between polymorphism and divergence (Kreitman and Hudson 1991).

Two peaks of elevated polymorphism can be observed in the graph, one at the 5' end and the other around site 450. As these two peaks are higher than the values predicted (fig. 3a), elevated polymorphism cannot be accounted for by reduced functional constraints (increased neutral mutation rate) in these particular regions (Kreitman and Hudson 1991). Moreover, these two peaks were also evident when divergence between *D. buzzatii* and two more distantly related species *D. borborea* or *D. richardsoni* were used to estimate the expected distribution of synonymous variation (not included in fig. 3a).

These peaks of elevated polymorphism do not seem to be due to linkage of *Est-A* to the inversion polymorphism, because the mean number of synonymous differences between arrangements *ST* and *J* (12.70 ± 1.95 , standard deviation computed by the bootstrap method) is not substantially larger than the mean number of differences between alleles within *ST* (12.58 ± 1.97) and greater than within *J* (8.46 ± 1.54). Moreover, the average number of differences between *ST* and *J* chromosomes at the sites where polymorphism appears to exceed neutral expectations is not higher than within arrangements (fig. 3b; JZ^3 alleles were excluded from this analysis).

The comparison between *D. buzzatii* and *D. koepferae* sequences revealed the presence of 12 shared polymorphisms, 6 synonymous (at sites 9, 45, 81, 450, 456, and 459) and 6 replacement (at sites 14, 32, 77, 94, 421, and 452) (fig. 1), i.e., segregating for the same variants in both species (transpecific polymorphisms). Most of these shared replacement polymorphisms are located in regions of elevated synonymous polymorphism, and may be the putative targets of balancing selection. Particularly, amino acid polymorphisms *Y/F* at 26, *H/N* at 32, and *A/D/N* at 151 may be good candidates because variants are segregating at relatively high frequencies and consist of nonconservative substitutions for at least one of the amino acid physico-chemical properties usually employed to evaluate relatedness in protein sequences (Taylor 1986).

In the sliding window plot the average numbers of synonymous differences between allele families, at or near the putative selected sites, are substantially higher than between sequences within allele families at both amino acid sites 32 (average numbers of pairwise synonymous differences within the allelic classes bearing *H* and *N* are 10.40 ± 1.93 and 8.68 ± 1.82 , respectively, and between classes is 12.05 ± 2.09) (fig. 3c) and 151 (average numbers of pairwise synonymous differences within the allelic classes bearing *D* and *A* are 9.55 ± 1.56 and 11.76 ± 2.10 , respectively and between classes the average is 14.24 ± 2.42) (fig. 3d; sequences with *N* at site 151 were not included). Sliding window graphs in which *Est-A* alleles are classified according to amino acid sites 26 and 31 showed patterns similar to those in figure 3c. This is not unexpected given the association observed between variants at sites 5, 26, and 32 for which 6 of 8 possible haplotypes were present in our sample. In contrast, polymorphic amino acid sites, not shared between species and segregating for variants in intermediate frequencies,

	111111	1111111222	2222222223	33
	11112222	2233556777	7789000122	2244557011
	5912373567	8912145023	7884027906	8918142905
	1259005469			67
Chu109s	YPAQPEPFFT	EREHANVPNH	EFRPQFNPLG	NEVQDGVAFI
Ota26s	.L.....G.....P..P.E.....K.....P.....R			
Trh10sP.A.A.E.K...KV.....PE.....R			
Ber1sY.....S...P.A...E.K...K...D...P.....R			
Ber36s	F.....GS...I..P...E.....V..A.....S.Q.R			
Cat2sG.....P.A..SE.....V.....P.....R			
Cat8sD..Y.G...D...A...E...K...A.....R			
Tir17s	...L.D..Y...SNSD...P...E.....K.....P.S.....R			
Tir20s	.S.....Y...SN...P..A.E...K...A...P.....R			
Ota4s	F.....Y...SN...N...E...K.L.....R			
Chu74s	F...A...Y...SN...LGY D...E...KVL.A.....R			
Chu41s	F...E...Y.G.SN...A.E...K.....L.....R			
Chu53sD..Y...SD...P.A.A.E...KV...D...V.....R			
Chu43sD..Y.D.SD...T.A.A.E...K.LN...L.P.Q...A...R			
Tir7j	F..L.D..Y...SN.....E.....K.....R			
Chu6j	..R.....G.....P.E...KV...A.....I...R			
Chu7j	..R.....G.....LP.E...K.....R			
Chu45j	F...DRRYR...SNS.IL.P...EC.V.KD...A...C.....R			
Chu118jG.....A.E...K.....W			
Chu120j	F..L.D..Y...SN.....A.E...K.....R			
Til1jG.....A...E...KV.....W			
Ota73jG...I..P.A..SE...KV.....R			
Qui7jG.....E..S...K.....W			
Qui19jD.....E...D.V.N.T.....R			
Trh6jG.....E...KV.....R			
Ota72jG.....E...K.....W			
Chu2ZG.....E...KV.....R			
Til3ZD..Y...SNS...P.A.A.E...KV...A...R			
Chu20ZY...SD...P...E...KV.N...P.....G.VR			
Chu38ZY...SD...P..SE...KV.N...P...A...R			
Caf1a	F...D..Y.D...SD...P.T..SE...VL.A.T..P.A...D.R			
Caf1b	F...D..Y.D...SD...P.T..SE...VL.A.T..P.A...D.R			
Caf2D...D.NSD...P.T..SE...SVL...T..P.A...D.R			
Fam1	F.G...D..Y.D...NSD...P.T..SE...VL.A.T..P.A...D.R			
Fam1b	F.G...D..Y.D...NSD...P.T..SE...VL.A.T..P.A...D.R			

FIG. 4.—Amino acid polymorphism at the *Est-A* gene region of *D. buzzatii*. Amino acids identical to the first reference sequence are indicated by a dot. Variants found in *D. koepferae* sequence at the polymorphic sites in *D. buzzatii* are indicated in the last rows. Blocks of sequence transferred from one gene arrangement to another are indicated as shaded boxes.

such as sites 28 and 77, did not show such peaks of elevated synonymous variation (data not shown). Likewise, we did not detect patterns of distribution of synonymous polymorphism similar to the one depicted in figure 3d when sequences were grouped according to amino acid variants segregating at sites close to residue 151 (results not shown).

Replacement Polymorphism and Amino Acid Variation

Ample replacement polymorphism was detected in our sample (fig. 4). At the level of the amino acid sequence some sites segregate for more than two amino acids, which in general were due to substitutions at different positions of the same codon, except codons 32, 73, and 215, in all of which the presence of three amino acid variants involved more than one substitution at the same position. Half of the amino acid substitutions were nonconservative changes for different properties such as charge, polarity, and volume. Replacement heterozygosity is extraordinarily high in the 5' half and in certain parts similar to nonsynonymous divergence between *D. buzzatii* and its sibling *D. koepferae*. In contrast, in the 3' half heterozygosity is markedly lower than interspecific divergence (see above).

Most sites segregating for replacement variants are singletons or low frequency variants, suggesting that most amino acid variants may be slightly deleterious.

Discussion

DNA Polymorphism at *Est-A*

The amount of nucleotide variation in the *Est-A* gene of *D. buzzatii* is relatively high when compared to available data for noncoding sequences of the breakpoints of inversion *J* (Cáceres, Puig, and Ruiz 2001). Moreover, when our estimates are compared to available data sets in other species, *Est-A* stands within the group of genes showing high levels of synonymous variation such as *Est-6* (Karotam, Boyce, and Oakeshott 1995), *boss* (Ayala and Hartl 1993), *Rh3* (Ayala, Chang, and Hartl 1993), and *Pgm* (Verrelli and Eanes 2000) in *D. simulans*; *Amy-d* in *D. melanogaster* (Inomata et al. 1995); *Adhr* in *D. pseudoobscura* (Schaeffer et al. 2001), and *Pgi* in the plant *Leavenworthia stylosa* (Filatov and Charlesworth 1999). Similarly, variation at nonsynonymous sites is strikingly high, only comparable to *Ref(2)P* a gene that confers viral resistance (Wayne, Contamine, and Kreitman 1996) in *D. melanogaster*, and two rapidly evolving anonymous genes, *Anon-1G5* and *Anon-1E9* in *D. simulans* (Schmid et al. 1999).

Low functional constraint can only partially account for the high level of nucleotide diversity at *Est-A*, because divergence between *D. buzzatii* and *D. koepferae* at both synonymous and (mainly) replacement sites ($K_S = 0.153$ and $K_R = 0.046$) is also relatively high when compared to published data for *Xdh* (Rodríguez-Trelles, Alarcón, and Fontdevila 2000) ($K_S = 0.123$ and $K_R = 0.022$). Recent or ancient population subdivision may also be ruled out as an alternative explanation because geographic structure is not apparent either in the clustering of alleles in the NJ tree (fig. 2) or in the estimates of genetic differentiation among populations grouped according to geographic origin. Furthermore, South American *D. buzzatii* populations are only structured for the inversion polymorphism and certain adaptive allozyme polymorphisms (Rodríguez et al. 2000), whereas mt-DNA (Rossi et al. 1996) failed to reveal geographic subdivision.

The third possible explanation is balancing selection. Population genetic theory predicts a window of elevated silent polymorphism near the selected site, the width of the window depending on the local recombination rate (Hudson 1990), leading to a local departure from the predicted correlation between silent polymorphism and divergence.

The result of the average sliding *G* test seems to give support to this hypothesis. The presence of two apparent peaks of synonymous variation, which seems to be coincident with nonconservative amino acid substitutions in the sliding window plot (fig. 3a), is also consistent with this explanation. However, it is worth noting that the graphs should only be considered as an illustration of the McDonald (1998) test. The first peak, near the 5' end of the region sequenced, would be associated with amino acid polymorphisms at sites 26 and/or 32, whereas the second peak seems to be centered on amino acid site 151.

Moreover, the six alleles carrying alanine (A) at codon 151 (Cat8, Tir20, Chu74, Chu45, Til3, and Ber36) appear ancestral in the NJ tree, whereas 10 of the 16 alleles that carry glutamic acid (E) at codon 31 and histidine (H) at codon 32 (Chu6, Chu7, Trh6, Chu2, Til1, Qui19, Chu118, Ota72, Ota73, and Qui7), which all happen to be on *J* or *JZ*³ chromosomes, form a cluster in NJ (fig. 2). Peaks of elevated polymorphism appear to be due to a greater average number of pairwise differences between alleles carrying different amino acids at sites 31 and/or 32 and 151 than between alleles within families (fig. 3c, 3d).

However, there are two issues that cast some doubts on our conclusion that the peaks of elevated polymorphism may be the footprints of balancing selection. The first is related to the problem of the use of several recombination rates, which raises the issue of multiple comparisons in the average sliding *G* test. If we correct for multiple comparisons using Bonferroni's correction, which in the present case is overconservative due to nonindependence, none of the tests would be significant. The second issue is that most of our analyses are applied on intra-allelic subsamples that are nonrandom subsamples of the total. In particular, under the hypothesis of balancing selection, we expect a lower polymorphism within allelic classes compared to the total sample and genetic differentiation between allelic classes (Innan and Tajima 1999). However, these effects are expected to vanish with the genetic distance from the selected site in a model with recombination, as can be observed in figure 3c and 3d. In fact, the values of the ratio of the sum of the average number of pairwise differences within allelic classes, as compared to the average pairwise number of differences between all sequences, were slightly lower than 2 (1.8 and 1.9 for sites 32 and 151, respectively), which is the expected value under neutrality and moderate rates of recombination (Innan and Tajima 1999). It should be noted, however, that under neutrality a similar intra-allelic effect to that under balancing selection is also expected (Innan and Tajima 1997, 1999).

The hypothesis that peaks of elevated silent polymorphism may be the footprints of balancing selection is also supported by the fact that polymorphic variants at the putative targets of selection, along with variants at neighboring sites, appear to be rather old because they are segregating for the same variants in *D. koepferae*, i.e., shared or transpecific polymorphisms. Well-known examples of polymorphisms shared between species are the *Mhc* in mammals (reviewed in Nei and Kumar 2000), self-incompatibility genes in plants (Charlesworth and Awadalla 1998), and *Pgi* in *L. stylosa* (Filatov and Charlesworth 1999). In *Drosophila* transpecific polymorphisms have been reported in the *D. simulans* species complex (Kliman et al. 2000), in *D. pseudoobscura* and close relatives (Wang et al. 1997), and in the *virilis* subgroup (Hilton and Hey 1997; Vieira 2002).

One possible explanation for shared polymorphisms in closely related species is recurrent mutation. However, the probability of 12 shared polymorphisms (6 synonymous and 6 nonsynonymous) conditional on the number of polymorphic sites in *D. buzzatii* (126, 64 synonymous and 62 replacement) and *D. koepferae* (30, 15 synonymous

and nonsynonymous) is quite low ($P < 10^{-7}$; $P = 8 \times 10^{-4}$ and $P = 10^{-6}$, for total, synonymous, and replacement polymorphism, respectively). Yet, this method is based on the assumption of constant mutation rates, which in the case of *Est-A* may have weak support, given the presence of 8 sites segregating for 3 variants in *D. buzzatii*. However, the presence of 8 sites with two hits in a sequence of 1,061 nucleotides conditional on a total of 137 mutations is not an unexpected event for a Poisson process (goodness of fit $\chi^2 = 0.58$, $df = 2$, $P = 0.75$). A similar conclusion can be reached for the 5 synonymous ($\chi^2 = 5.41$, $df = 2$, $P = 0.07$) and 3 nonsynonymous ($\chi^2 = 2.74$, $df = 2$, $P = 0.25$) sites with three variants.

Still, the number of sites with three variants may be an underestimate of sites with multiple hits. Under the Jukes and Cantor model, which assumes homogeneous rates among sites, one-third of double hits would be undetected. Hence, the number of sites with undetected double hits may be 4. If we add this quantity to the number of observed sites with three variants, the distribution of mutations along the sequence does not depart from the expectations of a Poisson process ($\chi^2 = 2.83$, $df = 2$, $P = 0.24$). However, it may be argued that this effect would be stronger if there is a transition/transversion bias. For instance, sites with two successive transitions would show only two variants and would thus not be detected. Assuming the Kimura 2 parameter model and using the transition/transversion ratio ($\kappa = ts/tv$), we can calculate the relative probability of such undetected double hits. Given that there are two hits on the site the relative probability of undetected double transitions is $((\kappa / 2 + \kappa)^2)$, and $(2/(2 + \kappa)^2)$ for double transversions. The remaining double hits that should lead to three detectable variants occur with a probability equal to $(2 + 4\kappa) / (2 + \kappa)^2$. Therefore, the expected ratio of undetected/detected double hits is $(2 + \kappa^2) / (2 + 4\kappa)$. If we use in the latter formula our estimate of $\kappa = 1.2$, calculated from our data, we obtain an estimate of the proportion of undetected double hits of 0.51. Again, as with Jukes and Cantor's model, the distribution of mutations does not depart from the expectations of a Poisson process ($\chi^2 = 2.94$, $df = 2$, $P = 0.23$). Yet, the mutational hypothesis cannot be disregarded because we cannot rule out that under more complex models (4–12 parameters models) the number of observed double hits would depart from the expectations of a Poisson process.

We further investigated the issue of rate heterogeneity by means of the Akaike (1974) information criterion (AIC) using the program Modeltest (Posada and Crandall 1998). The AIC is a useful measure that allows comparisons among alternative models of sequence evolution by means of goodness-of-fit tests, in which AIC is the maximum value of the likelihood function for a specific model using a number of independently adjusted parameters (Posada and Crandall 2001). On the one hand, the best model fitting our data for fourfold degenerate sites did not include the gamma distribution (a measure of rate heterogeneity among sites) but included different rates of transitions and transversions. On the other hand, the alpha parameter of the gamma distribution was included among the param-

eters of the best models fitting the data for 0-fold ($\alpha = 0.43$) and twofold ($\alpha = 0.71$) degenerate sites.

An alternative explanation for transpecific polymorphisms is the persistence of neutral ancestral variants in populations of large size. However, population genetic theory predicts that the expected time for the loss of a shared ancestral neutral polymorphism is 1.7 N generations (Clark 1997). Available information suggests a relatively old divergence between *D. buzzatii* and *D. koepferae*. Cáceres, Puig, and Ruiz (2001) reported that *D. buzzatii* and *D. martensis* shared their last common ancestor around 5.8 Myr on the basis of divergence at the breakpoints of inversion *J*. In addition, in a recent phylogenetic study of the *D. buzzatii* complex using the *Xdh* locus (Rodríguez-Trelles, Alarcón, and Fontdevila 2000), the number of divergent synonymous sites between *D. martensis* and *D. buzzatii* was 86.5, while 62 separated *D. buzzatii* and *D. koepferae*. This gives an estimate of ~ 4 Myr $[(62 \times 5.8) / 86.5]$ for the time of divergence between the latter two species and suggests that sufficient time has passed for most shared neutral ancestral polymorphisms to have become fixed.

Interspecific gene flow can also account for the presence of shared polymorphisms. *D. buzzatii* and *D. koepferae* are reproductively isolated by partial ecological isolation (Fanara, Fontdevila, and Hasson 1999) and postmating barriers due to sterility of male hybrid progeny (Naveira and Fontdevila 1986). Thus, hybrid females are a potential bridge for introgression. Recently, we have detected transpecific polymorphisms in the *Xdh* gene (R. Piccinali, M. Aguadé, and E. Hasson, unpublished results), which is also linked to the polymorphic second chromosome. This observation suggests that ancient or recent gene flow may be a plausible explanation for the extensive sharing of variants in the pair *D. buzzatii*–*D. koepferae*. Nevertheless, the distribution of silent polymorphism and divergence along the sequence of *Est-A*, compared to *Xdh*, suggests that the pattern observed in the present study is a particular feature of *Est-A*.

In order to disentangle the role of ancient or recent gene flow, balancing selection, recurrent mutation, persistence of ancestral polymorphisms, and the role played by adaptive inversion polymorphisms, it is necessary to compare nucleotide variation between gene regions linked to second chromosome inversions versus genes located in inversion free chromosomes.

Patterns of Nucleotide Variation Within and Between Arrangements

In general terms, the distribution of nucleotide variation among arrangements is in good agreement with expectations of population models of gene flux in inversion heterokaryotypes (Navarro et al. 1997). Genetic exchange between arrangements in the *Est-A* region has led to an extensive sharing of polymorphisms, as usually observed in genes located in central positions of inversions (Bénassi et al. 1993; Popadic, Popadic, and Anderson 1995; Hasson and Eanes 1996; Rozas et al. 1999). Conversely, in the neighborhood of the breakpoints of naturally occurring inversions, monophyly is unequivocal,

because genetic exchange has not erased the genealogical relationships between alleles contained in different arrangements (Wesley and Eanes 1994; Popadic and Anderson 1995; Babcock and Anderson 1996; Hasson and Eanes 1996; Munté, Aguadé, and Segarra 2000; Navarro-Sabaté, Aguadé, and Segarra 1999; Rozas et al. 1999). Unique origin is also well supported, even in cases in which transposable elements appear to have been involved in the origin of inversions such as *D. melanogaster In(2L)t* (Andolfatto, Wall, and Kreitman 1999) and *D. buzzatii* inversion *J* (Cáceres et al. 1999).

The comparison of patterns of nucleotide variation at *Est-A* and the breakpoints of inversion *J* clearly shows that the presence of the inversion polymorphism differentially affects gene regions located in the center and close to the breakpoints. Genetic exchange due to gene conversion and/or double crossovers tends to homogenize genetic variation among arrangements in the middle of an inversion, whereas in the breakpoints neutral variation evolves largely independently in inverted and ancestral chromosomes. In *D. buzzatii* available evidence suggest that the distribution of nucleotide variation within and between gene arrangements at both the breakpoints and *Est-A* is not in agreement with those predicted by theoretical work modeling the population genetics of inversions (Navarro, Barbadilla, and Ruiz 2000), in particular *2J*, a very common inversion that appears to be 1 Myr old (Cáceres, Puig, and Ruiz 2001). The lower nucleotide diversity, the greater linkage disequilibrium, the clustering of 10 derived alleles (*9J* and *1JZ*²) in the NJ tree, and, especially the extremely low sequence variation at the breakpoints reported in Cáceres, Puig, and Ruiz (2001) are not compatible with the idea of inversion *J* being an old lineage.

Possible explanations for the low level of variation in *J* chromosomes, such as recent origin and low effective population size, can be ruled out. A plausible alternative may be that the contemporary global frequency of arrangement *J* is not representative of its historical effective population size, but a reflection of a relatively recent expansion. The footprint of such an event may be an excess of rare variants, which is consistent with the results of Fu's test. But, the significant result obtained when this test was applied to the subsample of *ST* chromosomes, seems to indicate that the entire species had passed through a population expansion. However, the results of these tests should be interpreted with caution given that recombination tends to produce additional haplotypes. In this sense, the estimated recombination parameters suggest, assuming equal recombination rates in both arrangements, that the historical effective population size of *ST* has been at least 7–8 times larger than *J*, even though present average frequencies of both arrangements are almost equal. Thus, the most likely explanation seems to be that the frequency of arrangement *J* has risen in association with a recent expansion of the species.

Data on nucleotide sequence variation for actual neutral markers uncovering other regions along the rearranged segment are needed to understand the complex evolutionary history of the inversion polymorphism of *D. buzzatii*. The recently reported physical map based on sequenced RAPD markers (Laayouni, Santos, and Font-

devila 2000) may provide the tools to clarify the issues raised in this article.

Acknowledgments

We thank A. Fontdevila and members of Grupo de Biología Evolutiva (Universidad Autónoma de Barcelona, Spain) for hospitality and a pleasant atmosphere during a sabbatical stay. We wish to thank W. F. Eanes, M. Aguadé, A. Fontdevila, M. Santos, J. S. F. Barker, R. Piccinali, and H. Laayouni for critical reading of earlier versions of this paper and W. F. Eanes for revising the last version. We wish to thank two anonymous reviewers for valuable suggestions and S. Rossi, A. Navarro, G. Boente-Boente, R. Lombardo, and J. Vieira for helpful discussions and advice. The help of M. Peiró with in situ hybridizations is gratefully acknowledged. The first draft of this paper was written while E.H. was a fellow of the Ministerio de Educación, Cultura y Deporte of Spain. This work was supported by Universidad de Buenos Aires, CONICET, and Fundación Antorchas grants. E.H. is member of Carrera del Investigador Científico (CONICET). The results reported herein are part of the Ph.D. thesis that G.G. is doing with the financial support of a FOMEC fellowship.

Literature Cited

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Contr.* **19**:716–723.
- Andolfatto, P., F. DePaulis, and A. Navarro. 2001. Inversion polymorphism and nucleotide variability in *Drosophila*. *Genet. Res.* **77**:1–8.
- Andolfatto, P., and M. Kreitman. 2000. Molecular variation at the *In(2L)t* proximal break point site in natural populations of *Drosophila melanogaster* and *D. simulans*. *Genetics* **154**:1681–1691.
- Andolfatto, P., J. D. Wall, and M. Kreitman. 1999. Unusual haplotype structure at the proximal break point of *In(2L)t* in a natural population of *Drosophila melanogaster*. *Genetics* **153**:1297–1311.
- Aquadro, C. F., A. L. Weaver, S. W. Schaeffer, and W. W. Anderson. 1991. Molecular evolution of inversions in *Drosophila pseudoobscura*: the *amylase* gene region. *Proc. Natl. Acad. Sci. USA* **88**:305–309.
- Ayala, F. J., B. S. Chang, and D. L. Hartl. 1993. Molecular evolution of the Rh3 gene in *Drosophila*. *Genetica* **92**:23–32.
- Ayala, F. J., and D. L. Hartl. 1993. Molecular drift of the *bride-of-sevenless* (*boss*) gene in *Drosophila*. *Mol. Biol. Evol.* **10**:1030–1040.
- Babcock, C. S., and W. W. Anderson. 1996. Molecular evolution of the sex-ratio inversion complex in *Drosophila pseudoobscura*: analysis of the *Esterase-5* gene region. *Mol. Biol. Evol.* **13**:297–308.
- Bénassi, V., S. Aulard, S. Mazeau, and M. Veuille. 1993. Molecular variation of *Adh* and *P6* genes in an African population of *Drosophila melanogaster* and its relation to chromosomal inversions. *Genetics* **134**:789–799.
- Bénassi, V., F. DePaulis, G. Meghlaoui, and M. Veuille. 1999. Partial sweeping at the *Fbp2* locus in a West African population of *Drosophila melanogaster*. *Mol. Biol. Evol.* **16**:347–353.
- Betrán, E., J. Rozas, A. Navarro, and A. Barbadilla. 1997. The estimation of the number and the length distribution of gene

- conversion tracts from population DNA sequence data. *Genetics* **146**:89–99.
- Cáceres, M., M. Puig, and A. Ruiz. 2001. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res.* **11**:1353–1364.
- Cáceres, M., J. M. Ranz, A. Barbadilla, M. Long, and A. Ruiz. 1999. Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285**:415–418.
- Charlesworth, D., and P. Awadalla. 1998. Flowering plant self-incompatibility: the molecular population genetics of *Brassica* S-loci. *Heredity* **81**:1–9.
- Clark, A. G. 1997. Neutral behavior of shared polymorphism. *Proc. Natl. Acad. Sci. USA* **94**:7730–7734.
- DePaulis, F., L. Brazier, S. Mousset, A. Turbé, and M. Veuille. 2000. Selective sweep near in an African population of *Drosophila melanogaster*. *Genet. Res.* **76**:149–158.
- DePaulis, F., L. Brazier, and M. Veuille. 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.
- East, P., A. Graham, and G. Whittington. 1990. Molecular isolation and preliminary characterization of a duplicated esterase locus in *Drosophila buzzatii*. Pp. 389–406. in J. S. F. Barker, W. T. Starmer and R. J. MacIntyre, eds. *Ecological and Evolutionary Genetics of Drosophila*. Plenum Press, New York.
- Fanara, J. J., A. Fontdevila, and E. Hasson. 1999. Oviposition preference, viability, developmental time and body size in the cactophilic sibling species *Drosophila buzzatii* and *Drosophila koepferae* in association to their natural hosts. *Evo. Ecol.* **13**:173–190.
- Filatov, D. A., and D. Charlesworth. 1999. DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia Pgi* locus. *Genetics* **153**:1423–1434.
- Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**:915–925.
- Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**:693–709.
- Hasson, E., and W. Eanes. 1996. Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. *Genetics* **144**:1565–1575.
- Hasson, E., C. Rodríguez, J. J. Fanara, H. Naveira, O. A. Reig, and A. Fontdevila. 1995. The evolutionary history of *Drosophila buzzatii*. XXVI. Macrogeographic patterns in the inversion polymorphism in New World populations. *J. Evol. Biol.* **8**:369–384.
- Hey, J., and J. Wakeley. 1997. A coalescent estimator of the population recombination parameter. *Genetics* **145**:833–846.
- Hilton, H., and J. Hey. 1997. A multilocus view of speciation in the *Drosophila virilis* species group reveals complex histories and taxonomic conflicts. *Genet. Res.* **70**:185–194.
- Hudson, R. J. 1987. Estimating the genetic recombination parameter of a finite population model without selection. *Genet. Res.* **50**:245–250.
- Hudson, R. J., and N. L. 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., 1990. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**:1–44.
- Hudson, R. R., D. D. Boos, and N. L. Kaplan. 1992. A statistical test for detecting population subdivision. *Mol. Biol. Evol.* **9**:138–151.
- 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., M. Slatkin, and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**:583–589.
- Innan, H., and F. Tajima. 1997. The amounts of nucleotide variation within and between allelic classes, and the reconstruction of common ancestral sequence. *Genetics* **147**:1431–1444.
- . 1999. The effect of selection on the amounts of nucleotide variation within and between allelic classes. *Genet. Res.* **73**:15–28.
- Inomata, N., H. Shibara, E. Okuyama, and T. Yamazaki. 1995. Evolutionary relationships and sequence variation of α -amylase variants encoded by duplicated genes in the *Amy* locus of *Drosophila melanogaster*. *Genetics* **141**:237–244.
- Karotam, J., T. M. Boyce, and J. G. Oakeshott. 1995. Nucleotide variation at the hypervariable esterase 6 isozyme locus of *Drosophila simulans*. *Mol. Biol. Evol.* **12**:113–122.
- Kliman, R. M., P. Andolfatto, J. A. Coyne, F. DePaulis, M. Kreitman, A. J. Berry, J. McCarter, J. Wakeley, and J. Hey. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**:1913–1931.
- Kreitman, M., and R. R. Hudson. 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.
- Krimbas, C. B., and J. R. Powell. 1992. Introduction. Pp. 1–52 in C. B. Krimbas and J. R. Powell, eds. *Drosophila* inversion polymorphism. CRC Press, Boca Raton, FL.
- Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: Molecular evolutionary genetics analysis software. Arizona State University, Tempe, Arizona.
- Laayouni, H., M. Santos, and A. Fontdevila. 2000. Toward a physical map of *Drosophila buzzatii*. Use of randomly amplified polymorphic DNA polymorphisms and sequence tagged site landmarks. *Genetics* **156**:1797–1816.
- Ladeveze, V., S. Aulard, N. Chaminade, G. Periquet, and F. Lemeunier. 1998. Hobo transposons causing chromosomal break points. *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* **265**:1157–1159.
- Lewontin, R. C. 1995. The detection of linkage disequilibrium in molecular sequence data. *Genetics* **140**:377–388.
- Lytte, T. W., and D. S. Haymer. 1992. The role of transposable element hobo in the origin of endemic inversions in wild populations of *Drosophila melanogaster*. *Genetica* **86**:113–126.
- 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.
- Montgomery, E. A., B. Charlesworth, and C. H. Langley. 1987. A test of the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet. Res.* **49**:31–41.
- Munté, A., M. Aguadé, and C. Segarra. 2000. Nucleotide variation at the yellow gene region is not reduced in *Drosophila subobscura*: a study in relation to chromosomal polymorphism. *Mol. Biol. Evol.* **17**:1942–1955.
- Navarro, A., A. Barbadilla, and A. Ruiz. 2000. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in *Drosophila*. *Genetics* **155**:685–698.
- Navarro, A., E. Betrán, A. Barbadilla, and A. Ruiz. 1997. Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics* **146**:695–709.
- Navarro-Sabaté, A., M. Aguadé, and C. Segarra. 1999. The relationship between allozyme and chromosomal polymor-

- phism inferred from nucleotide variation at the *AcpH-1* gene region of *Drosophila subobscura*. *Genetics* **153**:871–889.
- Naveira, H., and A. Fontdevila. 1986. The evolutionary history of *Drosophila buzzatii*. XII. The genetic basis of sterility in hybrids between *D. buzzatii* and its sibling *D. koepferae* from Argentina. *Genetics* **114**:841–857.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nei, M., and S. Kumar. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford.
- Oakeshott, J. G., C. Claudianos, R. J. Russell, and G. C. Robin. 1999. Carboxyl/cholinesterases: a case study of the evolution of a successful multigene family. *Bioessays* **21**:1031–1042.
- Popadic, A., and W. W. Anderson. 1995. Evidence for gene conversion in the amylase multigene family of *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **12**:564–572.
- Popadic, A., D. Popadic, and W. W. Anderson. 1995. Interchromosomal exchange of genetic information between gene arrangements of the third chromosome of *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **12**:938–943.
- Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- . 2001. Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). *Mol. Biol. Evol.* **18**:897–906.
- Powell, J. R. 1997. *Progress and prospects in evolutionary biology: the Drosophila model*. Oxford University Press, New York.
- Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* **43**:223–225.
- Rodríguez, C., R. Piccinali, E. Levy, and E. Hasson. 2000. Contrasting population structures using allozyme variation and the inversion polymorphism in *Drosophila buzzatii*. *J. Evol. Biol.* **13**:976–984.
- . 2001. Gametic associations between inversion and allozyme polymorphisms in *Drosophila buzzatii*. *J. Hered.* **92**:382–391.
- Rodríguez-Trelles, F., L. Alarcón, and A. Fontdevila. 2000. Molecular evolution and phylogeny of the *buzzatii* complex (*Drosophila repleta* group): a maximum-likelihood approach. *Mol. Biol. Evol.* **17**:1112–1122.
- Rossi, M., Barrio, E., A. Latorre, J. E. Quezada-Diaz, E. Hasson, A. Moya, and A. Fontdevila. 1996. The evolutionary history of *Drosophila buzzatii*. XXX. Mitochondrial DNA polymorphism in original and colonizing populations. *Mol. Biol. Evol.* **13**:314–323.
- Rozas, J., and R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**:174–175.
- Rozas, J., C. Segarra, G. Ribó, and M. Aguadé. 1999. Molecular population genetics of the *rp49* gene region in different chromosomal inversions of *Drosophila subobscura*. *Genetics* **151**:189–202.
- Saitou, N., and M. Nei. 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- Schaeffer, S. W., C. Scott Walthour, D. M. Toleno, A. T. Olek, and E. L. Miller. 2001. Protein variation in ADH and ADH-related in *Drosophila pseudoobscura*: linkage disequilibrium between single nucleotide polymorphisms and protein alleles. *Genetics* **159**:673–687.
- Schmid, K. J., L. Nigro, C. H. Aquadro, and D. Tautz. 1999. Large number of replacement polymorphisms in rapidly evolving genes of *Drosophila*: implications for genome-wide surveys DNA polymorphism. *Genetics* **153**:1717–1729.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
- Taylor, W. R. 1986. The classification of amino acid conservation. *J. Theor. Biol.* **119**:205–218.
- Verelli, B. C., and W. F. Eanes. 2000. Extensive amino acid polymorphism at the *Pgm* locus is consistent with adaptive protein evolution in *Drosophila melanogaster*. *Genetics* **156**:1737–1752.
- Vieira, J. 2002. Two divergent species of the *virilis* group, *Drosophila littoralis* and *D. virilis*, share a replacement polymorphism at the fused locus. *Mol. Biol. Evol.* **19**:579–581.
- Wang, R. L., J. Wakeley, and J. Hey. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147**:1091–1106.
- Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**:256–276.
- Wayne, M. L., D. Contamine, and M. Kreitman. 1996. Molecular population genetics of *ref(2)P*, a locus which confers viral resistance in *Drosophila*. *Mol. Biol. Evol.* **13**:191–199.
- Wesley, C., and W. F. Eanes. 1994. Isolation and analysis of the breakpoint sequences of chromosome inversion *In(3L)Payne* in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **91**:3132–3136.

Herve Philippe, Associate Editor

Accepted November 8, 2002