

Molecular Population Genetics of the α -Esterase5 Gene Locus in Original and Colonized Populations of *Drosophila buzzatii* and Its Sibling *Drosophila koepferae*

R. V. Piccinali,¹ L. J. Mascord,² J. S. F. Barker,² J. G. Oakeshott,³ E. Hasson¹

¹ Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón 2, Ciudad Universitaria C1428EHA, Buenos Aires, Argentina

² School of Rural Science and Agriculture, Faculty of the Sciences, University of New England, Armidale, Australia

³ Division of Entomology, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia

Received: 22 September 2005 / Accepted: 21 July 2006 [Reviewing Editor: Dr. Richard Kliman]

Abstract. Several studies have suggested that esterase-2 (EST-2) may be the target of natural selection in the cactophilic fly *Drosophila buzzatii*. In this work, we analyzed nucleotide variation in a fragment of α -esterase5 ($\alpha E5$), the gene encoding EST-2, in original (Argentinian) and colonized (Australian) populations of *D. buzzatii* and in its sibling *D. koepferae*. Estimates of nucleotide heterozygosity in *D. buzzatii* were similar in Australia and Argentina, although we detected a loss of singletons in colonized populations, suggesting a moderate founder effect. Interspecific comparisons revealed that *D. buzzatii* was more polymorphic for nonsynonymous variation, whereas *D. koepferae* was more variable for synonymous and noncoding sites. The two major chromosomal arrangements (*2st* and *2j*) in *D. buzzatii* displayed similar levels of nucleotide variation, whereas *2jz*³ was monomorphic. The sequenced region allowed the discrimination of a greater number of EST-2 protein variants in the Australian sample than in the Argentinian sample. In *D. koepferae*, nucleotide variation in $\alpha E5$ does not depart from neutral expectations, although tests of population structure were significant for silent variation. In contrast, *D. buzzatii* has probably undergone a recent population expansion in its South American range. In addition, the McDonald and Kreitman test revealed an excess of nonsynonymous polymorphism in both original and colonized populations of this species.

Key words: $\alpha E5$ — Cactophilic *Drosophila* — Excess of nonsynonymous polymorphism — Founder effect — Population expansion

Introduction

The carboxyl/cholinesterases are a particularly variable gene family, with very high rates of molecular evolution and low amino-acid identities among its members (Cygler et al. 1992; Hemila et al. 1994; Oakeshott et al. 1999, 2005). Rapid sequence change has resulted in the diversification of carboxyl/cholinesterases both in sequence identity and biochemical function in correlation with ecological and behavioral changes in their host species (Oakeshott et al. 1999, 2005).

Two clusters of carboxyl/cholinesterases genes have been described in detail in *Drosophila* spp. These are the α - and β -esterase clusters, denominations that generally reflect the ability of the encoded enzymes to hydrolyze α - and β -naphthyl-acetate, respectively. The α -esterase cluster has been characterized in *D. melanogaster* (Robin et al. 1996) and in the cactophilic *D. buzzatii* (Robin et al. 2000). In the latter, the cluster is comprised of 11 genes encoding functional esterases, 8 of which are orthologous with genes of the α -cluster of *D. melanogaster*. These results indicate that most duplication events involved in the

origin of this family occurred previous to the divergence of the subgenera *Sophophora* and *Drosophila* (to which *D. melanogaster* and *D. buzzatii*, respectively, belong), that is, < 40 million years ago (Russo et al. 1995).

The EST-2 (esterase-2) protein in *D. buzzatii* is the major α -naphthyl acetate-staining esterase on starch gels. Like the majority of esterase enzymes, the natural substrate and biological function of EST-2 are unknown. The protein is expressed principally in the alimentary tract, midgut, and malpighian tubules (East 1982). Biochemical and physiological analyses have indicated that it belongs to the carboxyl/cholinesterases multigene family, and it has been suggested that it is orthologous to EST-9 in *D. melanogaster* (East et al. 1990; Oakeshott et al. 1993). Recently, Campbell et al. (2003) used proteomics and mass spectrophotometry to show that EST-9 is encoded by the $\alpha E5$ gene in the α -esterase cluster, and Mascord (2004) has used a genetic approach to show that EST-2 is encoded by the $\alpha E5$ ortholog in *D. buzzatii*. Mascord (2004) obtained the full sequence of the $\alpha E5$ gene from several *D. buzzatii* lines selected to cover EST-2 starch mobility alleles and showed that the surface charges of the predicted $\alpha E5$ proteins explained the relative mobilities of the EST-2 allozymes and that an EST-2 null line had a frame-shifting deletion in the $\alpha E5$ gene.

The α -esterase cluster has been mapped to the second chromosome of *D. buzzatii* between cytological bands *F5e* and *F6a*, a region close to, but outside of, the proximal break points of the two more common inversions, $2j$ and $2z^3$ (Ranz et al. 2003). The $2j$ arrangement is a simple paracentric inversion of the *D. buzzatii* standard sequence ($2st$), whereas the $2z^3$ inversion occurred on a $2j$ chromosome and gave rise to the $2jz^3$ arrangement, which includes the segment inverted by $2j$. The proximity of the α -esterase cluster to the break points of naturally occurring inversions probably explains, at least in part, the linkage disequilibrium of EST-2 variants with second-chromosome arrangements (Knibb et al. 1987; Betrán et al. 1995; Rodríguez et al. 2001).

The population genetics of allozyme variation of EST-2 in *D. buzzatii* has been examined in both native populations from Argentina and recently colonized areas of Australia and the Mediterranean Basin. Seven starch gel electromorphs (EST-2^a, EST-2^{a-}, EST-2^b, EST-2^c, EST-2^{c+}, EST-2^d, and EST-2^e) have been identified in Argentinian populations, with all except EST-2^{a-} detected also in Australia (J. S. F. Barker, 1974–1996, unpublished data). Furthermore, using sequential cellulose acetate gel electrophoresis, a more discriminating electrophoretic technique, the number of alleles that could be distinguished in two Australian populations was extended from 6 to 25 (Barker 1994).

Only one study into the population genetics of the variants detected by sequential cellulose acetate electrophoresis has been conducted, and this could not fully discriminate between natural selection and drift as factors accounting for the differences observed between populations (Barker 1994). However, several studies on the population genetics of the major starch gel variants have indicated that this protein may be the target of natural selection. Some of these major EST-2 variants show parallel latitudinal clines in Australia and Argentina, which are largely independent of inversion frequencies (Knibb and Barker 1988; Rodríguez et al. 2000). Populations are genetically structured for these variants (Sokal et al. 1987), and differentiation is correlated with ecologically relevant variables, particularly with the diversity of phytogeography and cactus in Argentina (Rodríguez et al. 2000). Other lines of evidence include responses to genetic perturbation experiments in natural populations (Barker and East 1980; Barker et al. 1989) and to thermal shocks (Watt 1981). Furthermore, the starch gel variation seems to be associated with differential attraction to and viability on diverse cactus hosts (Fernández-Iriarte et al. 2002) and with oviposition site preferences for different yeast species (Barker et al. 1981, 1986; Barker and Starmer 1999).

The recent confirmation that EST-2 in *D. buzzatii* is orthologous to EST-9 in *D. melanogaster* and encoded by $\alpha E5$ presents an opportunity to examine the evolutionary forces that have shaped the EST-2 variation at a molecular level. We report the results of a comparative study of $\alpha E5$ nucleotide variation in populations of *D. buzzatii* and its sibling species *D. koepferae*. Specifically, we compared nucleotide variability in original and colonized populations of *D. buzzatii*, to search for signs of a founder effect, and between *D. buzzatii* and *D. koepferae*. We also analyzed levels of nucleotide variation within and between the main chromosomal arrangements segregating in natural populations of *D. buzzatii* and within and between EST-2 electrophoretic alleles. Finally, we searched for evidence of population structure at this locus and investigated if $\alpha E5$ nucleotide variation is compatible with neutral expectations.

Materials and Methods

Drosophila Lines

Eighty-five *D. buzzatii* lines isogenic for the second chromosome and derived from wild flies collected in Argentina ($n = 39$) and Australia ($n = 46$) were used in the present study. Samples were collected in (1) 11 Argentinian localities spanning 5 phytogeographic regions in autumn 1995, spring 1996, and autumn 1997, and in (2) 2 Australian localities in summer 1984 and spring 1985 (Supplementary Table 1). Isogenic flies were obtained by crossing

wild male flies individually with virgin female flies of the balancer stock *Antp/D⁺* as described in Barker (1994) and Rodríguez et al. (2000). Each line was characterized cytologically and electrophoretically to ascertain the inversion karyotype and the EST-2 genotype, and specimens were stored at -80°C until DNA extraction. The lines from Argentina were a representative sample of the *2st*, *2j*, and *2jz³* inversions, and their frequencies were not significantly different ($\chi^2 = 0.09$; $p = 0.96$) from global estimates in previous surveys (Rodríguez et al. 2000). The isogenic lines from Australia were randomly selected from the 147 lines characterized by Barker (1994) for additional allozyme variation at EST-2 using sequential gel electrophoresis. Equal numbers of the *2st* and *2j* inversions for each site were selected at random, except that only 10 sequences from the *2st* lines from the Trinkey population were obtained. A constructed random sample (CRS; Hudson et al. 1994) was also analyzed for comparative purposes.

Twenty *D. koepferae* lines derived from flies collected in autumn 1997 and 1999 in eight localities of Argentina were also analyzed (Supplementary Table 1). Wild flies were captured by means of net sweeping on fermented banana baits, transported to the laboratory, sexed, and female flies placed in individual vials. After four generations of sib mating, progeny individuals of each isofemale line were frozen and stored at -80°C until DNA extraction. Neither the inversion karyotype nor the EST-2 genotype was known for these flies.

DNA Sequencing

Genomic DNA was obtained from individual flies using a Puregene kit (GENTRA Systems) according to the manufacturer's protocol. $\alpha E5$ fragments of 1.9 or 1.6 kb were polymerase chain reaction (PCR) amplified using primers 151+ (sense 5'CCGCAGCCTTACGATGATG3') and 2082- (antisense 5'AGACTATCCCAGACGAGCAG3') or primers 5.5b (sense 5'CTAACGGGCATCATAGAAGGCAG3') and 5.6b (antisense 5'GCAGTTGGGATTGG AAGTAGCTG3'), respectively. If the yield of amplifications was high, two 50- μL reactions were run in a 0.8% agarose gel and the DNA band purified using Qiaquick spin columns (Qiagen). If the yield of the first PCR was not as high, a smaller nested fragment was reamplified with primers 337+ (sense 5'GGACTGCCTTTATCTGAATG3') and 1998- (antisense 5'TGCTCCTGCTTCGCTGCTG3'), and this PCR product was purified in spin columns.

All PCR reactions were carried out in final volumes of 50 μL containing 8 μL dNTPs (1.25 mM each), 5 μL 10x reaction buffer, 1 μL MgCl_2 (1mM), 0.2 μL 5U/ μL *Taq* DNA polymerase (Invitrogen), 90 ng sense and antisense primers, and 50 to 100 ng DNA. Amplifications were performed according to the following PCR profile: one cycle of 1 minute at 94°C ; 35 cycles of denaturation (1 minute at 94°C), annealing (1 minute at 60°C), and extension (1.5 minutes at 72°C); and a final extension step of 3.5 minutes at 72°C . Purified DNAs were used as templates for direct sequencing of both strands using primers 5.21 (sense 5'GACCCAGCTGTGAGGTG3'), 5.11 (antisense 5'GTGGGAGCCGTGGCGTATG3'), and 1866- (antisense 5'TTCCACGACTCCGAGTTACG3') with the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit, version 2.0 (Perkin Elmer) in a Perkin Elmer ABI PRISM 3700 DNA Analyzer Automated Sequencer. This sequencing strategy allowed us to obtain the nucleotide sequence between positions 684 and 1486 of the published $\alpha E5$ sequence (Robin et al. 2000; GenBank accession no. AF216213). This fragment covers 45% of the coding region of the gene and includes part of the third and fourth exons (699 and 39 bp, respectively) and the complete fourth intron (64 bp). This region was selected for sequencing because it showed high sequence diversity among Australian allozyme alleles, and nucleotide differences in $\alpha E5$ alleles identified in this region distinguished particular EST-2 sequential gel electromorphs (Mascord 2004).

In *D. koepferae*, unlike *D. buzzatii*, individuals were not isogenic, and several heterozygous sites were detected per individual. A nucleotide position was classified as heterozygous if two overlapping signals were consistently observed at the same site in the chromatograms of two or three sequencing reactions. When the analytical software allowed it, data were treated as genotypic with unknown gametic phase in further analyses (Arlequin 2.001; Schneider et al. 2000). Otherwise, two sequences per individual were created by distributing the segregating sites at random between each of them (DnaSP 3.97; Rozas and Rozas 1999).

Partial and complete sequences of each line were aligned manually or using Clustal X (Higgins et al. 1994). Newly reported sequences have been deposited at GenBank under accession nos. DQ192063 to DQ192108 and DQ204622 to DQ204680.

Data Analyses

Estimates of nucleotide diversity were obtained using Watterson's (θ_W) and Tajima's (π) estimators, based on the number of segregating sites and the average number of pairwise differences per site, respectively (Watterson 1975; Tajima 1989). Interspecific divergence was calculated according to the method of Jukes and Cantor (1969).

The population recombination parameter was estimated using the method of Hudson and Kaplan (1985), based on the minimum number of recombination events in a sample (R_m). Nucleotide differentiation among inversion arrangements in *D. buzzatii* for the $\alpha E5$ gene region was analyzed by means of K_{ST^*} (Hudson et al. 1992), and S_{mm} (Hudson 2000) statistics. Significant associations between electrophoretic EST-2 variants and $\alpha E5$ haplotypes were tested using a linkage disequilibrium analysis based on D coefficients (Lewontin and Kojima 1960). Genealogical relationships between haplotypes were estimated as a haplotype network, based on the algorithm proposed by Templeton et al. (1992). Population structure was investigated by means of analysis of molecular variance (AMOVA; Excoffier et al. 1992) for total, nonsynonymous, synonymous, and silent variations. For these analyses, sequences of Argentinian *D. buzzatii* and *D. koepferae* were arranged in two hierarchical levels, the first according to the collection site and the second to the phylogeographic region in which each site is located (Supplementary Table 1). For the Australian sample, sequences were assembled according to collection sites only.

The tests of Tajima (1989), Fu and Li (1993), and Ramos-Onsins and Rozas (2002) were applied to determine whether the frequency spectrum of polymorphic variants departs from neutral expectations. We also employed the McDonald and Kreitman (1991) (MK) test to examine the prediction of the neutral theory of a correlation between polymorphism and divergence for the ratio of synonymous to nonsynonymous variation. The neutrality index (NI) proposed by Rand and Kann (1996) was used as a qualitative indicator of the degree and direction of departures from neutrality in the ratio of polymorphism to divergence. The index is defined as $\text{NI} = (P_r/F_r)/(P_s/F_s)$, where P_r and P_s are the number of replacement and synonymous polymorphic sites, respectively, and F_r and F_s are the number of replacement and synonymous fixed differences between species, respectively. Under strict neutrality, NI is expected to be equal to 1. A value >1 indicates an excess of amino-acid variation within species, whereas a value <1 indicates an excess of fixed amino-acid differences between species (Rand & Kann 1996).

All tests that required an outgroup were performed using the other species (i.e., *D. koepferae* for *D. buzzatii* and vice versa) or *D. borborema* (GenBank accession no. DQ453800), another member of the *D. buzzatii* complex (Ruiz and Wasserman 1993; Manfrin et al. 2001).

DnaSP 3.97 (Rozas & Rozas 1999) was used for all statistical analysis, except for AMOVAs and linkage disequilibrium analyses,

Table 1. Estimates of nucleotide heterozygosity in *D. buzzatii* and *D. koepferae* for all, synonymous, nonsynonymous and noncoding sites in the $\alpha E5$ gene region

Variation	Species	N	N°sites	S	η	Singl	θ_w (SD)	π (SD)
Total	<i>D. buzzatii</i> Arg	38	802	29	29	20 (68.97)	0.0086 (0.0029)	0.0051 (0.0026)
	<i>D. buzzatii</i> Aus	46	802	25	26	6 (24.00)	0.0071 (0.0024)	0.0066 (0.0032)
	<i>D. buzzatii</i> Aus CRS	33	802	24	25	9 (37.50)	0.0074 (0.0026)	0.0063 (0.0031)
	<i>D. koepferae</i>	40	804	34	37	10 (29.41)	0.0099 (0.0017)	0.0092 (0.0020)
Nonsynonymous	<i>D. buzzatii</i> Arg	38	570	14	14	10 (71.43)	0.0058 (0.0021)	0.0034 (0.0019)
	<i>D. buzzatii</i> Aus	46	571	16	17	4 (25.00)	0.0064 (0.0023)	0.0055 (0.0029)
	<i>D. buzzatii</i> Aus CRS	33	571	16	17	7 (43.75)	0.0069 (0.0026)	0.0050 (0.0027)
	<i>D. koepferae</i>	40	568	9	9	3 (33.33)	0.0037 (0.0012)	0.0025 (0.0014)
Synonymous	<i>D. buzzatii</i> Arg	38	168	13	13	8 (61.54)	0.0184 (0.0072)	0.0125 (0.0071)
	<i>D. buzzatii</i> Aus	46	167	8	8	1 (12.50)	0.0109 (0.0048)	0.0125 (0.0072)
	<i>D. buzzatii</i> Aus CRS	33	167	7	7	0 (0.00)	0.0103 (0.0048)	0.0126 (0.0073)
	<i>D. koepferae</i>	40	170	20	21	5 (25.00)	0.0277 (0.0062)	0.0271 (0.0075)
Noncoding	<i>D. buzzatii</i> Arg	38	64	2	2	2 (100)	0.0074 (0.0055)	0.0016 (0.0028)
	<i>D. buzzatii</i> Aus	46	64	1	1	1 (100)	0.0036 (0.0036)	0.0007 (0.0018)
	<i>D. buzzatii</i> Aus CRS	33	64	1	1	1 (100)	0.0038 (0.0039)	0.0010 (0.0021)
	<i>D. koepferae</i>	40	66	6	7	2 (33.33)	0.0249 (0.0087)	0.0205 (0.0109)

N, number of sequences; N°sites, number of sites; S, number of segregating sites; η , total number of mutations; Singl, number of singletons (percentages in parentheses); θ_w , Watterson's estimator; π , mean number of pairwise differences per site.

which are subroutines of the program Arlequin 2.001 (Schneider et al. 2000). The haplotype network was built with TCS 1.21 (Clement et al. 2000).

The significance of the AMOVAs, K_{ST}^* and S_{mn} tests were obtained by means of 10000 permutations of haplotypes (or genotypes) between chromosomes, alleles, populations or regions. The statistical significance of neutrality tests was calculated by 1000 coalescent simulations based on a Monte Carlo process (Hudson 1990), assuming the most conservative option of no recombination.

Results

Nucleotide Variation Within and Between Species

The fragment of the $\alpha E5$ gene region sequenced in this study was polymorphic at 29 and 25 sites in the Argentinian and Australian samples of *D. buzzatii*, respectively, and at 34 sites in the *D. koepferae* samples (Table 1; Supplementary Figs. 1 and 2). No shared polymorphisms were detected between the two species.

A 28-bp deletion between sites 337 and 364 was detected in the *D. buzzatii* line Tir-17. Although this deletion results in a frame-shift mutation, it did not appear to be lethal because several adult individuals homozygous for this variant were recovered by means of the balancer stock *Antp*/ Δ^5 (data not shown). Although this indicates that EST-2 function is not critical for survival in the laboratory, we note that null alleles only occur in natural populations at low frequencies (0.01% in Australia [Barker 1994] and 0.09% in Argentina [Rodríguez et al. 2001]), suggesting some selective value for EST-2 function in the field environment. This line was excluded from all analyses. We also detected a fixed indel difference of 2

bp (GC) between *D. koepferae* and *D. buzzatii* at intron position 725.

Argentinian and Australian *D. buzzatii* had 18 polymorphic sites in common (Supplementary Figure 1), which means that >60% of the polymorphic variants in the original area were also present in the founder population that reached Australia. Remarkably, the nine parsimony-informative sites detected in Argentina are all present in Australia. These results are in agreement with the predictions of population genetic theory that alleles at lower frequencies are more likely to be lost during a founder event (Hedrick 1983). Alternative hypotheses for these findings involve recent gene flow or recurrent mutation.

Levels of nucleotide heterozygosity estimated by means of θ_w and π are listed in Table 1. Watterson's estimate of nucleotide heterozygosity (θ_w) indicates that total polymorphism was only slightly higher in *D. koepferae* than in *D. buzzatii*. However, when the average number of pairwise differences per site (π) was considered, this region of $\alpha E5$ in *D. koepferae* showed a level of polymorphism >1.5 times higher than in *D. buzzatii*. This interspecific difference detected in $\alpha E5$ is different from the patterns observed in both the *Xdh* and *Est-A* gene regions, where *D. buzzatii* was twice as polymorphic as *D. koepferae* (Gómez and Hasson 2003; Piccinali et al. 2004). The interspecific difference in $\alpha E5$ was even more striking when nonsynonymous and synonymous variation were considered separately. The percentage of nonsynonymous polymorphic sites was high in both Argentinian (48%) and Australian *D. buzzatii* (64%), whereas the percentage of nonsynonymous polymorphisms was much lower in *D. koepferae* (26%)

(Supplementary Figs. 1 and 2). This species, in contrast, was more polymorphic at silent sites (11%, synonymous and noncoding) than *D. buzzatii* (6%).

Argentinian *D. buzzatii* exhibited a high proportion of singleton variants (close to 70%) in the $\alpha E5$ gene (Table 1), similar to the values reported for the *Xdh* locus in Argentinian *D. buzzatii* (Piccinali et al. 2004). Conversely, unique variants were noticeably lower in Australian *D. buzzatii* (24% of total variation; 25% synonymous and 13% nonsynonymous sites) and in *D. koepferae* (29%, 33%, and 25%, respectively). Results based on the Australian CRS were similar to those obtained for the complete Australian data set.

Estimates of the recombination parameter according to R_M were equal to 4/gene and 0.0050/site in the total sample, whereas the values for both continents analyzed separately were exactly the same and equal to 3/gene and 0.0037/site.

Haplotype diversity in the *D. buzzatii* dataset was quite similar in Argentina (0.936) and Australia (0.926), but haplotype composition was very different in the two continents. Only 7 of the 34 haplotypes detected were present in both samples (Supplementary Figure 1). The Argentinian *D. buzzatii* sample contained 23 haplotypes, with 1 major and widely distributed haplotype (haplotype 10; frequency 21%), recorded from the Northwest (lines *Til-3*, *Quil-7*, and *Quil-19*) and the Central West (lines *Chu-2*, *Chu-20*, *Chu-38*) to the southernmost limit of the species (lines *Ota-3* and *Ota-4*). It was also detected in the two major chromosomal arrangements and in $2jz^3$ (Supplementary Figure 1). Remarkably, it appeared to be absent in Australia. Conversely, the Australian sample, which contained 18 haplotypes, included 2 haplotypes with moderate frequencies (haplotypes 24 and 30, each one with frequencies of 17%) that were not present in Argentina.

Quite apart from the differences in *D. buzzatii* haplotypes between continents, the high frequencies of the nominated haplotypes in each continent are themselves inconsistent with neutral expectations. To show this, we used coalescent simulations (Hudson et al. 1994) to calculate for each continent the likelihood of having a subset of n sequences with a haplotype diversity equal to zero, given the observed mean heterozygosity in the total sample for that continent. The observed values of haplotypic diversity were much lower than those realized in the coalescent simulations in both Argentina and Australia ($p < 0.0001$). Thus, the presence of such haplotypes in relatively high frequencies in the respective continents is not compatible with neutral expectations.

Sequence divergence between *D. buzzatii* and *D. koepferae* was 4.7% for all sites, 1.5% for nonsynonymous sites, and 15.3% for synonymous sites. These estimates were similar to the divergence between

D. koepferae and *D. borborema* (4.5%, 1%, and 16.6% for total, nonsynonymous, and synonymous sites, respectively). However, divergence between *D. buzzatii* and *D. borborema* was slightly higher for all sources of variation (6%, 2%, and 23%). These results concur with previous work in finding that phylogenetic relationships among certain species of the *Drosophila buzzatii* complex are ambiguous. Some studies suggest that *D. buzzatii* and *D. koepferae* are sister groups (Durando et al. 2000; Manfrin et al. 2001) whereas others support a closer relationship between *D. koepferae* and *D. borborema* (Rodríguez-Trelles et al. 2000).

Differentiation Among EST-2 Electrophoretic Variants

Our samples contained six EST-2 starch gel variants in the Argentinian sequences (EST-2^a, EST-2^{a-}, EST-2^b, EST-2^c, EST-2^{c+}, and EST-2^d) and six in the Australian sequences (EST-2^a, EST-2^b, EST-2^c, EST-2^{c+}, EST-2^{cf}, and EST-2^d). This spectrum is broadly similar to previous findings (Knibb et al. 1987; Rodríguez et al. 2000), except that this sampling did not include the relatively uncommon EST-2^e variant but did include EST-2^{cf}, another rare variant, that shows faint staining around the position of the EST-2^c on starch gels (Knibb et al. 1987).

Six significant associations existed between haplotypes and starch gel electrophoretic variants in Argentina, and only two of these involved >1 individual (EST-2^c with haplotype 3 and EST-2^a with haplotype 9; Table 2). In Australia, 12 significant associations existed, of which 9 involved at least 2 individuals (Table 2). Most of the haplotypes involved in significant associations were only associated with a single electrophoretic variant in each sample, although haplotype 16 was found in both EST-2^c and EST-2^{cf} in Australia, and the only associations that were the same in the two samples were those involving EST-2^d. The most common Argentinian haplotype (10) was not significantly associated with any particular electrophoretic variant, but the two high-frequency Australian haplotypes (24 and 30) were associated exclusively with EST-2^a.

The haplotype network (Fig. 1) shows the genealogical relationships within and between EST-2 electrophoretic classes. Argentinian and Australian haplotypes were mixed throughout the network, showing that some Australian haplotypes are closer to Argentinian variants than to other Australian alleles, as expected because of the recent colonization of Australia by *D. buzzatii*. Sequences associated with EST-2^a, EST-2^{a-} or EST-2^b were scattered widely through the network, but haplotypes associated with EST-2^c, EST-2^{cf}, EST-2^{c+} and EST-2^d were more clustered.

Table 2. Linkage disequilibrium between EST-2 allozymes and $\alpha E5$ haplotypes^a

Sample	All	Hap	D (D')	N
Argentina	a	9	0.04 (1.00)*	2
	a-	21	0.03 (1.00)***	1
	c	3	0.08 (0.73)***	4
	c+	12	0.03 (1.00)***	1
	d	5	0.06 (1.00)***	1
	d	7	0.03 (1.00)***	1
Australia	a	24	0.09 (1.00)***	8
	a	30	0.09 (1.00)***	8
	b	2	0.03 (1.00)*	2
	b	28	0.03 (1.00)*	2
	b	34	0.03 (1.00)*	2
	b	2	0.03 (1.00)*	2
	c	3	0.04 (1.00)***	2
	c	16	0.02 (0.47)*	1
	cf	16	0.02 (1.00)***	1
	c+	2	0.10 (1.00)***	5
	d	5	0.06 (1.00)***	5
	d	7	0.02 (1.00)**	1

^aOnly significant values are shown.

All, allozymes; Hap, haplotype; D , Lewontin and Kojima (1960) coefficient; D' , standardized D coefficient (Lewontin 1974); N, number of observed individuals.

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

Differentiation Among Chromosomal Arrangements in *D. buzzatii*

Levels of nucleotide variation within the two major arrangements $2st$ and $2j$ were very similar ($2st$ in Australia: $\theta_w = 0.0062 \pm 0.0025$, $\pi = 0.0064 \pm 0.0033$, $2j$ in Australia: $\theta_w = 0.0080 \pm 0.0030$, $\pi = 0.0064 \pm 0.0032$, $2st$ in Argentina: $\theta_w = 0.0076 \pm 0.0033$, $\pi = 0.0055 \pm 0.0029$, and $2j$ in Argentina: $\theta_w = 0.0067 \pm 0.0027$, $\pi = 0.0054 \pm 0.0028$), whereas the sample of $2jz^3$ chromosomes was completely monomorphic. Of the total of 38 variable sites in the 2 continents, 9 and 8 were exclusive to $2st$ and $2j$, respectively, and the remaining 21 were found in both (Supplementary Table 1).

We can use the calculations of Innan and Tajima (1997) to show that these results for $2st$ and $2j$ are in agreement with theoretical expectations. The authors estimated the expected number of pairwise differences within and between two allelic classes when the ancestral allelic class is known and found that the amount of variability within each allelic class changes with their frequencies. The expected level of nucleotide variation within $2st$ and $2j$ in each continent can be obtained from the values reported in Table 2 of Innan and Tajima (1997), taking into account the respective frequencies of both arrangements and excluding $2jz^3$ chromosomes from the Argentinian sample. According to the accepted chromosomal phylogeny, arrangement $2st$ is ancestral to $2j$ (Ruiz and Wasserman 1993), and the average frequencies of

$2st$ in Australia and Argentina are 0.27 and 0.35, respectively. The expected differences within $2st$ and $2j$ in Australia are fairly equal (0.449 θ and 0.451 θ , respectively), which is exactly coincident with the observed values. In Argentina, the expected values are 0.546 θ and 0.403 θ , indicating that $2st$ is predicted to harbor 35% more variation than $2j$, but such a difference is still within the 95% confidence intervals of our estimates.

Innan and Tajima (1999) have also shown that the ratio of the sum of the average number of pairwise differences within allelic classes to the average pairwise number of differences between all sequences is expected to be 1 in the absence of recombination. In the $\alpha E5$ gene region, the values of the ratio were 1.96 and 1.76 in Australian and Argentinian samples, respectively. Such values suggest negligible levels of differentiation and moderate recombination between arrangements. This is concordant with the high proportion of shared polymorphic variants between $2st$ and $2j$ chromosomes.

In fact, differentiation among arrangements within Argentina and Australia was nonsignificant after Bonferroni's correction (data not shown), although there was a certain degree of differentiation across the continents. Argentinian $2jz^3$ was significantly differentiated from Australian $2st$ ($K_{ST^*} = 0.092$, $p = 0.001$; $S_{nn} = 0.985$, $p < 0.0001$) and $2j$ ($K_{ST^*} = 0.053$, $p = 0.004$; $S_{nn} = 1.000$; $p < 0.0001$) and Argentinian $2j$ from Australian $2st$ ($K_{ST^*} = 0.049$, $p = 0.004$; $S_{nn} = 0.832$; $p < 0.0001$). However, comparisons involving $2jz^3$ should be interpreted with caution because of the absence of variation within it as well as its small sample size (Charlesworth 1998).

Geographic Differentiation

Because our samples consisted of sequences of different geographic origin, and given that electrophoretic surveys of genetic variation have shown that *D. buzzatii* populations are significantly differentiated for EST-2 variant frequencies (Knibb and Barker 1988; Rodríguez et al. 2000), we also examined nucleotide variation at $\alpha E5$ in *D. buzzatii* and *D. koepferae* for patterns of population subdivision.

The AMOVAs showed that genetic differentiation among populations within and among regions for total, nonsynonymous, synonymous, and silent variation was not significant either in Argentinian or Australian *D. buzzatii* (results not shown). Because the small number of sequences sampled per population and the large within-population variance can mask a structured pattern, we also repeated the analyses in Argentina, including only those populations with at least four sequences (Termas de Río

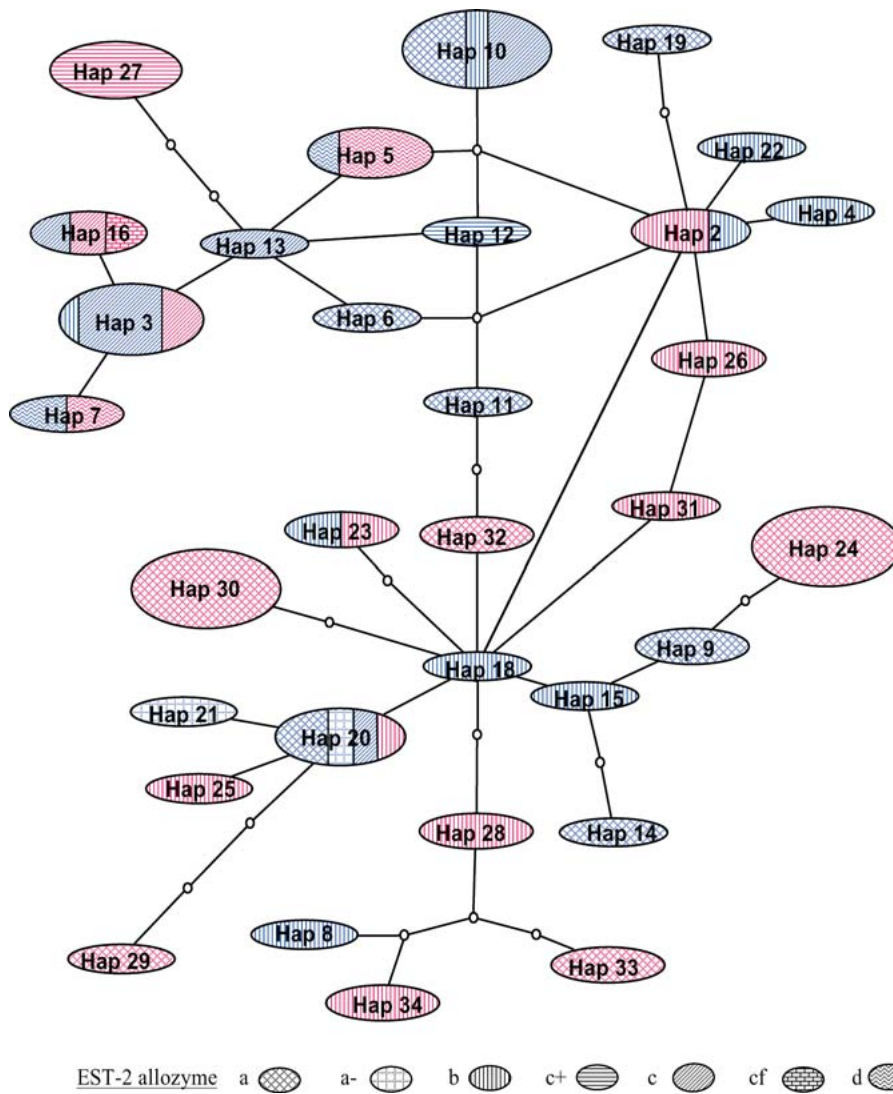


Fig. 1. Haplotype network for *D. buzzatii* $\alpha E5$ sequences. The network was built with the algorithm of Templeton et al. (1992) as implemented in TCS 1.21 (Clement et al. 2000). Each branch represents a mutational step. Small empty circles represent hypothetical haplotypes. The sizes of circles enclosing observed haplotypes are proportional to their frequencies in the whole sample. Blue color: Argentinian haplotypes. Red color: Australian haplotypes. Haplotype 1 was excluded because the number of mutational steps separating it from the closest neighbor was too high, compromising the confidence limits of 95% for the statistical parsimony analysis (Templeton et al. 1992). The Australian variant EST-2^{c+} has an unusual triple-banding pattern around the position of the EST-2^c variant on starch gels, whereas the rare variant EST-2^{cf} shows faint staining at the EST-2^c position.

Hondo, Salta, Catamarca and Tilcara, which belong to four different phylogeographic regions). However, the results were again not significant.

D. koepferae did not exhibit signs of population structure either when total or nonsynonymous variation was considered, but a pattern of population structure emerged from silent variation. Differentiation between regions was significant ($\sigma_a^2 = 0.671$, $\phi_{CT} = 0.166$, $df = 3$; $p = 0.044$) and accounted for 17% of total silent variation, whereas differentiation between populations within regions was not significant ($\sigma_b^2 = -0.286$, $\phi_{SC} = -0.085$, $df = 4$; $p = 0.822$).

Neutrality Tests

Fu and Li's *D* and *F* statistics were negative and highly significant for total, nonsynonymous, and synonymous variation in Argentinian populations of *D. buzzatii* (Table 3), pointing to an excess of low-frequency variants. These tests were not significant in

either Australian *D. buzzatii* or *D. koepferae*, and the results were congruent when no outgroup was used (results not shown). The *R*₂ statistic, which compares the observed number of singletons with those expected under a major recent population growth event, was significant only for the total variation in the Argentinian *D. buzzatii* sample.

The MK test was significant when either Argentinian or Australian *D. buzzatii* were compared with *D. koepferae* or *D. borborema* as outgroups (Table 4). In all of these comparisons, the ratio of nonsynonymous polymorphism to nonsynonymous fixed differences was higher than the ratio of synonymous polymorphism to synonymous fixed differences (NI values 3 to 6.2), pointing to an excess of nonsynonymous polymorphism or a deficit of nonsynonymous fixed differences. The Australian CRS gave results qualitatively similar to those for the whole Australian sample (*D. koepferae*: G-value: 11.97, p : 0.0005, NI: 7.63; *D. borborema*: G-value: 13.71, p : 0.0002, NI: 7.53).

Table 3. Tajima's (D_T), Fu and Li's (D_{FL} and F_{FL}) and Ramos-Onsins and Rozas' (R_2) tests for *D. buzzatii* and *D. koepferae*

Species	Variation	D_T	D_{FL}	F_{FL}	R_2
<i>D. buzzatii Arg</i>	Total	-1.394	-3.740*	-3.501*	0.063*
	Nonsyn	-1.257	-3.069**	-2.945**	0.070
	Syn	-1.007	-2.549*	-2.431*	0.075
	Sil	-1.225	-3.020**	-2.889**	0.069
<i>D. buzzatii Aus</i>	Total	-0.240	-0.105	-0.187	0.102
		(-0.512)	(-1.098)	(-1.105)	(0.099)
	Nonsyn	-0.467	-0.168	-0.329	0.096
		(-0.928)	(-1.503)	(-1.617)	(0.088)
	Syn	0.477	0.609	0.672	0.132
		(0.696)	(0.554)	(0.710)	(0.151)
	Sil	0.179	0.033	0.097	0.118
		(0.346)	(-0.020)	(0.114)	(0.133)
<i>D. koepferae</i>	Total	-0.276	-1.022	-0.906	0.100
	Nonsyn	-0.399	-1.267	-1.170	0.096
	Syn	-0.029	-0.500	-0.402	0.111
	Sil	-0.055	-0.911	-0.727	0.108

* $p < 0.05$; ** $p < 0.01$.

Nonsyn, nonsynonymous variation; Syn, synonymous variation; Sil, silent variation.

In parenthesis: the values for the Australian CRS.

Table 4. McDonald and Kreitman test for the $\alpha E5$ gene region in *D. buzzatii* and *D. koepferae*

Species	Nonsyn	Syn	G	p	NI	Sil	G	p	NI
<i>D. buzzatii Arg</i>									
Polymorphic sites	14	13				15			
Fixed diff. <i>D. koepferae</i>	6	17	3.506	0.061	3.05	21	4.225	0.040	3.27
Fixed diff. <i>D. borborema</i>	10	31	5.351	0.021	3.34	32	4.567	0.033	2.99
<i>D. buzzatii Aus</i>									
Polymorphic sites	16	8				9			
Fixed diff. <i>D. koepferae</i>	6	17	8.009	0.005	5.67	21	9.577	0.002	6.22
Fixed diff. <i>D. borborema</i>	10	31	11.385	0.001	6.20	32	10.719	0.001	5.69
<i>D. koepferae</i>									
Polymorphic sites	9	20				26			
Fixed diff. <i>D. buzzatii</i>	6	17	0.154	0.695	1.28	21	0.102	0.750	1.20
Fixed diff. <i>D. borborema</i>	5	20	0.862	0.353	1.80	23	0.563	0.453	1.60

Nonsyn, nonsynonymous variation; Syn, synonymous variation; Sil, silent variation; G, value of the G-statistic; p, probability value of G; Diff, differences; NI, Neutrality index.

Significant results in bold.

Discussion

Founder Effect in the Colonization of Australia

D. buzzatii was inadvertently introduced into Australia between 1930 and 1935 during the biological control program for prickly pear cacti (*Opuntia* spp.) (Barker 1982). Results of surveys of inversion, allozyme and mtDNA restriction site polymorphisms indicate that Australian populations have been subject to a moderate founder effect (Barker et al. 1985; Knibb et al. 1987; Halliburton and Barker 1993), whereas original South American populations of *D. buzzatii* have undergone a recent population expansion (Rossi et al. 1996; Rodríguez et al. 2000; DeBrito et al. 2002; Piccinali et al. 2004).

Theoretical and experimental studies have shown that founder events can cause the loss of low-fre-

quency alleles and drastic shifts in allele frequencies in newly founded populations (Hedrick 1983; Fontdevila 1989; Rozas et al. 1990; Ridley 1996; Austerlitz et al. 2000; Pascual et al. 2001). The present survey of nucleotide variation at the $\alpha E5$ gene region indicates that these predictions are met in the case of *D. buzzatii*. Despite the extensive sharing of nucleotide polymorphisms and the similar levels of nucleotide variation in the two continents, a considerable loss of singletons has accompanied the colonization of Australia by *D. buzzatii*. In fact, the percentage of singleton variants decreased dramatically from 70% in Argentina to 24% in Australia. Moreover, we also detected important shifts in the frequency of certain variants. For example, singleton variants at positions 348, 501, 532, 574, 575, and 589 in Argentina increased to higher frequencies during colonization of Australia.

Intriguingly, the most frequent and widespread haplotype in Argentina is not present in Australia, although two closely related haplotypes (separated by two mutational steps) are present there. It may be that the genetic diversity of the region from which the Australian population originated was not represented in the Argentinian sample analyzed. However, this seems unlikely for two reasons. First, the sample analyzed included localities such as Ticucho and Termas de Río Hondo, which are close to the research stations from which substantial amounts of cactus material were introduced into Australia during the prickly pear control program (Barker 1982). Second, surveys of molecular genetic variation of mitochondrial and nuclear genes have failed to reveal significant geographic subdivision in Argentinian *D. buzzatii* (Rossi et al. 1996; Piccinali et al. 2004). The most likely explanation for the changes in the haplotypic composition between continents is that it is the result of genetic drift associated with the founder event, but some forms of natural selection cannot be ruled out (see below).

Nucleotide Variation in D. buzzatii and D. koepferae

Our survey of nucleotide variation in $\alpha E5$ in *D. buzzatii* and *D. koepferae* revealed that *D. koepferae* was more polymorphic at silent sites, whereas *D. buzzatii* was more polymorphic at nonsynonymous sites. Although stochastic variance can produce great changes in levels of nucleotide variation (see, for example, Machado et al. 2002), this result was surprising because previous studies based on the *Xdh* and *Est-A* genes showed that *D. buzzatii* was twice as polymorphic as *D. koepferae* for nonsynonymous, synonymous, and silent variation (Piccinali et al. 2004; Hasson unpublished data). These earlier results were attributed to interspecific differences in the historical population size or in the recombinational environment in which a gene is located because of the presence of different polymorphic inversions (Piccinali et al. 2004). However, in view of the departures observed in $\alpha E5$ from neutral expectations (see the MK test results), the peculiar pattern in this locus may be the outcome of some form of diversifying selection in *D. buzzatii* populations (see below).

Electrophoretic Alleles

As might be expected, several significant associations between $\alpha E5$ haplotypes and EST-2 variants were detected by starch gel electrophoresis. The correlations were not perfect because only part of the $\alpha E5$ gene has been sequenced in our study, and starch electrophoresis does not detect all the protein varia-

tions observable by more sensitive sequential cellulose acetate gel electrophoresis (Barker 1994). This latter technique has shown that EST-2^a and EST-2^b are more heterogeneous on cellulose acetate gels, at least in Australian populations (Barker 1994). In agreement with these observations, the statistical parsimony analysis showed that the haplotypes associated with these two allelic classes were those more widely distributed across the haplotype network. In contrast, most of the EST-2^c, EST-2^{c+}, and EST-2^d haplotypes from both continents were closely related, strongly suggesting that most of them were derived from one ancestral sequence. These observations may indicate that the EST-2^a and EST-2^b classes have accumulated a higher number of variants and recombination events with time and that they are ancestral to EST-2^c, EST-2^{c+}, and EST-2^d classes.

Finally, the observation of stronger associations between haplotypes and electrophoretic alleles in Australia than Argentina is probably another consequence of the population bottleneck that occurred during the colonization of Australia.

Differentiation Between Chromosomal Arrangements

Electrophoretic surveys have shown that certain variants of the EST-2 protein are in linkage disequilibrium with second-chromosome inversions (East et al. 1987; Rodríguez et al. 2001), but differentiation in the $\alpha E5$ region between chromosomal inversions within continents was not significant after Bonferroni's correction. In addition, levels of nucleotide variation within arrangements were similar for *2st* and *2j* in both Argentina and Australia, and in agreement with theoretical expectations, whereas between and within variability suggest a moderate recombination rate between arrangements (Innan and Tajima 1999).

Despite the similar levels of nucleotide variation within *2st* and *2j*, our sample of four *2jz*³ chromosomes was monomorphic for the sequenced $\alpha E5$ gene region. However, the same chromosomes have shown sequence variation in other genes (Gomez and Hasson 2003; Laayouni et al., 2003; Piccinali et al. 2004). There are two alternative explanations for the lack of variation in *2jz*³ chromosomes, both based on the proximity of $\alpha E5$ to the proximal breakpoint of inversion *2z*³. One is that this proximity makes the region more vulnerable to selective sweeps. The second, and more plausible explanation, is that *2z*³ is a relatively recent inversion (Cáceres et al. 2001; Laayouni et al. 2003), which has simply not existed long enough to acquire variation near its breakpoints by recombination. A similar picture has been reported for *2q*⁷, another young inversion in *D. buzzatii* derived from a *2j* chromosome by a single paracentric inversion. All

chromosomes sharing this arrangement were identical for the sequences of the regions encompassing the breakpoints of the inversion (Casals et al. 2003).

Population Genetic Structure

Our survey of nucleotide variation in the $\alpha E5$ gene revealed that natural populations of *D. koepferae* are genetically structured for silent sites and that variation is hierarchically structured among phylogeographic regions. A similar population genetic structure was detected in a previous study based on synonymous variation in the *Xdh* gene (Piccinali et al. 2004), indicating that this feature is not locus specific and that populations of *D. koepferae* behave as semi-isolated populations with limited gene flow. However, the $\alpha E5$ gene, as well as *Xdh* and other nuclear and mitochondrial genes (Rossi et al. 1996; Gómez and Hasson 2003; Piccinali et al. 2004), failed to show signs of population structure in *D. buzzatii*. Although the existence of population subdivision in Argentinian *D. buzzatii* cannot be excluded, especially because of the small sample size analyzed per population, our results suggest that the pattern of structuring is weaker than in *D. koepferae*. This absence of population structure in *D. buzzatii* could reflect an extensive gene flow between populations and/or recent divergence from an ancestral population (Rodríguez et al. 2000).

Departures from Neutrality in the $\alpha E5$ Gene in *D. buzzatii*

Overall, our results suggest that populations of *D. koepferae* are in drift–mutation equilibrium. However, in Argentinian *D. buzzatii*, there is evidence that nucleotide variation departs significantly from neutrality toward an excess of singletons at synonymous and nonsynonymous sites (Fu and Li tests; Table 3). As discussed elsewhere (Piccinali et al. 2004), similar departures have been reported in other nuclear loci (Gómez and Hasson 2003; Laayouni et al. 2003) and in the mitochondrial genome (Rossi et al. 1996; DeBrito et al. 2002), and the available evidence suggests that these departures are not merely a sampling artifact. One possible explanation for the genome-wide excess of singletons in Argentina may be a recent population expansion. This interpretation is also supported by the results of the R_2 test, which is one of the more powerful tests for detecting changes in population size (Ramos-Onsins and Rozas 2002).

Another pattern emerging from our data is an excess of replacement polymorphism in both native and colonized populations of *D. buzzatii*. Three alternative explanations have been offered to explain such a pattern. First, a relaxation of natural selection can lead to an increase in the amount of nonsynonymous

polymorphism not affecting interspecific divergence (Kennedy and Nachman 1998). Second, the so-called nearly neutral model contends that nonsynonymous substitutions are slightly deleterious and can persist for longer periods of time because of small selection coefficients and generally do not contribute to divergence because the fixation of these mutations is highly unlikely (Ohta 1992). The third hypothesis is Gillespie's model of episodic selection (Gillespie 1991), which proposes that particular haplotypes, carrying initially advantageous amino-acid variants that may rapidly increase in frequency but cannot reach fixation because of changes in the intensity and direction of natural selection.

The relaxation of natural selection may be a plausible explanation for $\alpha E5$ variability in *D. buzzatii*, but only in Australia, because an excess of nonsynonymous variants is an expected feature in populations that have experienced a recent population bottleneck (Fay and Wu 2002). However, the signature of a population expansion, as seems to have occurred in Argentina, is usually an excess of divergence, which can be confounded with positive selection (Eyre-Walker 2002; Fay and Wu 2002).

Rand and Kann (1996) proposed a method to test the nearly neutral model. One prediction of the nearly neutral theory is that mildly deleterious polymorphisms are not expected to reach appreciable frequencies, whereas the frequencies of neutral variants may range from rare to intermediate (Ohta 1992). Thus, according to this line of reasoning, an appropriate test of the model would be to compare the frequency spectra of silent versus replacement polymorphisms. Because silent polymorphisms may reach high frequencies and exhibit a wide frequency range, Tajima's D would be expected to be less extreme than for replacement sites. On one hand, in Argentina, nonsynonymous and synonymous sites have fairly similar average frequencies for less common variants (mean = 0.09 for both), and the variance (var) for nonsynonymous sites (var = 0.02) is twice as large as that for synonymous sites (var = 0.01). Similarly, Tajima's D , although not significant, is quite similar for both types of sites (Table 3). On the other hand, in Australia, the observed pattern is closer to the expectations of the model. In effect, the mean and variance of the frequency of synonymous variants (mean = 0.18; var = 0.03) are greater than for nonsynonymous sites (mean = 0.12, var = 0.01), and Tajima's D s are positive and negative for synonymous and nonsynonymous sites, respectively (Table 3). Therefore, the nearly neutral model seems to be another suitable explanation for the excess of nonsynonymous polymorphism at the $\alpha E5$ locus in Australia but not in Argentina.

Finally, we consider whether the excess of replacement polymorphism in the *D. buzzatii* of both

continents is in agreement with the predictions of some form of heterogeneous selective regime. EST-2 is proposed to be involved in digestion and/or detoxification (East 1982), and habitat selection (Fernández-Iriarte et al. 2002), and *D. buzzatii* is the only species of the buzzatii complex that lives associated with prickly pear cacti; the other species breed and feed in columnar cacti (Hasson et al. 1992; Manfrin and Sene 2006). Several species of *Opuntia* serve as hosts for *D. buzzatii* in both Argentina and Australia, although there is very little concordance between the species between the two continents (Kircher 1982; Hasson et al. 1992). The different *Opuntia* species also constitute chemically different habitats for *D. buzzatii* with respect to the kind and quantity of alcohols, ketones, alkaloids, and esters (East 1982; Kircher 1982). Therefore, we suggest that the observed excess of nonsynonymous polymorphism in *D. buzzatii* on both continents might be related to the shifts to heterogeneous complements of host plants that *D. buzzatii* has experienced since its divergence from the other *D. buzzatii* complex species and during the expansion of its range. In particular, given the different complements of *Opuntia* species on the two continents, such an explanation could also account for the finding of different high-frequency haplotypes in the two data sets. Although the idea clearly requires testing in further more specific ecological genetic analyses, we propose that the high level of nonsynonymous variation we found in the $\alpha E5$ gene is related, at least in part, to a role it has played in the diversification of the species into a complex cactophilic niche in the arid lands of South America and, more recently, in Australia.

Acknowledgments. R. Kliman and two anonymous reviewers are acknowledged for helpful comments on an earlier version of this manuscript, as is Eva Zinkovsky for help with the final version of Fig. 1. R. P. was supported by a CONICET doctoral fellowship and a Fundación Antorchas postdoctoral fellowship. L. J. M. was supported by an Australian Postgraduate Award. E. H. is member of Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Twenty Argentinian *D. buzzatii* and all *D. koepferae* sequences were obtained in Argentina by R. P. with UBA, CONICET, Fundación Antorchas, and ANPCyT grants to E. H. The remaining 19 Argentinian *D. buzzatii*, all of the Australian *D. buzzatii*, and the *D. borborema* sequences were obtained by L. J. M. and financed by an Australian Research Council grant to J. S. F. B.

References

- Austerlitz F, Mariette S, Machon N, Gouyon P-H, Godelle B (2000) Effects of colonization processes on genetic diversity: Differences between annual plants and tree species. *Genetics* 154:1309–1321
- Barker JSF (1982) Population genetics of *Opuntia* breeding *Drosophila* in Australia. In: Barker JSF, Starmer WT (eds) Ecological genetics and evolution. The cactus-yeast-*Drosophila* model system. Academic, Sydney, Australia, pp 209–224
- Barker JSF (1994) Sequential gel electrophoretic analysis of esterase-2 in two populations of *Drosophila buzzatii*. *Genetica* 92:165–175
- Barker JSF, East PD (1980) Evidence of selection following perturbation of allozyme frequencies in a natural population of *Drosophila*. *Nature* 284:166–168
- Barker JSF, Starmer WT (1999) Environmental effects and the genetics of oviposition preference for natural yeast substrates in *Drosophila buzzatii*. *Hereditas* 130:145–175
- Barker JSF, Toll GL, East PD, Widders PR (1981) Attraction of *Drosophila buzzatii* and *D. aldrichi* to species of yeasts isolated from their natural environment. II. Field experiments. *Aust J Biol Sci* 34:613–624
- Barker JSF, de Sene FM, East PD, Pereira MAQR (1985) Allozyme and chromosomal polymorphism of *Drosophila buzzatii* in Brazil and Argentina. *Genetica* 67:161–170
- Barker JSF, Vacek DC, East PD, Starmer WT (1986) Allozyme genotypes of *Drosophila buzzatii*: Feeding and oviposition preferences for microbial species, and habitat selection. *Aust J Biol Sci* 39:47–58
- Barker JSF, East PD, Christiansen FB (1989) Estimation of migration from a perturbation experiment in natural populations of *Drosophila buzzatii* Patterson & Wheeler. *Biol J Linn Soc* 37:311–334
- Betrán E, Quezada-Díaz JE, Ruiz A, Santos M, Fontdevila A (1995) The evolutionary history of *Drosophila buzzatii*. XXXII. Linkage disequilibrium between allozymes and chromosome inversions in two colonizing populations. *Heredity* 74:188–199
- Cáceres M, Puig M, Ruiz A (2001) Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res* 11:1353–1364
- Campbell PM, Robin GCdeQ, Court LN, Dorrian SJ, Russell RJ, Oakeshott JG (2003) Developmental expression and gene/enzyme identifications in the alpha esterase gene cluster of *Drosophila melanogaster*. *Insect Mol Biol* 12:459–471
- Casals F, Cáceres M, Ruiz A (2003) The Foldback-like transposon *Galileo* is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*. *Mol Biol Evol* 20:674–685
- Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Mol Biol Evol* 15:538–543
- Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1660
- Cyglar M, Schrag JD, Ergon F (1992) Advances in structural understanding of lipases. *Biotech Genet Eng Rev* 10:143–184
- DeBrito RA, Manfrin MH, Sene FM (2002) Mitochondrial DNA phylogeography of Brazilian populations of *Drosophila buzzatii*. *Genet Mol Biol* 25:161–171
- Durando CM, Baker RH, Etges WJ, Breed WS, Wasserman M, DeSalle R (2000) Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. *Molecular Phylogenetics and Evolution* 16:296–307
- East PD (1982) Non-specific esterases of *Drosophila buzzatii*. In: Barker JSF, Starmer WT (eds) Ecological genetics and evolution. The cactus-yeast-*Drosophila* model system. Academic, Sydney, Australia, pp 323–338
- East P, Graham A, Whittington G (1990) Molecular isolation and preliminary characterisation of a duplicated esterase locus in *Drosophila buzzatii*. In: Starmer WT, Barker JSF, MacIntyre RJ (eds) Ecological and evolutionary genetics of *Drosophila*. Plenum, New York, NY, pp 389–406
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of the molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479–491

- Eyre-Walker A (2002) Changing effective population size and the McDonald-Kreitman test. *Genetics* 162:2017–2024
- Fay JC, Wyckoff GJ, Wu CI (2002) Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* 28:1024–1026
- Fernández-Iriarte PJ, Rodríguez C, Hasson E (2002) Inversion and allozyme polymorphism show contrasting patterns of microgeographical population structure in a natural population of *Drosophila buzzatii* from Argentina. *J Evol Biol* 15:226–234
- Fontdevila A (1989) Founder effects in colonizing populations: The case of *Drosophila buzzatii*. In: Fontdevila A (ed) *Evolutionary biology of transient unstable populations*. Springer-Verlag, Berlin, Germany, pp 74–95
- Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709
- Gillespie JH (1991) *The causes of molecular evolution*. Oxford University Press, New York, NY
- Gómez GA, Hasson E (2003) Transpecific polymorphisms in an inversion linked esterase locus in *Drosophila buzzatii*. *Mol Biol Evol* 20:410–423
- Halliburton R, Barker JSF (1993) Lack of mitochondrial DNA variation in Australian *Drosophila buzzatii*. *Mol Biol Evol* 10:484–487
- Hasson E, Naviera H, Fontdevila A (1992) The breeding sites of the Argentinian species of the *Drosophila mulleri* complex (subgenus *Drosophila-repleta* group). *Revista Chilena De Historia Natural* 65:319–326
- Hedrick PW (1983) *Genetics of populations* Science. Books International, Boston, MA
- Hemila H, Koivula TT, Palva I (1994) Hormone-sensitive lipase is closely related to several bacterial proteins, and distantly related to acetylcholinesterase and lipoprotein lipase: Identification of a superfamily of esterases and lipases. *Biochem Biophys Acta* 1210:249–253
- Higgins DG, Thompson JD, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* 7:1–44
- Hudson RR (2000) A new statistic for detecting genetic differentiation. *Genetics* 155:2011–2014
- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164
- Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographical subdivision. *Mol Biol Evol* 9:138–151
- Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ (1994) Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* 136:1329–1340
- Innan H, Tajima F (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
- Innan H, Tajima F (1999) The effect of selection on the amount of nucleotide variation within and between allelic classes. *Genet Res* 73:15–28
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) *Mammalian protein metabolism*. Academic, New York, NY, pp 21–132
- Kennedy P, Nachman MW (1998) Deleterious mutations at the mitochondrial ND3 gene in South American marsh rats (*Holochilus*). *Genetics* 150:359–368
- Kircher HW (1982) Chemical composition of cacti and its relationship to Sonoran Desert *Drosophila*. In: Barker JSF, Starmer WT (eds) *Ecological genetics and evolution*. The cactus-yeast-drosophila model system. Academic, Sydney, Australia, pp 143–158
- Knibb WR, Barker JSF (1988) Polymorphic inversion and esterase loci complex on chromosome 2 of *Drosophila buzzatii*. II. Spatial variation. *Aust J Biol Sci* 41:239–246
- Knibb WR, East PD, Barker JSF (1987) Polymorphic inversion and esterase loci complex on chromosome 2 of *Drosophila buzzatii*. I. Linkage disequilibria. *Aust J Biol Sci* 40:257–269
- Laayouni H, Hasson E, Santos M, Fontdevila A (2003) The evolutionary history of *Drosophila buzzatii* XXXV. Inversion polymorphism and nucleotide variability in different regions of the second chromosome. *Mol Biol Evol* 20:931–944
- Lewontin RC, Kojima K (1960) The evolutionary dynamics of complex polymorphisms. *Evolution* 14:458–472
- Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: The case of *Drosophila pseudoobscura* and close relatives. *Mol Biol Evol* 19:472–488
- Manfrin MH, DeBrito ROA, Sene FM (2001) Systematics and evolution of the *Drosophila buzzatii* (Diptera: Drosophilidae) cluster using mtDNA. *Ann Entomol Soc Am* 94:333–346
- Manfrin MH, Sene FM (2006) Cactophilic drosophila in South America: A model for evolutionary studies. *Genetica* 126:57–75
- Mascord LJ (2004) Molecular population genetics of the *esterase-2* locus in Australian and Argentinian populations of *Drosophila buzzatii*. Doctoral thesis, University of New England, Armidale, New South Wales, Australia
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654
- Oakeshott JG, van Papenrecht EA, Boyce TM, Healy MJ, Russell RJ (1993) Evolutionary genetics of *Drosophila* esterases. *Genetica* 90:239–268
- Oakeshott JG, Claudianos C, Russell RJ, Robin GC (1999) Carboxyl/cholinesterases: A case study of the evolution of a successful multigene family. *BioEssays* 21:1031–1042
- Oakeshott JG, Claudianos C, Campbell PM, Newcomb RD, Russell RJ (2005) Biochemical genetics and genomics of insect esterases. In: Gilbert LI, Iatrou K, Gill SS (eds) *Comprehensive molecular insect science-pharmacology*. Series 2, vol. 5. Elsevier, Oxford, UK, pp 309–381
- Ohta T (1992) The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst* 23:263–286
- Pascual M, Aquadro CF, Soto V, Serra L (2001) Microsatellite variation in colonizing and palearctic populations of *Drosophila subobscura*. *Mol Biol Evol* 18:731–740
- Piccinali R, Aguadé M, Hasson E (2004) Comparative molecular population genetics of the *Xdh* locus in the cactophilic sibling species *Drosophila buzzatii* and *D. koepferae*. *Mol Biol Evol* 21:141–152
- Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19:2092–2100
- Rand DM, Kann LM (1996) Excess amino acid polymorphism in mitochondrial DNA: Contrasts among genes from *Drosophila*, mice, and humans. *Mol Biol Evol* 13:735–748
- Ranz JM, González J, Casals F, Ruiz A (2003) Low occurrence of gene transposition events during the evolution of the genus *Drosophila*. *Evolution* 57:1325–1335
- Ridley M (1996) *Evolution*, 2nd ed. Blackwell Science, Cambridge, MA
- Robin GCDeQ, Medveczky KM, Russell RJ, Oakeshott JG (1996) Duplication and divergence of the genes of the alpha-esterase cluster of *Drosophila melanogaster*. *J Mol Evol* 43:241–252
- Robin GCDeQ, Claudianos C, Russell RJ, Oakeshott JG (2000) Reconstructing the diversification of α -esterases: Comparing the gene clusters of *Drosophila buzzatii* and *D. melanogaster*. *J Mol Evol* 51:149–160

- Rodríguez C, Piccinali R, Levy E, Hasson E (2000) Contrasting population genetic structures using allozymes and the inversion polymorphism in *Drosophila buzzatii*. *J Evol Biol* 13:976–984
- Rodríguez C, Piccinali R, Levy E, Hasson E (2001) Gametic associations between inversion and allozyme polymorphisms in *Drosophila buzzatii*. *J Heredity* 92:382–391
- Rodríguez-Trelles F, Alarcón L, Fontdevila A (2000) Molecular evolution and phylogeny of the *buzzatii* complex (*Drosophila repleta* group): a maximum-likelihood approach. *Mol Biol Evol* 17:1112–1122
- Rossi MS, Barrio E, Latorre A, Quezada-Díaz JE, Hasson E, Moya A, Fontdevila A (1996) The evolutionary history of *Drosophila buzzatii*. XXX. Mitochondrial DNA polymorphism in original and colonized populations. *Mol Biol Evol* 13:314–323
- Rozas J, Rozas R (1999) DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175
- Rozas J, Hernández M, Cabrera VM, Prevosti A (1990) Colonization of America by *Drosophila subobscura*: Effect of the founder event on the mitochondrial DNA polymorphism. *Mol Biol Evol* 7:103–109
- Ruiz A, Wasserman M (1993) Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. *Heredity* 70:582–596
- Russo CAM, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of Drosophilid species. *Mol Biol Evol* 12:391–404
- Schneider S, Roessli D, Excoffier L (2000) Arlequin ver 2.000: A software for population data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland. Available at: <http://anthro.unige.ch/arlequin>. Accessed: February 21, 2006
- Sokal RR, Oden NL, Barker JSF (1987) Spatial structure in *Drosophila buzzatii* populations: Simple and directional spatial autocorrelation. *Am Nat* 129:122–142
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619–633
- Watt AW (1981) The genetics of temperature tolerance in *Drosophila*. In: Gibson JB, Oakeshott JG (eds) Genetic studies of drosophila populations. Australian National University, Canberra, Australia, pp 139–146
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276