The Evolutionary History of *Drosophila buzzatii*. XXXV. Inversion Polymorphism and Nucleotide Variability in Different Regions of the Second Chromosome

Hafid Laayouni, Esteban Hasson,*† Mauro Santos,* and Antonio Fontdevila**

*Grup de Biologia Evolutiva, Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain; and †Departamento de Ecología Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

Inversions are portions of a chromosome where the gene order is reversed relative to a standard reference orientation. Because of reduced levels of recombination in heterokaryotypes, inversions have a potentially important effect on patterns of nucleotide variability in those genomic regions close to, or included in, the inverted fragments. Here we report sequence variation at three anonymous regions (STSs) located at different positions in relation to second-chromosome inversion breakpoints in 29 isochromosomal lines derived from an Argentinean population of *Drosophila buzzatii*. In agreement with previous findings in Drosophila, gene flux (crossing over and/or gene conversion) between arrangements seems to appreciably increase as we approach the middle sections of inversion *2j*, and patterns of nucleotide variability within, as well as genetic differentiation between chromosome arrangements, are comparable to those observed at the molecular marker outside the inverted fragments. On the other hand, nucleotide diversity near the proximal breakpoint of inversion *2j* is reduced when contrasted with that found at the other regions, particularly in the case of derived inverted chromosomes. Using the data from the breakpoint, we estimate that the inversion polymorphism is approximately 1.63 *N* generations old, where *N* is the effective population size. An excess of low-frequency segregating polymorphisms is detected; mostly in the ancestral *2st* arrangement and probably indicating a population expansion that predates the coalescent time of inversion *2j*. Heterogeneity in mutation rates between the markers linked to the inversions may be sufficient to explain the different levels of nucleotide diversity observed. When considered in the context of other studies on patterns of variation relative to physical distance to inversion breakpoints, our data appear to be consistent with the conclusion that inversions are unlikely to be ''long-lived'' balanced polymorphisms.

Introduction

The study of chromosomal inversion polymorphisms in the genus Drosophila has a long and venerable tradition in population genetics. First considered as neutral genetic markers in the foundation book of the modern synthesis (Dobzhansky 1937), extensive sampling efforts in southern California over 6 years from 1939 to 1946 highlighted seasonal cycles in chromosomal arrangement frequencies of *D. pseudoobscura* and brought Dobzhansky (1947) to a fundamental turning point in his previous thinking that has pervaded much subsequent work. Experimental results with population cages, detection of altitudinal and latitudinal clines of inversion frequencies, uncovering of linkage disequilibria between allozyme loci and polymorphic inversions, and fitness-components studies in the wild have convinced many researchers that inversion polymorphisms are maintained by strong balancing selection, most probably heterosis (see Krimbas and Powell 1992 for a review). Because the basic mechanistic effect of paracentric inversions (the most common) in Drosophila is to reduce the frequency of recombination during gametogenesis in inversion heterozygotes (heterokaryotypes), several investigators have proposed that inversions allow preserving allelic combinations that have been built up by selection and, hence, reduce the frequency of harmful products of crossover events (the so-called ''recombination load'' [see Charlesworth and Barton 1996]).

Key words: *Drosophila buzzatii*, inversion polymorphism, nucleotide variability, sequence-tagged sites.

E-mail: hafid.laayouni@uab.es.

Mol. Biol. Evol. 20(6):931–944. 2003
DOI: 10.1093/molbev/msg099

There is, however, little consensus about the extent or importance of genetic interactions in the evolutionary process after more than half a century of debate (Whitlock et al. 1995). On the other hand, the advent of molecular population genetics has provided compelling evidence that the level of nucleotide diversity is positively correlated with the regional rate of recombination (Begun and Aquadro 1992; Aguadé and Langley 1994). This pattern can be explained by variation-reducing selection on neutral variability in different recombination environments; namely, ''selective sweep'' events (Maynard Smith and Haigh 1974; Kaplan, Hudson, and Langley 1989) or recurrent selection against deleterious mutations (''background selection'' [Charlesworth, Morgan, and Charlesworth 1993]). Conversely to the footprint left by these directional selection processes, the presence of an old balanced polymorphism will cause, in the long term, an increase of neutral diversity at closely linked sites (Hudson and Kaplan 1988; Hudson 1990). Because differences in recombination rates outside and within the chromosome segments covered by inversions are expected in heterokaryotypes (see Roberts 1976), the long-standing assumption of strong balancing selection acting on inversion polymorphisms can now be critically tested by analyzing patterns of nucleotide variation at molecular markers located in different regions relative to an inverted chromosomal fragment. To be more specific, enhanced levels of total nucleotide variability and differentiation between chromosome arrangements are expected at molecular markers closely linked to inversion breakpoints if the polymorphism has a long history of balancing selection (Strobeck 1983; Navarro, Barbadilla, and Ruiz 2000; Andolfatto, Depaulis, and Navarro 2001).

Previous studies of loci closely linked to inversion breakpoints in Drosophila reveal patterns of nucleotide

variation consistent with reduced levels of gene flux (crossing over and/or gene conversion [see Navarro et al. 1997]) near breakpoints relative to loci located at longer distances (reviewed in Andolfatto, Depaulis, and Navarro 2001). The situation, however, is more complicated in the presence of a complex inversion system because of considerable gene flux between arrangements, as in the case of the *rp49* gene region of chromosome *O* in *D. subobscura* (Rozas et al. 1999) or the *Est-5* and *Hsp83* genes on the right arm of the X chromosome in *D. pseudoobscura*, which segregates for the *sex ratio* inversion system (Kovacevic and Schaeffer 2000). By focusing on those works that have studied levels of nucleotide variation in chromosomal elements free of complex inversion systems (e.g., inversions *In(3L)Payne* and *In(2L)t* in *D. melanogaster*) another clear pattern emerges; namely, nucleotide diversity at markers located close to inversion breakpoints is substantially lower in the derived inverted chromosome relative to that found in the standard one (Wesley and Eanes 1994; Hasson and Eanes 1996; Andolfatto, Wall, and Kreitman 1999; Depaulis, Brazier, and Veuille 1999; Depaulis et al. 2000), and total nucleotide variability at those markers is generally lower than that found at loci far apart (>1,000 kb [see fig. 2 in Andolfatto, Depaulis, and Navarro 2000]). Overall, the data are inconsistent with these two inversions being an ancient polymorphism maintained by strong balancing selection. Theoretical work modeling inversions as balanced polymorphisms concludes that an increase of neutral diversity at sites closely linked to breakpoints is in fact expected, provided that balancing selection does not fluctuate with time and the age of inversions is greater than N generations, where N is the effective population size (Navarro, Barbadilla, and Ruiz 2000). The age of *In(3L)Payne* is approximately 0.36 Myr (Hasson and Eanes 1996) or approximately 1.2 N generations, assuming an effective population size of approximately 3×10^6 (Kreitman 1983) and about 10 generations per year in *D. melanogaster*, and the age of *In(2L)t* is approximately 0.3N generations (Andolfatto, Wall, and Kreitman 1999). Therefore, it could be claimed that the life span of these inversions is shorter than the required time to achieve equilibrium.

Here we report patterns of DNA polymorphism and interspecific divergence at three anonymous sequence-tagged sites (STSs) of the second chromosome of *D. buzzatii*. This semicosmopolitan species provides a valuable model system for several reasons. First, as with most species in the *repleta* group, its karyotype consists of five acrocentric chromosomes and a dot chromosome (Wasserman 1982), and most natural populations are polymorphic for chromosomal arrangements *2st* and *2j* that differ by the single paracentric inversion *2j* (Fontdevila et al. 1981, 1982; Barker et al. 1985; Hasson et al. 1995), which captures approximately 26% of the chromosome element and can be considered as an average Drosophila inversion (Federer, Steel, and Wallace 1967). Second, a number of studies have shown that these chromosomal arrangements differentially affect developmental time and adult body size, thus suggesting antagonistic pleiotropic effects on fitness-related traits (see Betrán, Santos, and Ruiz 1998; Fernández-Iriarte and Hasson 2000). Third, inversion *2j* has recently been estimated to be approximately1 Myr old (Cáceres, Puig, and Ruiz 2001), which suggests that the polymorphism may be old enough for the DNA variability to have reached the mutation-drift equilibrium. Fourth, a physical map based on RAPD polymorphic fragments and STSs is already available (Laayouni, Santos, and Fontdevila 2000). Fifth and finally, there is a relatively wealthy biogeographical knowledge of the *repleta* group. Thus, Carson and Wasserman (1965), Vilela, Sene, and Pereira (1980), and Fontdevila et al. (1982) suggested the Argentinean Chaco as the most likely center of origin of *D. buzzatii*. Hence, we have chosen a population from Chumbicha that belongs to the Argentinean Chaco Dominium (Cabrera 1976) and likely represents an ancestral population closer to mutation-drift equilibrium, enabling the selective forces determining DNA sequence variation to be more easily elucidated. This study aims therefore to address the following questions: (1) How are nucleotide diversity levels related to distance from inversion breakpoints in the second chromosome of *D. buzzatii*? (2) What are the levels of gene flux between gene arrangements? (3) What is the presumably selective history of inversion *2j*? Adaptive and demographic or historic factors can in principle be distinguished since the latter are expected to have a genome-wide effect, whereas indirect selection will affect a relatively small fraction of the genome.

Materials and Methods

Drosophila Stocks

D. buzzatii flies analyzed in the present study were sampled at Chumbicha (Catamarca Province, North Western Argentina; 28°9′S, 66°3′W) in 1995. In this locality *D. buzzatii* is polymorphic for the two common cosmopolitan second-chromosome arrangements *2st* and *2j*, as well as for the rare cosmopolitan $2jz^3$ with frequencies equal to 0.45, 0.45, and 0.10, respectively (Rodríguez et al. 2000). Arrangements *2j* and $2jz^3$ share inversion *2j* and are derived compared with *2st*, which is the species ancestral gene order (Ruiz, Fontdevila, and Wasserman 1982). A total of 29 isochromosomal lines for the second chromosome (15 *2st*, 10 *2j*, and four $2jz^3$), established using the $Antp/\Delta^5$ balancer stock (kindly provided by J. S. F. Barker), were studied in this work (see Rodríguez et al. 2000 for details). A stock of *D. koepferae* from Sierra de San Luis (Argentina) was used as outgroup. Both species belong to the *buzzatii* cluster of the *buzzatii* complex of the *mulleri* subgroup (see Rodríguez-Trelles, Alarcón, and Fontdevila 2000).

Sequence-Tagged Site (STS) Landmarks

Three anonymous regions previously mapped to the polytene chromosomes by in situ hybridization were chosen to study the effect of the polymorphic inversions on the level and distribution of nucleotide diversity. Going from distal to proximal ends on a *2st* gene arrangement of the second chromosome, the mapping of the STSs was as follows (fig. 1): outside-distal *70.09.1sts* on *2*(C7e) and

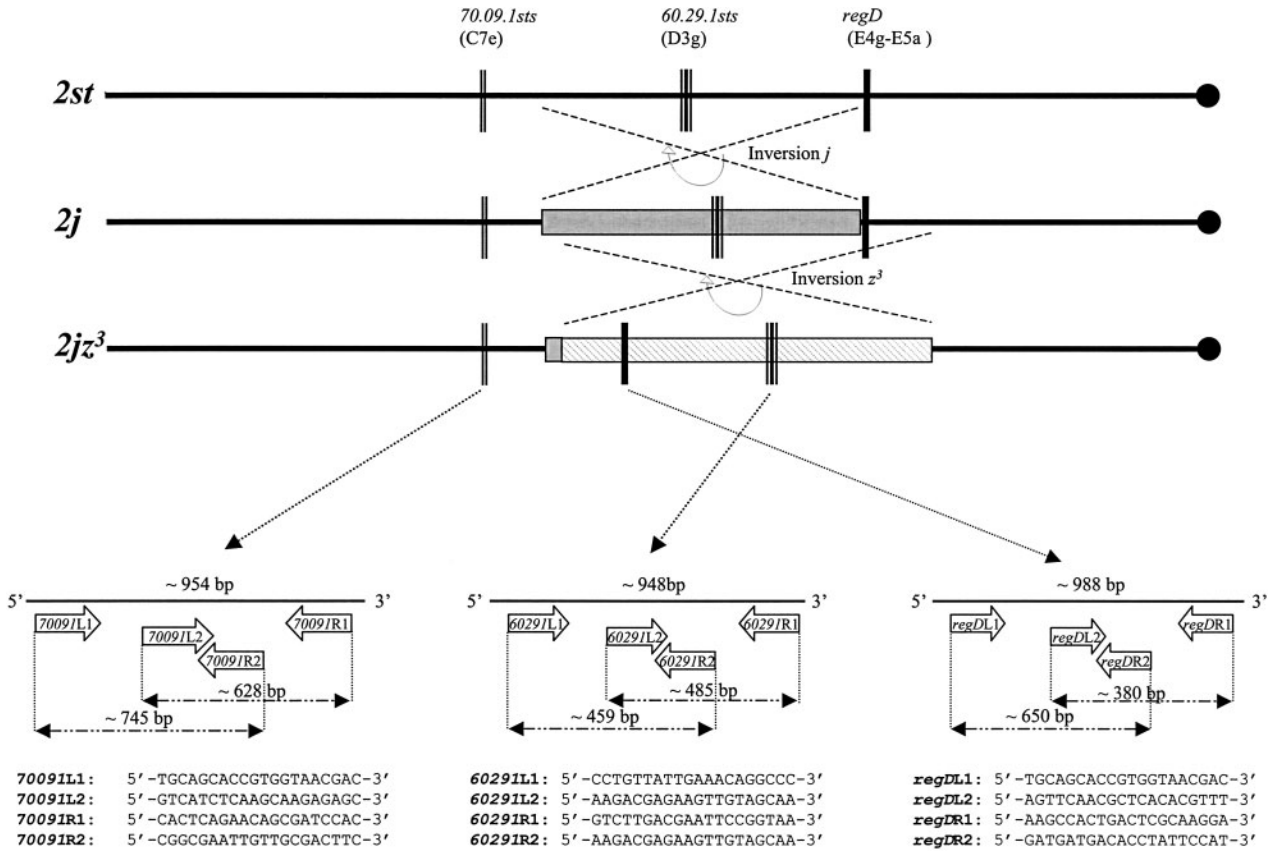


FIG. 1.—Schematic representation for the location of *70.09.1sts*, *60.29.1sts*, and *regD* markers mapped on second-chromosome arrangements of *D. buzzatii*. Connecting lines match the approximate cytological position of the breakpoints of inversion *2j* on a *2st* arrangements and the breakpoints of inversion $2z^3$ on a *2j* arrangement. The distal breakpoint of inversion $2z^3$ is approximately four cytological bands more proximal than that for inversion *2j* (see Laayouni, Santos, and Fontdevila 2000 for details). The sequences of all 20-nucleotide-long primers used for PCR amplification of the three regions investigated are shown, as well as the total lengths of the amplified fragments.

approximately 600 kb away (i.e., about 12 polytene chromosome bands and assuming approximately 50 kb per band as in *D. hydei* [see Laird 1973; Hartl et al. 1994]) from the distal breakpoint of inversion *2j*; inside-middle *60.29.1sts*, a randomly amplified polymorphic DNA fragment (RAPD) now converted to STS (see below) on *2*(D3g) and approximately in the center of the chromosomal segment covered by inversion *2j* (Laayouni, Santos, and Fontdevila 2000); and *regD* on *2*(E4g-E5a) and very close (27 bp away) to the proximal breakpoint of inversion *2j* (Cáceres et al. 1999). The first two STSs rendered significant ''hits'' with *D. melanogaster* nucleotide sequences located on the homologous Mueller/Sturtevant/Novitski chromosomal element *E* (= *2* in *D. buzzatii* and arm *3R* in *D. melanogaster* [see Powell 1997, p. 307]) when using the Blast program (Altschul et al. 1997) in the GenBank database (http://www.ncbi.nlm.nih.gov/BLAST/). Thus, *70.09.1sts* gives a match (a Blastn score of $P < 10^{-16}$) with sequence AE003725, and *60.29.1sts* (a Blastn score of $P < 10^{-26}$) with sequence AE003727.

The *70.09.1sts* locus shows a high similarity to a putative gene product described in *D. melanogaster* (GC5237, AE003725, Blastx score of $P < 4 \times 10^{-87}$), and it seems to have at least one short exon of 75 bp (at positions 649 to 723). *regD* also shows consistent similarity to another gene product described in *D. melanogaster* (CG13617, AE003748 Blastx score of $P < 4 \times 10^{-40}$) and seems to contain a short exon of 132 bp (at positions 286 to 417). However, given the small length of hypothetical coding regions in both sequences and uncertainties in the location of putative splicing sites, we did not analyze the number of synonymous and nonsynonymous substitutions.

DNA Isolation and Sequencing

DNA was extracted from three to five individuals of each isochromosomal line following Latorre, Moya, and Ayala (1986) after some modifications (see Laayouni, Santos, and Fontdevila 2000). The *70.09.1sts* sequence previously reported (Laayouni, Santos, and Fontdevila 2000) (EMBL/GenBank accession number AF288346) was employed to design 20-nucleotide-long primers to be used in polymerase chain reactions (PCR) (see fig. 1 and below). Gel-purified (10 to 100 ng) RAPD *60.29.1* was directly cloned in pMOS Vector (Amersham Pharmacia, England), and its DNA sequence was determined on both strands by the dideoxynucleotide chain termination method (Sanger, Nicklen, and Coulson 1977) using an ALF sequencer (Pharmacia Biotech, Piscataway, N.J.). Amplification primers were designed after RAPD *60.29.1* was converted to STS *60.29.1sts* (fig. 1). The proximal breakpoint *regD* sequence derived from a *2st* gene

arrangement (Cáceres et al. 1999) (EMBL/GenBank accession number AF162796) was used to design the primers *regD*L1 and *regD*R2 (see fig. 1) that amplified a fragment of approximately 650 bp. This fragment was then used as a probe for screening a *D. buzzatii* phage library (kindly provided by P. García). Plating of libraries, hybridization, and detection of positive clones were carried out following standard methods (Ausubel et al. 1998). Three positive clones were obtained, and their sequence was employed to design new primers for the amplification of a longer region towards the direction of the centromere. Figure 1 gives the sequences of the primers used for PCR amplifications and the total length of the regions investigated, as well as the strategies followed to determine overlaps between partial fragments.

PCR reactions were carried out in a final volume of 25μl, including 1× activity buffer (GIBCO BRL, Gaithersburg, Md.), 2 mM $MgCl_2$ (3 mM in the case of *60.29.1sts*), 200 μM of each dNTP (Boehringer Mannheim, Indianapolis, Ind.), 0.5 μM primer, template DNA (~30 to 40 ng), and 0.8 units of Taq polymerase (GIBCO BRL). Amplifications were run in a MJ Research Inc. (Watertown, Mass.) thermocycler programmed as follows: a preliminary 5-min denaturation at 94°C; 35 cycles of 30 sec at 94°C (denaturation), 1.5 min at specific PCR annealing temperatures (primers *70091*L1-R1, *70091*L1-R2, and *70091*L2-R1 at 60°C; *60091*L1-R1at 53°C; *60291*L1-R2 and *60291*L2-R1 at 58°C; *regD*L1-R1 at 55°C; *regD*L1-R2 at 53°C; and *regD*L2-R1 at 58°C), and 1.5 min at 72°C (extension); and a final extension at 72°C for 5 min followed by storage at 4°C. Because of failures in the amplifications, two additional primers were used in some cases: 5′-AGTTCAACGCTCACACGTTT-3′ instead of *60291*L2, and 5′-GCCGTGGCATGTGTGTGTGT-3′ instead of *60291*R1. Electrophoresis was performed in 1.4% agarose gels (SeaKem), and in all cases, the fragments amplified with the most external primers were gel purified and reamplified with the appropriate internal primers (see fig. 1). These amplified products were also gel purified and directly sequenced on both strands.

D. koepferae sequences were obtained using the same primers and procedures as those previously described for *D. buzzatii* with minor modifications in PCR conditions. In the case of sequence *60.29.1sts*, only one fragment (*60291*L1-R2 [see fig. 1 and below]) could be obtained.

Sequences collected in this study have been deposited into GenBank under accession numbers AY134753 to AY134842.

DNA Sequence Variation

A total of 2.6 kb were sequenced in each of the 29 *D. buzzatii* isochromosomal lines from Chumbicha. The multiple alignment of the three STSs analyzed (excluding insertion/deletion polymorphisms) included 943 sites for *70.09.1sts*, 411 sites for *60.29.1sts*, and 973 sites for *regD*. The difference between the approximately 948 bp for the total length of inside-middle *60.29.1sts* and the shorter fragment used in the analyses (amplified with primers *60291*L1-R2 [see fig. 1]) was due to technical difficulties when sequencing eight lines amplified for the overlapping second fragment (i.e., there is a gap of approximately 200 bp with many ambiguities that could not be resolved after several efforts). However, inspection of the complete fragment in the remaining 21 lines shows that molecular population estimates remain qualitatively the same (data not shown) and, hence, the shorter fragment can be considered as fully informative. Insertions/deletions (indels) were relatively frequent in all regions (*D. koepferae* generally had longer sequences), particularly in *60.29.1sts* with eight indels between *D. buzzatii* and *D. koepferae* (four gaps in *D. buzzatii* ranging from 4 bp to 38 bp at sites 68 to 72, 96 to 97, 132 to 133, and 267 to 268 and four gaps in *D. koepferae* ranging from 2 bp to 16 bp at sites 178 to 179, 375 to 376, 380 to 382, and 97 to 112). Line chu31 (*2j*) presented a deletion of 3 bp, and chu74 (*2st*) presented two deletions of 4 bp and 28 bp, respectively. In *70.09.1sts* 16 single indels were detected within *D. buzzatii* lines plus a large gap of 23 bp in *D. koepferae* (at sites 865 to 887). Finally, five single-base indels were observed in *regD* among *D. buzzatii* lines (one was only present in *2st*), in addition to two gaps of 4 bp (at sites 476 to 478) and 28 bp (at sites 446 to 447) in *D. buzzatii* and a gap of 3 bp (at sites 727 to 730) in *D. koepferae*. Length polymorphisms within and between chromosomal arrangements were also relatively common in *regD*, with an indel of 4 bp (at sites 494 to 495) in 13 out of 15 *2st* chromosomes and in all *2j* chromosomes (in which the indel was 2 bp longer) and an indel of 4 bp (at sites 691 to 692) present only in *2j* chromosomes. Isochromosomal line chu40 (*2st*) had a deletion of 9 bp (at sites 648 to 653). Indels were excluded from further analyses.

Data Analyses

Sequences were multiply aligned using the default option of the program ClustalW (version 1.6) (Thompson, Higgins, and Gibson 1994) and thereafter aligned manually to minimize the number of differences. Sequences from line chu3 of *D. buzzatii* were used as reference in all cases. Nucleotide polymorphisms were estimated by commonly used measures of DNA variability as number of segregating nucleotide sites (S), nucleotide diversity (π) or average number of nucleotide differences per site (Nei 1987), and heterozygosity per site (θ) expected under the infinite-site model at mutation-drift equilibrium given the observed S value (Watterson 1975).

Different neutrality tests were performed to determine whether the observed data conformed to the predictions of the neutral model of molecular evolution. Tajima's (1989) D statistic relies on intraspecific data and tests the null hypothesis that two estimates of the neutral mutation parameter, one derived from the average number of pairwise nucleotide differences and the other based on the number of segregating sites in the sample, are equal. Fu and Li (1993) proposed a set of tests that also rely on the comparison of different estimates of the mutation parameter. Their D statistic is based on the standardized difference between the total number of mutations and the number of mutations in the external branches of the genealogy, and their F statistic compares the standardized

difference between the average number of pairwise differences and the number of mutations in external branches of the genealogy. An outgroup species is needed to estimate the number of mutations in external branches, and the sequences of *D. koepferae* were used for this purpose. Negative values of the tests indicate an excess of rare variants and are consistent with a selective sweep or with a recent population growth. On the other hand, positive values indicate an excess of intermediate frequency variants and are usually associated with stable population subdivision (either spatial or genetical). Fu's (1997) F_S statistic tests the probability of having no fewer than the number of observed alleles in the sample given that $\theta = \pi$. This statistic tends to be negative when there is an excess of recent mutations (or rare alleles) and may be used to better detect population growth, but it is not conservative against recombination. For increasing levels of recombination (see below) the behavior of statistical tests that use information of the mutation (segregating site) frequency is better than that for the F_S statistic. Therefore, we have also used Ramos-Onsins and Rozas' (2002) R_2 statistic, which is based on the difference between the number of singleton mutations and the average number of nucleotide differences. Lower values of R_2 are expected under a scenario of recent population growth.

Intragenic recombination substantially affects the power of the statistical tests employed, and it is more realistic to assume $C > 0$ because there is evidence of recombination in our data set. However, the use of estimators of the population recombination rate, $C = 4Nc$, from nucleotide polymorphism data (Hudson 1987) are problematic because they tend to be unreliable unless large stretches of sequence information are available from many individuals (Wall 2000). The analyses were performed using a lower bound C_m estimate of the recombination parameter (i.e., minimum number of detected recombination events [see Hudson and Kaplan 1985]) and an independent estimate (C_{map}) based on the comparison of the physical and genetic maps, which represents our best guess of the true crossing-over rate. Thus, assuming that the size of the euchromatic portion for the second chromosome is approximately 28 Mb as in the homologous element E (chromosomal arm 3R) of *D. melanogaster* (Adams et al. 2000) and that the map length is approximately 138.5 cM (Schafer et al. 1993), an estimate for the recombination rate $c = 2.5 \times 10^{-8}$ per bp per generation (sex averaged) is obtained, assuming that recombination is approximately constant along this acrocentric chromosome in *D. buzzatii* (c.f. with the average value of 1.1×10^{-8} per bp per generation estimated in *D. melanogaster* females [see Lindsley and Zimm 1992]). Watterson's (1975) estimate of the neutral mutation parameter $4Nu$ is 0.0297 on average (table 1), and the neutral mutation rate in Drosophila is assumed to be 3×10^{-9} per site per generation (Andolfatto and Przeworski 2000). Therefore, the estimate for the effective population size in *D. buzzatii* is $N \approx 2.5 \times 10^6$, and $C_{map} \approx 4 \times (2.5 \times 10^6) \times (2.5 \times 10^{-8}) = 0.25$ per bp per generation for *70.09.1sts*, assuming that recombination is not inhibited in heterokaryotypes at the nonheterozygous distal portion of the second chromosome. On the other hand, we can expect a reduced level of

Table 1
Nucleotide Polymorphism and Divergence (*K*) Between *D. buzzatii* and *D. koepferae* (Outgroup) in the Three Analyzed Regions

	70.09.1sts (943 bp)				*60.29.1sts* (411 bp)				*regD* (973 bp)			
	Total	*2st*	*2j*	$2jz^3$	Total	*2st*	*2j*	$2jz^3$	Total	*2st*	*2j*	$2jz^3$
Sample size	29	15	10	4	29	15	10	4	29	15	10	4
S (η)	110 (121)	78 (85)	61 (64)	35 (35)	66 (70)	52 (55)	42 (46)	29 (30)	71 (72)	47 (48)	11 (11)	3 (3)
Singletons	57	54	32	34	34	32	14	23	31	26	4	3
θ	0.0297	0.0253	0.0227	0.0201	0.0409	0.0386	0.0323	0.0338	0.0186	0.0148	0.0040	0.0017
π	0.0197	0.0181	0.0210	0.0186	0.0319	0.0323	0.0349	0.0338	0.0148	0.0104	0.0036	0.0015
K	0.0559	—	—	—	0.1263	—	—	—	0.0607	—	—	—
C_m	21	12	14	0	13	10	9	0	6	6	0	0
C_{map}	236	236	236	236	43	21	21	1	101	49	49	2
C_{hud}/θ	1.18				27.83				0.72			
Tajima's D	−1.533**	−1.505**	−0.671	−0.773	−1.008	−0.902	−0.071	−0.335	−0.809	−1.329	−0.358	−0.754
Fu and Li's D	−2.471***	−2.578***	−0.842	−1.428	−1.077	−1.073	0.615	1.039	−1.769	−1.693	−0.215	−1.201
Fu and Li's F	−2.598***	−2.729**	−0.961	−1.539	−1.284	−1.263	0.500	0.907	−1.769	−1.919	−0.388	−1.279
Fu's F_s	−15.275***	−4.883*	−1.856	0.978	−16.099**	−5.957*	−2.278	3.414	−10.101	−7.335*	−2.974	−0.288
Ramos-Onsins and Rozas's R_2	—	0.0788***	0.1154	0.2947	—	0.0909*	0.1560	0.2919	—	0.0782**	0.1457	0.2764

NOTE.—S indicates the number of segregating sites; η indicates the total number of mutations; θ indicates the heterozygosity per site given S; π indicates nucleotide diversity; C_m and C_{map} are estimates of the population recombination parameter given per sequence; C_{hud}/θ is Hudson's (1987) recombination parameter relative to the neutral mutation parameter. The corresponding statistics was considered to be significant when $P < 0.05$ for both recombination values (C_m and C_{map}). Asterisks indicate those statistically significant values that reject the null hypothesis of selective neutrality: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

recombination in heterokaryotypes at loci linked to the inversions (which seems to be the case for *regD* [see below]) and, therefore, the previous C_{map} estimate has to be correspondingly multiplied by the expected frequency of homozygous individuals. Thus,

$$C_{\text{map}} = 0.25 \times \left(1 - \sum_{i \neq j} 2p_i p_j\right)$$

when considering the total sample, and

$$C_{\text{map}} = 0.25 \times p_i^2$$

when considering a given gene arrangement. The added contribution of gene conversion is obviously ignored in our previous estimates for the total rate of exchange.

Genetic differentiation between chromosomal arrangements was estimated by means of the average number of nucleotide substitutions per site (d_{XY}) and the net number of nucleotide differences per site (d_{A}) (Nei 1987). The extent of genetic differentiation was also analyzed by means of F_{ST} (Hudson, Slatkin, and Maddison 1992) and K_{ST} (Hudson, Boos, and Kaplan 1992), and the statistical significance was obtained by means of permutations tests implemented in the program ProSeq 2.9 (Filatov 2001). Linkage disequilibrium was analyzed between pairs of informative sites (i.e., those sites where the less frequent variant is present at least twice in the sample) and also by the sign test on *D* (Lewontin 1995). Statistical significance was tested by means of Fisher's exact test after correcting for multiple comparisons with the Bonferroni procedure (Weir 1996).

The DnaSP version 3.98 software (Rozas and Rozas 1999) was used to obtain most estimates of nucleotide variability reported in this work, to perform neutrality tests, and to detect gene conversion tracts from the algorithm proposed by Betrán et al. (1997). Critical values for statistical tests were obtained by computer simulations (1,000 replicates) using the coalescent algorithm described in Hudson (1990) and implemented in DnaSP. In all cases, the most conservative estimate of the population recombination parameter (i.e., C_{m} versus C_{map}) was used. Phylogenetic analyses employed to reconstruct the genealogies of the studied isochromosomal lines were conducted using MEGA version 2.1 (Kumar et al. 2001).

Results

Sequence Polymorphism Within Second-Chromosome Arrangements

Estimates of nucleotide variation for each STS are given in table 1 for the whole sample and for the different chromosome arrangements. A total of 247 polymorphic sites (263 mutations) were detected: 110 in *70.09.1sts*, 66 in *60.29.1sts*, and 71 in *regD*. Both the outside-distal and the inside-middle STSs display analogous levels of nucleotide variation among the three chromosomal arrangements, which were also equivalent to the corresponding estimates in the total data set. Thus, the average of the ratios of π in *2j* and *2jz*^{3} to π in *2st* is 1.08, very close to the expected value of 1 if the variation in the derived chromosome arrangements is a subset of that found in *2st* chromosomes. In contrast, nucleotide diversity in *regD* is, respectively, three and seven times lower in *2j* and *2jz*^{3} when compared with *2st* (see also Cáceres, Puig, and Ruiz 2001). Estimates of θ and π were similar within *2j* and *2jz*^{3} arrangements, but for *2st*, the relatively high number of singletons results in a larger θ when compared with π.

Gene Flux

The recombination parameter 4*Nc* (Hudson 1987) can be standardized relative to the neutral mutation parameter *4Nu* to estimate the expected number of recombination events per mutation under neutral equilibrium assumptions. An independent measure of this quantity (*c*/*u*) is based on the laboratory estimates of the recombination rate in the second chromosome of *D. buzzatii* ($c = 2.5 \times 10^{-8}$ per bp per generation [see above]) and the neutral mutation rate ($u \approx 3 \times 10^{-9}$ per site per generation), which yields a *c*/*u* ratio of approximately 8.3 for the outside-distal marker (assuming that recombination is not inhibited for this marker in heterokaryotypes [see above]). From the actual arrangement frequencies at Chumbicha and assuming no recombination in heterokaryotypes, the ratio becomes approximately 4.9 for the markers linked to the inversions. With the exception of the inside-middle marker *60.29.1sts*, which renders the highest C_{hud}/θ ratio (5.7 times higher than *c*/*u*, but this figure should be taken with caution because the sequence is only 411 bp long), C_{hud}/θ ratios are about seven times lower than expected (table 1). Because the recombination parameter is a summary of the amount of linkage disequilibrium in a sample (Hudson 1987), lower than expected C_{hud}/θ ratios are likely due to the linkage disequilibrium introduced by the inversions (see below).

The algorithm proposed by Betrán et al. (1997) was used to detect gene conversion tracts among arrangements. The probabilities of a site being informative of a conversion event (ψ) were relatively high for *regD* between the ancestral and the two derived chromosomal arrangements (table 2). However, no gene conversion tracts were detected for this region. Four conversion tracts were identified for *70.09.1sts*, but the one between *2st* and *2j* (line chu3) could correspond to a double crossover event because the length of the tract is unusually large (766 bp). The other three conversion tracts were identified between *2st* and *2jz*^{3} with an average length of 146 bp. Finally, for *60.29.1sts,* three conversion tracts were identified between *2st*-*2jz*^{3} and four were identified between *2j*-*2jz*^{3}, with an average length of 32 bp (figs. 2*a* and *b*).

The relative position of *regD* in *2j* and *2jz*^{3} arrangements (fig. 1) suggests that gene flux should be higher between them than between *2st* and *2j* or *2st* and *2jz*^{3}. In the first case (*2st*-*2j*), gene flux is virtually suppressed near the proximal breakpoint of the inversion, whereas in the second case (*2st*-*2jz*^{3}), the cause for an absence of gene flux might be the involvement of *regD* in the double loop formed in heterokaryotypes differing by the two overlapping inversions (recall that no conversion tracts have been uncovered for *regD* despite a relatively high probability of detection). On the other hand, gene

arrangements do not form separate clusters and are mixed in *70.09.1sts* (fig. 2*a*) and *60.29.1sts* (fig. 2*b*) trees, which confirms that gene flux inside (center) the inverted region is relatively high or even higher to that occurring rather far apart from the distal breakpoint.

Linkage Disequilibrium and Genetic Differentiation

For each STS, linkage disequilibrium was analyzed for all pairs of informative polymorphic sites in the total data set as well as within each chromosomal class. In the proximal-breakpoint *regD*, 195 out of 780 (25%) comparisons showed a significant association by Fisher's exact test, and this number dropped to 91 (12%) after correcting for multiple contrasts. However, these percentages are not very informative, because in some cases no significant association can be detected even with extreme disequilibrium (Lewontin 1995). Therefore, linkage disequilibrium was also analyzed with the sign test on *D* (Lewontin 1995), first for the actual ordering of 40 informative sites in the whole sample (39 independent pair comparisons) and thereafter by recording the proportion of goodness-of-fit tests that rendered a statistically significant *G*-value after 200 random ordering permutations (including the actual ordering). Because linkage disequilibrium decays with physical distance due to intragenic recombination, we could expect that most random orderings would render a lower *G* statistic than the actual ordering (which was indeed the case [data not shown]). The observed number of negative *D*s was always larger than expected (26 versus 19.03 for the actual ordering; $G = 5.20$, $P < 0.05$), and 68% of *G*-values were statistically significant. Thus, a significant excess of repulsion linkage in *regD* was detected, which may be partitioned into within-inversion and between-inversion components (Nei and Li 1973; Ruiz et al. 1991; Navarro et al. 1996). Within *2st*, 20 independent pair comparisons are possible, and the observed number of negative *D*s was slightly larger than expected in most random orderings (14 versus 11.25 for the actual ordering; $G = 1.64$, $P > 0.05$), with only 6% of the *G*-values being statistically significant. A similar result was found within the *2j* arrangement with six independent pair comparisons, and, therefore, the conclusion is that the disequilibrium detected for *regD* is mostly due to the genetic differentiation between arrangements *2st* and *2j* (see below). Similarly, linkage disequilibria were also analyzed for the outside-distal *70.09.1sts*, with 44 independent pair comparisons in the total data set ($G = 0.58$ for the actual ordering; $P > 0.05$), and the inside-middle *60.29.1sts*, with 27 independent pair comparisons in the total data set ($G = 4.22$ for the actual ordering; $P < 0.05$). There was no evidence of linkage disequilibrium in the whole sample or within-inversion components for *70.09.1sts*, but there was some excess of repulsion linkage for *60.29.1sts* in the entire sample (not within chromosomal arrangements), with 14% of the *G*-values being significant after 200 random ordering permutations. In conclusion, statistically significant overall genetic associations between-inversions were mainly detected for *regD*.

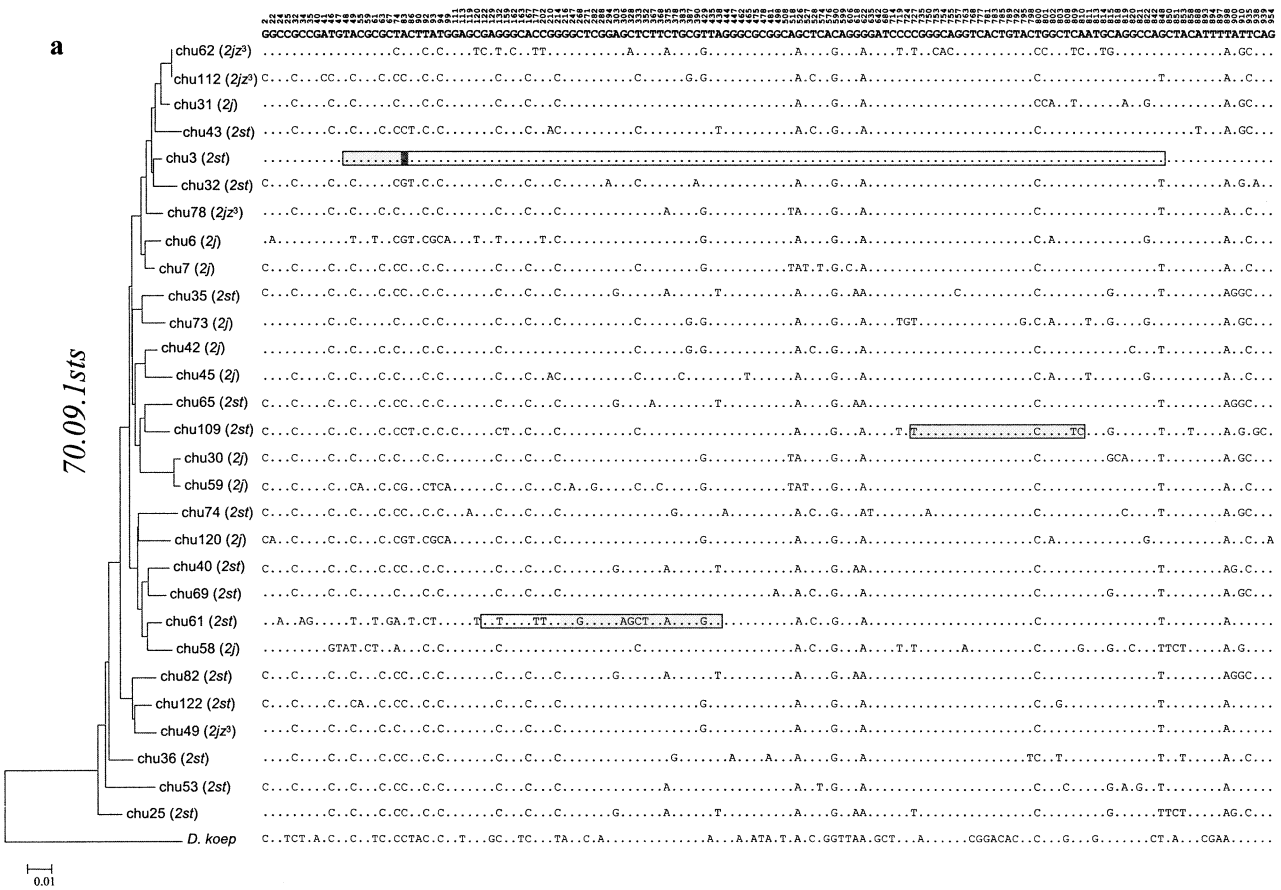
The average number of nucleotide differences between sequences differing in chromosomal arrangement

Table 2
Genetic Differentiation Between Gene Arrangements for the Three Analyzed Regions

	70.09.1sts			*60.29.1sts*			*regD*[a]		
	2st-2j	*2st-2j²³*	*2j-2j²³*	*2st-2j*	*2st-2j²³*	*2j-2j²³*	*2st-2j*	*2st-2j²³*	*2j-2j²³*
η	115	95	75	67	58	54	71	65	12
Shared (nonshared) mutations	33 (51, 31)	23 (60, 12)	23 (40, 12)	24 (31, 12)	18 (37, 3)	22 (24, 8)	0 (47, 11)	1 (47, 2)	1 (10, 1)
F_{ST}	0.0894**	0.0257	−0.0123	0.0609*	0.0928	0.0734	0.6861***	0.7459***	0.2370*
K_{ST}	0.0211**	0.0119	0.0021	0.0136*	0.0339*	0.0412*	0.2363***	0.2215***	0.0987**
d_{XY}	0.0214	0.0187	0.0193	0.0326	0.0321	0.0367	0.0222	0.0225	0.0030
d_A	0.0020	0.0004	−0.0002	0.0016	0.0025	0.0019	0.0152	0.0165	0.0007
ψ	0.0041	0.0036	0.0003	0.0041	0.0039	0.0038	0.0154	0.0080	0.0000

NOTE.—η is the total number of mutations; F_{ST} and K_{ST} are estimates of genetic differentiation; d_{XY} is the average number of nucleotide substitutions per site; d_A is the net number of nucleotide substitutions per site; ψ is the probability of detecting a converted tract. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

[a] The difference between η and the total number of shared and nonshared mutations is the number of fixed differences between arrangements.


a
70.09.1sts
chu62 (2jz³)
chu112 (2jz³)
chu31 (2j)
chu43 (2st)
chu3 (2st)
chu32 (2st)
chu78 (2jz³)
chu6 (2j)
chu7 (2j)
chu35 (2st)
chu73 (2j)
chu42 (2j)
chu45 (2j)
chu65 (2st)
chu109 (2st)
chu30 (2j)
chu59 (2j)
chu74 (2st)
chu120 (2j)
chu40 (2st)
chu69 (2st)
chu61 (2st)
chu58 (2j)
chu82 (2st)
chu122 (2st)
chu49 (2jz³)
chu36 (2st)
chu53 (2st)
chu25 (2st)
D. koep
0.01
b
60.29.1sts
chu61 (2st)
chu62 (2jz³)
chu6 (2j)
chu3 (2st)
chu58 (2j)
chu42 (2j)
chu69 (2st)
chu32 (2st)
chu109 (2st)
chu43 (2st)
chu120 (2j)
chu31 (2j)
chu45 (2j)
chu73 (2j)
chu49 (2jz³)
chu78 (2jz³)
chu30 (2j)
chu122 (2st)
chu112 (2jz³)
chu7(2j)
chu59 (2j)
chu36 (2st)
chu74 (2st)
chu25 (2st)
chu65 (2st)
chu40 (2st)
chu35 (2st)
chu82 (2st)
chu53 (2st)
D. koep
0.01

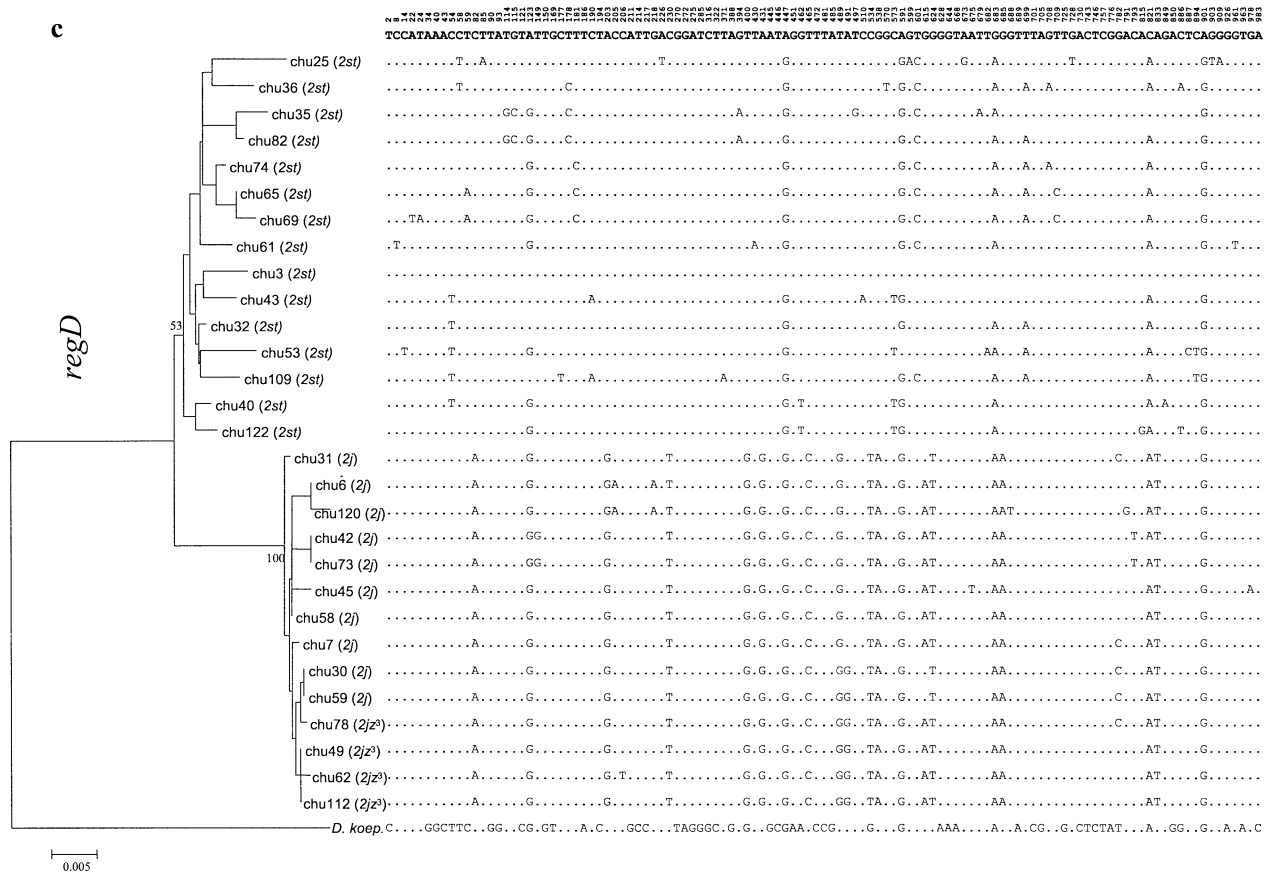


FIG. 2 (Continued)

was 68.67 for *70.09.1sts* (72.1% of the total number of mutations), 38.33 for *60.29.1sts* (64.1%), and 48.67 for *regD* (96.7%), which give an estimate of the average number of nucleotide substitutions per site between arrangements (d_{XY}) equal to 0.0198, 0.0338, and 0.0159, respectively (table 2). For the outside-distal and the inside-middle STSs no fixed differences between arrangements were observed, whereas the situation for the proximal-breakpoint *regD* was dramatically different, with 13 and 15 fixed differences between *2st-2j* and *2st-2jz*^{3}, respectively, and 46 segregating sites in the ancestral *2st* that were monomorphic in the derived arrangements. Significant differences among all arrangements were detected for *60.29.1sts* and *regD* using K_{ST} as the test statistic but only between *2st* and *2j* for *70.09.1sts*. Except for *regD*, however, the observed number of shared polymorphisms among arrangements is substantial (table 2). We have applied the hypergeometric distribution to estimate the probability that the number of shared polymorphisms between *2st-2j* could be explained by independently arising parallel mutations (Rozas and Aguadé 1994). For *70.09.1sts*, the expected number of shared polymorphisms would be less than or equal to five (cf. 33 shared mutations observed), and for *60.29.1sts* it would be less than or equal to four (24 observed). There is, therefore, extensive genetic exchange between arrangements as previously pointed out.

Figures 2*a–c* show neighbor-joining trees (Saitou and Nei 1987) for each STS analyzed. Genealogies of all lines were reconstructed using *D. koepferae* as the outgroup. In the *regD* tree (fig. 2*c*), sequences corresponding to the *2st* arrangement form a unique group but with a weak bootstrap value (53%), whereas those corresponding to the derived *2j* and *2jz*^{3} arrangements form a monophyletic group with some subclustering for *2jz*^{3}. The high bootstrap value of the monophyletic cluster including all *2j* and *2jz*^{3} sequences corroborates the proposed unique origin of inversion *2j* (Cáceres et al. 1999; see also Cáceres, Puig, and Ruiz 2001) and strongly argues for a monophyletic origin of inversion *2z*^{3} on an ancestral *2j* chromosome. This is further sustained by the fact that all fixed differences between *2st* and *2j* overlap with a fixed

←

FIG. 2.—Nucleotide polymorphisms and genealogies obtained by the neighbor-joining method for (*a*) *70.09.1sts*, (*b*) *60.29.1sts*, and (*c*) *regD* markers of *D. buzzatii* and *D. koepferae* (outgroup). Nucleotides identical to the reference sequence (chu3) are shown as a dot. Conversion tracts are indicated by white boxes between *2st* and *2j*, pale gray boxes between *2st* and *2jz*^{3}, and dark gray boxes between *2j* and *2jz*^{3}. Hatched position at *70.09.1sts* in line chu3 is shared by the two conversions tracts detected.

difference between *2st* and *2jz*[3] (table 2), and, likewise, the distribution of indels between arrangements fully agrees with the distribution of fixed and shared polymorphisms.

An important concern regarding the *regD* tree in figure 2*c* is the mutational instability of the region close to the inversion breakpoint. Although such instability seems to be restricted to the transposable element sequences implicated in the origin of inversion *2j* (Cáceres, Puig, and Ruiz 2001), we have tested the hypothesis of an accelerated substitution rate in this lineage. Relative-rate tests between lineages (using a *regD* phylogeny derived from maximum parsimony implemented in PAUP* 4.0b10 for Macintosh [Swofford 2000]) using RRTree version 1.1 program (Robinson et al. 1998; Robinson-Rechavi and Huchon 2000) suggest that standard (*st*) and inverted (*j*) lineages evolve at the same rate (Kimura's [1980] two-parameter model: mean $Kst = 0.0601$, mean $Kj = 0.0666$, $P = 0.135$). Therefore, mutational instability at the proximal breakpoint of inversion *2j* seems to happen only at the inserted transposable element sequences since no increased rate of nucleotide change at *regD* is detected. Consistent with the higher level of polymorphism in *2st* chromosomes, the tree is deeper in the branches connecting these sequences.

Neutrality Tests

The observed distributions of mutations between and within arrangements were contrasted with those predicted under the null hypothesis of selective neutrality. Tajima's (1989) *D* statistic was consistently negative and significantly so for the outside-distal marker *70.09.1sts* in the total sample as well as within *2st* gene arrangements, indicating a higher than expected number of low-frequency variants. An excess of unique polymorphisms was also confirmed for this marker with Fu and Li's (1993) tests. Interestingly, Fu's (1997) F_S statistic was negative and statistically significant for all STSs when considering the *2st* ancestral gene arrangement, suggesting that the excess of rare alleles may be the result of a population expansion or hitchhiking.

The possibility of an expansion event was further investigated by means of the R_2 statistic described in Ramos-Onsins and Rozas (2002). They advocate its use for small sample sizes, particularly in the presence of recombination. The last row in table 1 gives the R_2 estimates, which were always statistically significant in the subsample of *2st* chromosomes. The null hypothesis of constant population size, however, cannot be rejected for the subsamples including the derived arrangements. Therefore, the most plausible conclusion seems to be that *D. buzzatii* has passed through a population expansion and that this event has likely preceded the last event of directional selection that swept nucleotide variation in the population of *2j* chromosomes (see below).

A caveat to the above conclusion is that sample size affects the power to reject the constant size model (see figure 2 in Ramos-Onsins and Rozas 2002), and the number of lines used in this work for arrangement *2st* was 15, whereas the number of lines for arrangement *2j* was only 10. We have, therefore, generated independent subsamples by obtaining random combinations of 15 *2st* arrangements taken 10 at a time and calculated the R_2 statistic for each subsample. The results indicate that R_2 estimates remain statistically significant for *70.09.1sts* and *regD* in most cases, although for the shorter fragment *60.29.1.sts* most estimates felt in the 95% confidence limits but close to the lower limit and in some cases (~15%) remained significant. Therefore, different sample sizes between *2st* and *2j* do not seem to account for the discrepancies we observed.

Evolutionary History of Inversion *2j*

The age of inversion *2j* can be estimated from the net number of nucleotide differences per site between *2st* and *2j* gene arrangements in *regD*. Given that the average number of nucleotide differences (d_{XY}) between those arrangements is 0.0222 (table 2) and between *D. buzzatii* and *D. koepferae* the average number is 0.0601, the net number of nucleotide substitutions between *2st-2j* is $d_A = 0.0152$ and between *D. buzzatii–D. koepferae* is $d_A = 0.0525$.

Using the *Xdh* sequences from *D. buzzatii* (GenBank accession numbers AF226958 to AF226959), *D. koepferae* (GenBank accession numbers AF226964 to AF226965), and *D. hydei* (GenBank accession numbers AF226974 to AF226975) reported in Rodríguez-Trelles, Alarcón, and Fontdevila (2000), rates of synonymous substitutions are $Ks(D.b.\text{-}D.h.) = 0.4792$ and $Ks(D.b.\text{-}D.k.) = 0.1351$. Divergence time between *D. buzzatii* and *D. hydei* is approximately 14.26 Myr (Russo, Takezaki, and Nei 1995), and, therefore, the divergence time between *D. buzzatii* and *D. koepferae* can be estimated as approximately 4.02 Myr. As a result, the age of inversion *2j* can be estimated as approximately 1.16 Myr (in agreement with the figure reported by Cáceres, Puig, and Ruiz 2001).

Assuming that *regD* has diverged at a neutral rate, we can now obtain a mutation rate for this region from the net number of nucleotide substitutions between *D. buzzatii* and *D. koepferae*; namely, $u \approx 6.53 \times 10^{-9}$ per site per year or about 1.31×10^{-9} per site per generation, assuming five generations per year in *D. buzzatii*. (Field data from a natural population in Spain point to a generation time in this species about two times longer than in *D. melanogaster* at the warmest months of the year [Quezada-Díaz et al. 1997; M. Santos unpublished observations].) This value is approximately twice the lower bound for the neutral mutation rate reported in Andolfatto and Przeworski (2000) and would render an $N \approx 3.5 \times 10^6$ from the number of segregating variant sites at *regD*, which roughly corresponds to approximately 1.63 *N* generations for the age of inversion *2j*. In other words, the second-chromosome polymorphism in *D. buzzatii* can be considered as to be a middle-aged polymorphism according to Andolfatto, Depaulis, and Navarro (2001). Finally, the amount of nucleotide diversity can be used to estimate the age of the sampled alleles for each chromosomal arrangement (Rozas et al. 1999), which yields approximately 0.79 Myr for *2st* and approximately 0.27 Myr for *2j* (or ~0.38 *N* generations). It is interesting to note that in previous studies, coalescent

times were grossly underestimated, probably due to the inclusion of a large number of alleles sampled in recently colonizing populations from Australia and Spain (Cáceres, Puig, and Ruiz 2001).

Our previous estimate of N from the net number of nucleotide substitutions between *D. buzzatii* and *D. koepferae* increases the C_{map} estimates given in table 1 (see *Materials and Methods*) and, hence, are not a conservative assumption for Fu's (1997) F_S test and may be not conservative for Ramos-Onsins and Rozas' (2002) R_2 test. However, after using a corrected estimate for the recombination parameter, the statistical significance given in table 1 for F_S and R_2 values still remains. Therefore, the previous claim that *D. buzzatii* has passed through a population expansion seems to be sound.

Discussion

We have surveyed nucleotide variation for three anonymous markers that map at different positions according to the second-chromosome arrangements of *D. buzzatii*. The most striking observations were (1) the extensive gene flux for the marker located at the central position of inversion *2j*; (2) the reduced levels of nucleotide variability and extensive linkage disequilibrium resulting from strong genetic differentiation between arrangements at the proximal breakpoint region *regD*, suggesting that gene flux between arrangements is significantly suppressed at breakpoint regions; and (3) the significant excess of low-frequency segregating polymorphisms, mostly in the ancestral *2st* arrangement, which is probably a consequence of a population expansion that predates the coalescent time of inversion *2j*. In the following, we discuss the possible evolutionary forces underlying these patterns and the maintenance of the inversion polymorphism in nature.

Linkage disequilibria were almost absent at markers far apart from the *2j* inversion breakpoints, but some genetic differentiation between *2st* and *2j* arrangements still remains, even at the outside-distal *70.09.1sts* (table 2). This marker is located approximately 600 kb from the distal breakpoint and was used here in the belief that rates of exchange between arrangements were high enough as to make the nucleotide diversity of the marker independent of the evolutionary fate of inversions. Our wrong a priori conviction rested on the following reasoning: map distances in *D. buzzatii* for the second chromosome are about twice those estimated in *D. melanogaster* (see above), and theoretical models predict that recombination increases in the uninverted distal segment (Navarro et al. 1997). However, in accordance with empirical data from *D. melanogaster* (Andolfatto, Depaulis, and Navarro 2001), it seems that inversion effects on levels of nucleotide diversity may extend as far as 600 kb from breakpoints in *D. buzzatii*.

Patterns of divergence between chromosomal arrangements suggest that inversion *2j* in *D. buzzatii* can be considered as a middle-aged polymorphism (~1.63 N generations). However, a model of strong balancing selection that does not fluctuate with time (Navarro, Barbadilla, and Ruiz 2000) could be ruled out because patterns of nucleotide diversity place the age of the sampled *2j* alleles at approximately 0.27 Myr (~0.38 N generations). As for the rare cosmopolitan $2jz^3$ and $2jq^7$ arrangements, whose frequencies are generally low in New World natural populations (Hasson et al. 1995), it could be the case that historical frequencies of inversion *2j* have also remained quite low during most (~75%) of the time, and its rise in frequency, likely due to selection, has only occurred recently. This scenario is coherent with the reduced levels of nucleotide variation and the strong genetic differentiation between standard and inverted chromosomes (with no shared polymorphisms) observed at *regD* and also with the lower level of nucleotide variation in derived relative to ancestral chromosomes often found in Drosophila (Wesley and Eanes 1994; Hasson and Eanes 1996; Andolfatto, Wall, and Kreitman 1999; Depaulis et al. 2000). As previously pointed out, inversion $2z^3$ arose from a *2j* chromosome and changed the location of *regD* relative to the inversion loops formed in heterokaryotypes. This relocation of *regD* would be expected to affect the level of genetic exchange between arrangements, but the most we can say is that gene flux seems to be absent despite the relatively high probability of detecting genetic exchange not only between the ancestral arrangement and *2j* but also between *2st* and the other derived arrangement $2jz^3$.

Under the infinite-site model with no recombination and no selection, the average number of pairwise differences within an allelic class (i.e., standard or inverted sequences at the breakpoints) is expected to be proportional to the frequency of the allelic class times the neutral mutation parameter, and the ratio of the sum of the average number of pairwise differences within allelic classes to that number in the total sample is expected to equal 1 (Innan and Tajima 1997, 1999). For *regD*, this ratio for *2st* and *2j* gene arrangements is 0.93, close to the neutral expectation but in the predicted direction (i.e., <1) if inversion *2j* were maintained by balancing selection. On the other hand, Andolfatto, Depaulis, and Navarro (2001) have shown that total nucleotide diversity is lower at markers closer to inversion breakpoints, an ''unexpected'' result under neutrality. As can be seen in table 1, nucleotide diversity is 2.2 times larger for the inside-middle marker *60.29.1sts* than for *regD*, in agreement with the previous findings. However, the net number of nucleotide substitutions between *D. buzzatii* and *D. koepferae* for *60.29.1sts* is $d_A = 0.1109$, which yields an estimate of the neutral mutation rate $u \approx 2.76 \times 10^{-9}$ per site per generation; that is, 2.1 times larger than that estimated for *regD* (see above) and consistent with the previous ratio between nucleotide diversities. Thus, the level of polymorphism within species does not seem to significantly differ from the divergence level between species, and heterogeneity in mutation rates seem to be sufficient to explain the different levels of nucleotide diversity observed. This conclusion is reinforced from the estimated synonymous nucleotide diversities and substitutions for locus *Esterase-A* ($\pi = 0.0521$, $K_{(D.\ buzzatii-\ D.\ koepferae)} = 0.153$) (Gómez and Hasson 2003), which maps on *2*(D4f-h) and about seven polytene chromosome bands far from *60.29.1sts*. Those values are, respectively, about 3.5 and 2.5 times

larger than the corresponding estimates for *regD*. To summarize, the present results do not provide conclusive evidence for a putative long-term selective advantage of heterokaryotypes.

In the dynamics of inversion polymorphisms, selection may be operating at different levels, and several hypotheses have been proposed to explain the establishment of a given inversion in the population. Under a ''genic selection'' model, an inversion would be advantageous, provided it contains few or no deleterious alleles (Nei, Kojima, and Schaffer 1967; Santos 1986). However, deleterious mutations will accumulate in the inverted chromosome fragment and their frequencies eventually reach mutation-selection equilibrium. At this time, the inversion would become selectively neutral, and, therefore, it is obvious that other factors should be taken into account in order to explain the equilibrium frequencies of gene arrangements in natural populations of Drosophila. Data obtained for inversion *In(3L)Payne* in *D. melanogaster* showed no significant departure from neutral equilibrium (Wesley and Eanes 1994; Hasson and Eanes 1996; see also Innan and Tajima 1999) despite patterns suggesting selective sweeps. Similar patterns were observed in other species from the *obscura* group (Rozas and Aguadé 1990, 1993; Babcock and Anderson 1996). As stated by Andolfatto, Depaulis, and Navarro (2001), if selective sweeps are frequent in natural populations and differentially affect chromosome arrangements, how inversions remain polymorphic is unclear unless we assume that independently occurring selective sweeps cause a balanced polymorphism pattern. This explanation was put forward by Kirby and Stephan (1996) for small sequences of DNA where the absence of recombination is due to the small genetic distance, and it is known as the ''traffic hypothesis.''

While the hypothesis that the second-chromosome inversion polymorphism in *D. buzzatii* has a ''long-lived'' history of heterosis is not compatible with our observations, it is unclear what patterns of nucleotide diversity to expect from other types of balancing selection that can lead to shifts in inversion frequencies. Inversion polymorphisms have been classified in earlier works as ''rigid'' and ''flexible'' (Carson 1965; Sperlich and Pfriem 1986), somewhat suggesting that the frequency of polymorphic inversions may or may not remain approximately constant in natural populations. That shifts of second-chromosome arrangement frequencies in *D. buzzatii* occur in nature is strongly suggested by the latitudinal and altitudinal gradients observed in the native distribution range of the species (Hasson et al. 1995). If the antagonistic pleiotropic effects on fitness-related traits play an important role in the maintenance of the inversion polymorphism in nature (see above), it is not at all surprising that shifts in inversion frequencies occur depending on ecological gradients (for instance, arrangement *2j* is almost fixed in Northern Monte localities of Argentina [see Hasson et al. 1995]). Theoretical models and computer simulations, in addition to further information from molecular markers inside inversion *2j* and closer to the distal and proximal breakpoints than inside-middle *60.29.1sts*, as well as far apart from the inversion breakpoint than our outside-distal marker, may help to better understand the dynamic role played by demographic history, selection, and rate of gene flux in shaping nucleotide variation connected to inversion polymorphisms.

Acknowledgments

We would like to thank Constantina Rodríguez for fly samples, Pilar García for valuable technical advice, Arcadi Navarro, Francisco Rodríguez-Trelles, and Julio Rozas for very helpful discussions and comments on earlier versions of the manuscript, and Julio Rozas for sharing with us unpublished manuscripts and providing the DnaSP version 3.98 software before its public release. The constructive criticisms of two anonymous reviewers were of invaluable help in avoiding some embarrassing mistakes and improving the manuscript considerably. E.H. wishes to thank the Dirección General de Universidades of the Ministerio de Educación, Cultura y Deportes (Spain) for financial support during a sabbatical leave. Fieldwork was supported by grants from Universidad de Buenos Aires and Fundación Antorchas to E.H. This work was supported by grants BOS2000-0295-C02-01 from the Ministerio de Ciencia y Tecnología (Spain) and 2001SGR-00207 from the Direcció General de Recerca (Generalitat de Catalunya) to the G.B.E.

Literature Cited

Adams, M. D., S. E. Celniker, R. A. Holt et al. (192 co-authors). 2000. The genome sequence of *Drosophila melanogaster*. Science **287**:2185–2195.

Aguadé, M., and C. Langley. 1994. Polymorphism and divergence in regions of low recombination in Drosophila. Pp. 67–76 *in* B. Golding, ed. Non-neutral evolution: theories and molecular data. Chapman & Hall, New York.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped Blast and PSI-Blast: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

Andolfatto, P., F. Depaulis, and A. Navarro. 2001. Inversion polymorphism and nucleotide variability in Drosophila. Genet. Res. **77**:1–8.

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

Andolfatto, P., J. D. Wall, and M. Kreitman. 1999. Unusual haplotype structure at the proximal breakpoint of *In(2L)t* in a natural population of *Drosophila melanogaster*. Genetics **153**:1297–1311.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (eds.). 1998. Current protocols in molecular biology. John Wiley & Sons, New York.

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.

Barker, J. S. F., F. M. Sene, P. D. East, and M. A. Q. Pereira. 1985. Allozyme and chromosomal polymorphism of *Drosophila buzzatii* in Brazil and Argentina. Genetica **67**:161–170.

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

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.

Betrán, E., M. Santos, and A. Ruiz. 1998. Antagonistic pleiotropic effect of second-chromosome inversions on body size and early life-history traits in *Drosophila buzzatii*. Evolution **52**:144–154.

Cabrera, A. 1976. Regiones fitogeográficas de la Argentina. Enciclopedia argentina de agricultura y ganadería. Fac. 1 edn. ACME., SACI, Buenos Aires.

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.

Carson, H. L. 1965. Chromosomal morphism in geographically widespread species of Drosophila. Pp. 503–531 *in* H. G. Barker and G. L. stebbins, eds. Genetics of colonizing species. Academic Press, New York.

Carson, H. L., and M. Wasserman. 1965. A widespread chromosomal polymorphism in a widespread species, *Drosophila buzzatii*. Am. Nat. **99**:111–115.

Charlesworth, B., and N. H. Barton. 1996. Recombination load associated with selection for increase recombination. Genet. Res. **67**:27–41.

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

Depaulis, F., L. Brazier, S. Mousset, A. Turbe, and M. Veuille. 2000. Selective sweep near the *In(2L)t* inversion breakpoint 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.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia University Press, New York.

———. 1947. Adaptive changes induced by natural selection in wild populations of Drosophila. Evolution **1**: 1–16.

Federer, W. T., G. D. Steel, and B. Wallace. 1967. Mathematical models for chromosomal inversions. Genetics **55**: 783–795.

Fernández-Iriarte, P., and E. Hasson. 2000. The role of the use of different host plants in the maintenance of the inversion polymorphism in the cactophilic *Drosophila buzzatii*. Evolution **54**:1295–1302.

Filatov, D. 2001. Processor of sequences manual. University of Birmingham. Available at http://www.biosciences.bham.ac.uk/labs/filatov/proseq.html.

Fontdevila, A., A. Ruiz, G. Alonso, and J. Ocaña. 1981. The evolutionary history of *Drosophila buzzatii*. I. Natural chromosomal polymorphism in colonized populations of the Old World. Evolution **35**:148–157.

Fontdevila, A., A. Ruiz, J. Ocaña, and G. Alonso. 1982. The evolutionary history of *Drosophila buzzatii*. II. How much has chromosomal polymorphism changed in colonization? Evolution **36**:843–851.

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.

Gómez, G. A., and E. Hasson. 2003. Transpecific polymorphisms in an inversion linked esterase locus in *Drosophila buzzatii*. Mol. Biol. Evol. **20**:410–423.

Hartl, D. L., D. I. Nurminsky, R. W. Jones and E. R. Lozovskaya. 1994. Genome structure and evolution in Drosophila: applications of the framework P1 map. Proc. Nat. Acad. Sci. USA **91**:6824–6829.

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. Naviera, O. A. Reig, and A. Fontdevila. 1995. The evolutionary history of *Drosophila buzzatii*. XXVI. Macrogeographic patterns of inversion polymorphism in New World populations. J. Evol. Biol. **8**:369–384.

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

———. 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 geographic subdivision. Mol. Biol. Evol. **9**:138–151.

Hudson, R. R., 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.

———. 1988. The coalescent process in models with selection and recombination. Genetics **120**:831–840.

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 the common ancestral sequence in a population. Genetics **147**:1431–1444.

———. 1999. The effect of selection on the amount of nucleotide variation within and between allelic classes. Genet. Res. **73**:15–28.

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

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

Kirby, D. A., and W. Stephan. 1996. Multi-locus selection and the structure of variation at the *white* gene of *Drosophila melanogaster*. Genetics **114**:635–645.

Kovacevic, M., and S. W. Schaeffer. 2000. Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. Genetics **156**:155–172.

Kreitman, M. 1983. Nucleotide polymorphism at the *Alcohol Dehydrogenase* locus of *Drosophila melanogaster*. Nature **304**:412–417.

Krimbas, C. B., and J. R. Powell. 1992. Drosophila inversion polymorphism. CRC Press, Boca Raton, Fla.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

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.

Laird, C. D. 1973. DNA of Drosophila chromosomes. Ann. Rev. Genet. **7**:177–204.

Latorre, A., A. Moya, and F. J. Ayala. 1986. Evolution of mitochondrial DNA in *Drosophila subobscura*. Proc. Natl. Acad. Sci. USA. **83**:8649–8653.

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

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

Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. Genet. Res. **23**:23–35.

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, A., E. Betrán, C. Zapata, and A. Ruiz. 1996. Dynamics of gametic disequilibrium between loci linked to chromosome inversions: the recombination-redistributing effect of inversions. Genet. Res. **67**:67–76.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., K. Kojima, and H. E. Schaffer. 1967. Frequency changes of new inversions in populations under mutation-selection equilibria. Genetics **57**:741–750.

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

Powell, J. R. 1997. Progress and prospects in evolutionary biology: the Drosophila model. Oxford University Press, New York.

Quezada-Díaz, J. E., H. Laayouni, A. Leibowitz, M. Santos, and A. Fontdevila. 1997. Breeding structure of *Drosophila buzzatii* in relation to competition. Genet. Sel. Evol. **29**: 367–382.

Ramos-Onsins, S., and J. Rozas. 2002. Statistical properties of neutrality tests against population growth. Mol. Biol. Evol. **19**:2092–2100.

Roberts, P. A. 1976. The genetics of chromosome aberration. Pp. 67–184 *in* M. Ashburner and E. Novitski, eds. The genetics and biology of Drosophila, Vol. 1a. Academic Press, London.

Robinson M., M. Gouy, C. Gautier, and D. Mouchiroud. 1998. Sensitivity of the relative-rate test to taxonomic sampling. Mol. Biol. Evol. **15**:1091–1098.

Robinson-Rechavi, M., and D. Huchon. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics **16**:296–297.

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.

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.

Rozas, J., and M. Aguadé. 1990. Evidence of genetic exchange in the *rp49* region among polymorphic chromosome inversions in *Drosophila subobscura*. Genetics **126**:417–426.

———. 1993. Transfer of genetic information in the *rp49* region of *Drosophila subobscura* between different chromosomal gene arrangements. Proc. Natl. Acad. Sci. USA. **90**: 8083–8087.

———. 1994. Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of Drosophila. Proc. Natl. Acad. Sci. USA. **91**:11517–11521.

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.

Ruiz, A., A. Fontdevila, and M. Wasserman. 1982. The evolutionary history of *Drosophila buzzatii*. III. Cytogenetic relationships between two sibling species of the *buzzatii* cluster. Genetics **101**:503–518.

Ruiz, A., M. Santos, A. Barbadilla, J. E. Quezada-Díaz, E. Hasson, and A. Fontdevila. 1991. Genetic variance for body size in a natural population of *Drosophila buzzatii*. Genetics **128**:739–750.

Russo, C. A. M., N. Takezaki, and M. Nei. 1995. Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. **12**:391–404.

Saitou, N., and M. Nei. 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

Santos, M. 1986. The role of genic selection in the establishment of inversion polymorphism in *Drosophila subobscura*. Genetica **69**:35–45.

Schafer, D. J., D. K. Fredline, W. R. Knibb, M. M. Green, and J. S. F. Barker. 1993. Genetics and linkage mapping of *Drosophila buzzatii*. J. Hered. **84**:188–194.

Sperlich, D. and P. Pfriem. 1986. Chromosomal polymorphisms in natural and experimental populations. Pp. 257–309 *in* M. Ashburner, H. L. Carson, and J. N. Thompson, Jr., eds. Genetics and biology of Drosophila, Vol. 3e. Academic Press, London.

Strobeck, C. 1983. Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. Genetics **103**: 545–555.

Swofford, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sinauer Associates, Sunderland, Mass.

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

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

Vilela, C. A., F. Sene, and M. A. Q. R. Pereira. 1980. On the Drosophila fauna of Chaco and east slopes of the Andes in Argentina. Rev. Bras. Biol. **40**:837–841.

Wall, J. D. 2000. A comparison of estimators of the population recombination rate. Mol. Biol. Evol. **17**:156–163.

Wasserman, M. 1982. Evolution of the *repleta* group. Pp. 61–139 *in* M. Ashburner, H. L. Carson, and J. N. Thompson, Jr., eds. The genetics and biology of Drosophila, Vol. 3b. Academic Press, New York.

Watterson, G. A. 1975. On the number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7**:256–276.

Weir, B. S. 1996. Genetic data analysis II. Sinauer Associates, Sunderland, Mass.

Wesley, C. S., 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**:3132–3126.

Whitlock, M. C., P. C. Phillips, F. B.-G. Moore, and S. J. Tonsor. 1995. Multiple fitness peaks and epistasis. Annu. Rev. Ecol. Syst. **26**:601–629.

Wolfgang Stephan, Associate Editor

Accepted January 29, 2003