J Mol Evol (2002) 54:569–586
DOI: 10.1007/s00239-001-0044-7

JOURNAL OF MOLECULAR EVOLUTION

Rapid Concerted Evolution via Gene Conversion at the *Drosophila hsp70* Genes

Brian R. Bettencourt,[1] Martin E. Feder[1,2]

[1] Department of Organismal Biology and Anatomy, University of Chicago, 1027 East 57th Street, Chicago, IL 60637, USA
[2] Committee on Evolutionary Biology, University of Chicago, 1027 East 57th Street, Chicago, IL 60637, USA

Received: 25 June 2001 / Accepted: 10 October 2001

Abstract. We analyzed nucleotide variation in the *hsp70* genes of *Drosophila melanogaster* (five genes) and *D. simulans* (four genes) to characterize the homogenizing and diversifying roles of gene conversion in their evolution. Gene conversion within and between the 87A7 and 87C1 gene clusters homogenize the *hsp70* coding regions; in both *D. melanogaster* and *D. simulans,* same-cluster paralogues are virtually identical, and large intercluster conversion tracts diminish 87A7/87C1 divergence. Same-cluster paralogues share many polymorphisms, consistent with frequent intracluster conversion. Shared polymorphism is highly biased toward silent variation; homogenizing conversion interacts with purifying selection. In contrast to the coding regions, some *hsp70* flanking regions show conversion-mediated diversification. Strong reductions of nucleotide variability and linkage disequilibria among conversion-mediated sites in *hsp70Ab* and *hsp70Bb* alleles sampled from a single natural population are consistent with a selective sweep. Comparison of the *D. melanogaster* and *D. simulans hsp70* genes reveals whole-family fixed differences, consistent with rapid propagation of novel mutations among duplicate genes. These results suggest that the homogenizing and diversifying roles of conversion interact to drive dynamic concerted evolution of the *hsp70* genes.

Correspondence to: Brian R. Bettencourt, Biology Department, Pennsylvania State University, 208 Mueller Lab, University Park, PA 16802, USA; *email:* brb11@psu.edu

Key words: *hsp70* — *Drosophila* — Gene conversion — Concerted evolution

Introduction

Gene conversion plays contrasting roles in the evolution of multigene families; it can transmit novel sequence among genes but also constrain divergence (Ohta 1983, 1989). The first of these roles is evident during rapid evolution (and divergence) under strong selection for diversified roles. Examples include genes involved in olfaction (Sharon et al. 1999), immune response [Attacins (Lazzaro and Clark 2001), HLA (Takahata and Satta 1998), MHC (Edwards et al. 1998)], sex or reproductive isolation [*cdic/sdic* (Nurminsky et al. 1998), mating type loci (Haber 1998)], multiallelic systems [opsins (Shyue et al. 1995; Zhao et al. 1998)], and tissue- or time-specific expression [amylases (Inomata et al. 1995; Popadic et al. 1996)]. An important theme emerges from these studies: conversion can actually facilitate diversification of duplicated genes by protecting coding sequences (CDSs) from mutation and eventual pseudogene status (Walsh 1987) while allowing evolution of regulatory regions toward novel expression patterns. In contrast, conversion can also act to conserve the sequence of highly functionally constrained gene families [e.g., rDNA genes (Schlotterer and Tautz 1994; Polanco et al. 1998, 2000) and histones (Liu et al. 1987; reviewed by Ohta 1990)]. Thus, gene conversion may limit the independent evolution of gene copies toward new roles (Walsh 1987).

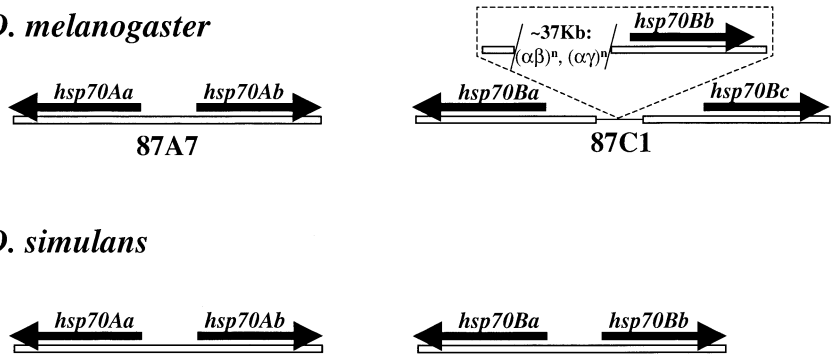


Fig. 1. Nomenclature and organization of *hsp70* genes in *D. melanogaster* and *D. simulans.* The 87C1 gene cluster of *D. melanogaster* is disrupted by the αβ and αγ repeats (Lis et al. 1978).

These studies suggest that a balance of the two opposing effects of conversion, homogenization and diversification, determines the degree of divergence among gene family members. Homogenization may win out in ancient, constrained gene families whose function is well entrenched and thus present little sequence variation for conversion to distribute, while diversification may win out in newly duplicated or rapidly diverging genes subject to strong selection. Here we report a unique instance in which gene conversion promotes both homogenization and rapid change among copies of a truly ancient gene, *hsp70.*

hsp70 is an ancient gene whose sequence and function have changed relatively little throughout billions of years. Indeed, *hsp70* ranks among the most highly conserved gene/protein systems (Gupta and Golding 1996; Karlin and Brocchieri 1998). Hsp70 is a molecular chaperone that prevents formation of cytotoxic aggregates of stress-denatured proteins (Agashe and Hartl 2000; Mayer et al. 2000). In *Drosophila melanogaster,* Hsp70 expression is critical for induced tolerance of natural heat stress (Feder et al. 1997). Five virtually identical genes (at two chromosomal loci, henceforth "gene clusters": 87A7–*hsp70Aa, hsp70Ab,* 87C1–*hsp70Ba, hsp70Bb, hsp70Bc*) encode Hsp70 in *D. melanogaster* (Fig. 1). All five genes are expressed massively after heat and other stresses (Simon et al. 1985; Sharma and Lakhotia 1995) and are otherwise transcriptionally inactive (Feder et al. 1992). Clearly, because these genes are ancient with a critically important phenotype, they ought to be subject to the homogenizing effect of conversion. Indeed, surveys of restriction-site variation among the five *hsp70* genes revealed very low paralogous divergence, suggesting frequent gene conversion (Leigh Brown and Ish-Horowicz 1981).

However, several studies hint that the *hsp70* genes vary and, thus, perhaps are subject to conversion's diversifying role. Natural populations of *D. melanogaster* display correlated variation in Hsp70 expression and thermotolerance (Krebs and Feder 1997), experimental evolution at varied temperatures effectively alters both Hsp70 expression and thermotolerance (Bettencourt et al. 1999), and different *hsp70* gene arrangements segregate in natural populations (Leigh Brown and Ish-Horowicz 1981)—all strong hints that *hsp70* is a target of natural selection. Furthermore, the *hsp70* genes proliferated rapidly via duplication of an ancestral two-*hsp70* gene cassette in the *melanogaster* species subgroup and subsequent tandem duplication of a single gene in *melanogaster* alone (Leigh Brown and Ish-Horowicz 1981; Konstantopoulou et al. 1998; Bettencourt and Feder 2001). This suggests historical positive selection as the species evolved a cosmopolitan distribution and expanded thermal niche (see Feder and Krebs 1998; Krebs and Bettencourt 1999).

Thus, the opposing forces of constraint and diversification are likely acting on the *hsp70* genes. The two forces could act together, however: homogenization via conversion could spread favorable mutations among the *hsp70* genes and thus respond to the tugs of both constraint and selection. That gene conversion could rapidly propagate mutations through a gene family, fueling change via concerted evolution, has long been suspected (Ohta and Dover 1984; Basten and Ohta 1992). To determine whether and how this process affects the *hsp70* genes, we examined the sequence and organization of multiple *hsp70* genes/alleles in *D. melanogaster* and a related species.

First, we characterize nucleotide polymorphism in a broad sample of alleles of all five *hsp70* genes to assess inter- and intracluster divergence and the relative homogenizing/diversifying roles of inter- and intracluster gene conversion. Next, we examine the nature of both past and present concerted evolution at the *hsp70* genes, first, by a population genetic analysis of *hsp70* alleles sampled from a single wild population and, second, by a comparison of the tempo and mode of change among and between the *hsp70* genes of *D. melanogaster* and *D. simulans*.

Materials and Methods

Fly Stocks, Genomic Clones, and Sampling. We sampled *hsp70* variation both broadly, from diverse *Drosophila* strains, and more deeply, from a single natural population. Four *D. melanogaster* isofemale strains were donated by C.-I. Wu (University of Chicago): Z(H)1 (Zimbabwe), ZZ30 (Zimbabwe), FrV_3-1 (France), and QD18 [Japan (JPN)]. Strains A25, B25, A28, and B28 are Oregon R-derived lines reared at

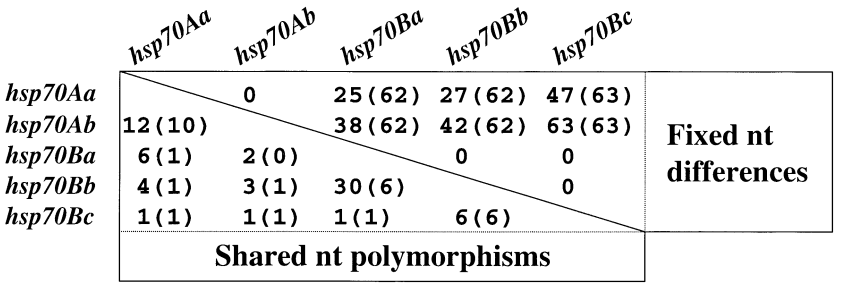


Fig. 2. Numbers of fixed nucleotide differences (*above diagonal*) and shared SNPs (*below diagonal*) between coding sequences of *D. melanogaster hsp70* genes. Numbers in parentheses reflect the exclusion of alleles bearing intercluster conversion tracts [$hsp70Aa^{122}$, $hsp70Aa^{PA47}$, $hsp70Ba^{A28}$, $hsp70Ba^{Z(H)1}$, $hsp70Bb^{B28}$, $hsp70Bb^{Z(H)1}$].

25 and 28°C for more than 20 years (Bettencourt et al. 1999). The AUS population was founded in the Wu laboratory by pooling 20 isofemale lines from Yepoon, Australia (collected by A. Hoffmann in 1994). Eight *D. melanogaster* third chromosome-isogenic strains derived from a single natural population in Pennsylvania, USA [3CPA2, 3CPA35, 3CPA43, 3CPA47, 3CPA61, 3CPA81, 3CPA86, and 3CPA126 (PA2, PA35, etc.)] were donated by A.G. Clark (Pennsylvania State University). *Drosophila simulans* isofemale strain DSR (Riverside, CA) was provided by T. Karr (University of Chicago). The *D. melanogaster* genomic clones 122 and 56H8, both derived from the 87A7 *hsp70* gene cluster of Oregon R genomes (Moran et al. 1979; Goldschmidt-Clermont 1980), were provided by M. Goldschmidt-Clermont and J.T. Lis, respectively.

DNA extraction. For each species/strain, genomic DNA was extracted from 75–80 adults by a standard phenol/chloroform method and stored at −20°C. These samples served as PCR templates for amplification products to be cloned (with the exception of the coisogenic PA strains, whose amplification products were sequenced directly). Additionally, single-fly DNA preparations were prepared according to Gloor et al. (1993) and stored at −20°C. These preparations served as PCR templates for amplification products to be sequenced directly.

Nomenclature. Figure 1 diagrams the arrangement and nomenclature of the *D. melanogaster* and *D. simulans hsp70* genes. Only the *melanogaster* genes (*hsp70Aa, -Ab, -Ba, -Bb, -Bc*) have names in the literature. We assigned names to the four *D. simulans* genes (*Dsimhsp70Aa, -Ab, -Ba, -Bb*) based on their cytological homology with the *D. melanogaster* loci (see Leigh Brown and Ish-Horowicz 1981).

Amplification of D. melanogaster hsp70 *Genes.* At the onset of this project, GenBank contained few *D. melanogaster hsp70* sequences. *hsp70Ab* and *hsp70Bb* were completely sequenced (accession Nos. J01103 and J01104). For the remaining three genes (*hsp70Aa, hsp70Ba, hsp70Bc*), only noncoding sequences were available (accession Nos. K01292, J01103, K01293, K01294, J01104, K01295). Using these sequences, whose similarity decreases with the distance from coding regions, we designed specific primers for each *D. melanogaster hsp70* gene (*hsp70Aa, -Ab, -Ba, -Bb, -Bc).* Primers and amplification conditions for each gene can be found at http://pondside.uchicago.edu/~feder/JMETableA.html. Sequences are located in GenBank under accession Nos. AF295933–AF295957 and AF350452–AF350491. For the JPN, B28, and PA47 strains, use of an alternative upper primer was required to amplify *hsp70Ab* successfully (see http://pondside.uchicago.edu/~feder/JMETableA.html).

Amplification of D. simulans hsp70 *Genes.* We used the annotated sequence of *D. melanogaster* chromosome arm 3R, provided by Celera Inc., and the GeneScene Java applet (http://www.fruitfly.org/annot/geneseen-launch-static.html) to amplify the *hsp70* genes of *D. simulans*. Using both the genomic sequence and predicted transcripts, we designed primers in predicted genes nearest to the 87A7 and 87C1 *hsp70* clusters to be used in PCR amplification together with a primer that begins on the first codon of the conserved *hsp70* CDS. These primers were used previously to amplify the *D. orena* and *mauritiana hsp70* genes (Bettencourt and Feder 2001). Sequences are located in GenBank under accession Nos. AF295963–AF295978. Primers and amplification conditions can be found at http://pondside.uchicago.edu/~feder/JMETableB.html.

Preparation and Cloning of Products. PCR products were cleaned and/or gel-purified using Qiagen spin columns and suspended in water. Products amplified from single-fly templates, or from bulk DNA of the coisogenic PA strains, were directly sequenced. Alternatively, products amplified from multiple-fly templates (see above) were cloned into the pGEM-T or pGEM-T Easy vector (Promega) according to the manufacturer's instructions. Individual plasmid clones were prepared with Qiagen spin columns, suspended in water, and sequenced.

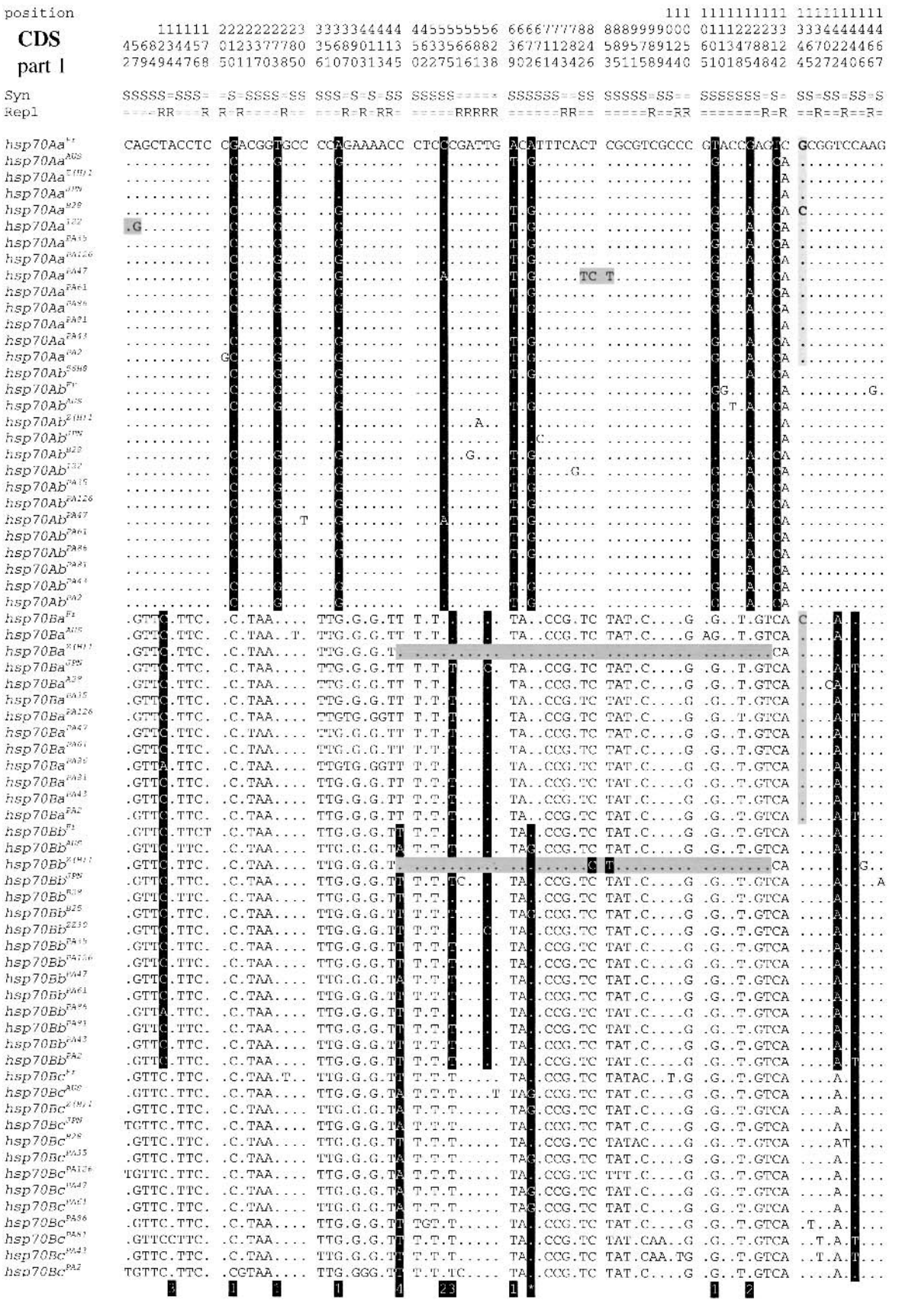
DNA Sequencing and Analysis. All genes/alleles were sequenced using ABI Prism cycle sequencing kits (Perkin Elmer) according to the manufacturer's instructions. Automated sequencing was conducted on ABI 377 sequencers located at the Cancer Research Center DNA Sequencing Facility and the Genome Sequencing Center (University of Chicago).

Internal sequencing primers (available upon request) were targeted to sites conserved in *hsp70Ab* and *hsp70Bb* (accession Nos. J01103 and J01104) to provide complete double-stranded coverage. Primers based on vector sequence (M13 forward/reverse, T7, SP6) were also used when applicable.

Assembly, Alignment, and Analysis of Sequences. Sequences were assembled manually and aligned using CLUSTAL X (Jeanmougin et al. 1998). Only regions of double-stranded coverage were analyzed. Alleles were named according to their source [i.e., $hsp70Aa^{122}$ from the 122 clone and $hsp70Bb^{Z(H)1}$ from the Z(H)1 strain]. All singleton polymorphisms found on products amplified without a proofreader enzyme were verified or rejected as PCR error by sequencing additional clones or PCR products. Software packages utilized to analyze the aligned sequences included DAMBE (X. Xia), DNASP (J. Rozas), PHYLIP (J. Felsenstein), and SITES (J. Hey). All analyses were restricted to aligned regions, and gaps were excluded.

Results

The hsp70 *Coding Sequences of* D. melanogaster *Are Nearly Identical and Evolve in Concert.* The 87A7 and 87C1 *hsp70* clusters have diverged little, and paralogues within either cluster lack fixed differences. Since the 87C1 cluster arose by duplication of 87A7 approximately 12–15 mya (Konstantopoulou et al. 1998; Bettencourt and Feder 2001), the 87C1 *hsp70* genes apparently diverged from those at 87A7 by 36.3 ± 15.5 (mean ± SE)


position
CDS
part 1
Syn
Repl

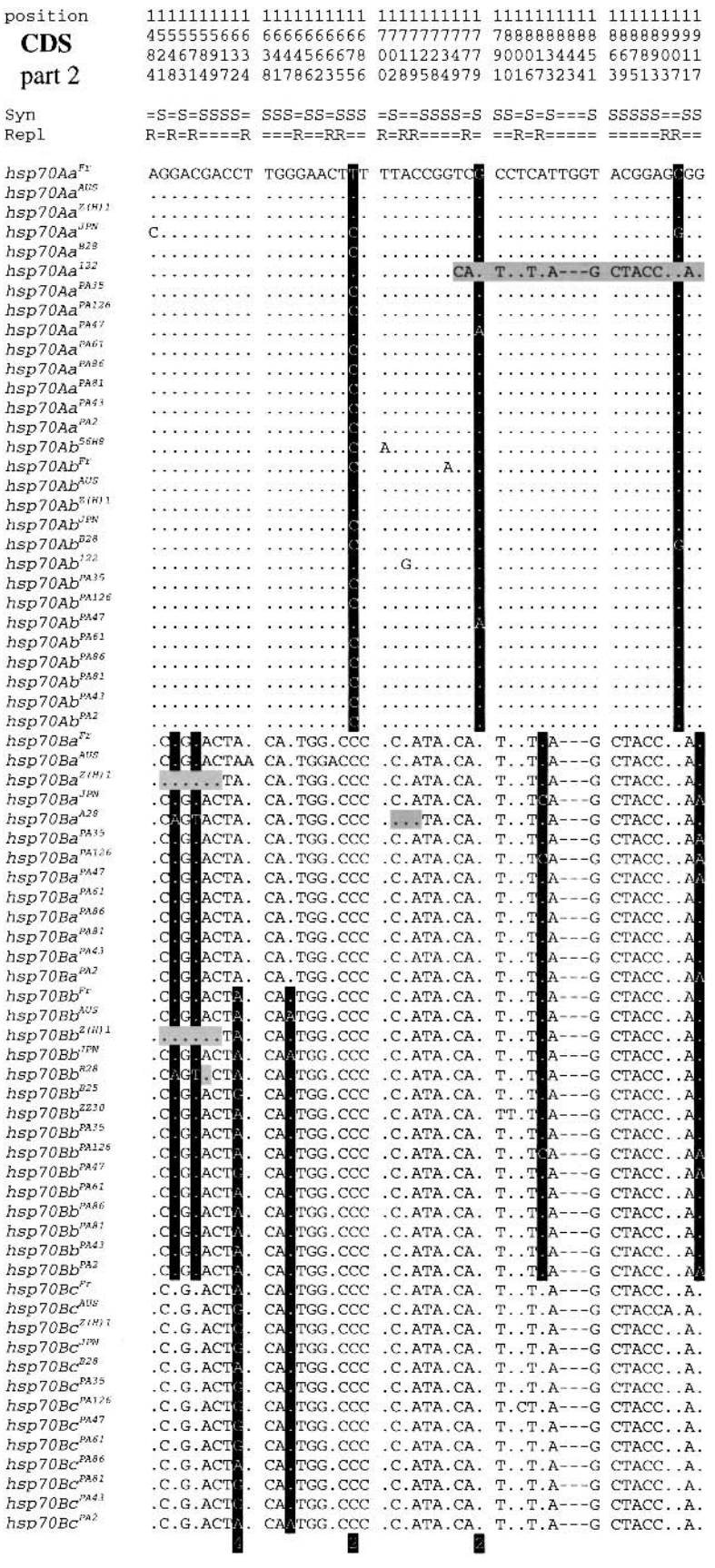


Fig. 3. Nucleotide polymorphism and divergence in the *hsp70* coding sequences of *D. melanogaster*. Nucleotide numbering (position) begins at the first base of the start codon. Synonymous and replacement substitutions are labeled S and R, respectively. Nucleotides identical to the first sequence are indicated by *dots;* deletions, by *dashes. Gray bars* indicate intercluster conversion tracts (*dark* if identified by DNASP, *light* if by eye). *Black columns* mark polymorphisms involved in intracluster conversion; linkage groups of these sites are indicated by *numbers in black boxes* below the sequences.

synonymous and 6.3 ± 2.5 nonsynonymous substitutions in their 1926- to 1929-bp CDSs, or by only 2.2% (Fig. 2). The apparent rates (k_S and k_{NS}) at which these differences arose, especially at nonsynonymous sites, are very low: $k_S = 2.94 \pm 1.26$ and $k_{NS} = 0.16 \pm 0.0063$ substitution/site/10^9 years. In comparison, the average k_S for *Drosophila* is 15.60 ± 5.50/site/10^9 years and the lowest k_{NS} previously reported for *Drosophila* is 0.23 ± 0.07/site/10^9 years [for arrestin B (Li 1997)]. Paralogues within either the 87A7 or the 87C1 cluster are still more homogeneous. The *hsp70Aa/Ab* gene pair at 87A7 is at least 500 million years old (Benedict et al. 1993); however, no nucleotide differences are fixed between the CDSs of *hsp70Aa* and those of *hsp70Ab* (Fig. 2). Similarly, no differences are fixed among the three *hsp70* CDSs at 87C1 (Fig. 2). Clearly, these five genes are not diverging from one another and have undergone intensive homogenization.

Inter- and Intracluster Gene Conversion Can Have Contrasting Effects on Coding and Flanking Regions

Intercluster Conversion Reduces 87A7/87C1 Coding Sequence Divergence. In the *hsp70* CDSs, DNASP identifies six intercluster conversion tracts (horizontal dark-gray bars in Fig. 3). These conversion tracts eliminate 40 87A7/87C1 differences (39 substitutions and a 3-bp indel). Visual inspection suggests three additional intercluster conversion tracts (horizontal light-gray bars in Fig. 3), which eliminate a few otherwise fixed 87A7/87C1 differences (3, 3, and 1, respectively). This homogenization confounds estimates of substitution rates based on 87A7/87C1 divergence (see above). The pattern of concerted evolution and slowed divergence is robust, however: exclusion of the six sequences containing intercluster conversion tracts increases k_S and k_{NS} only slightly (to 4.32 ± 0.0041 and 0.23 ± 0/site/10^9 years, respectively). Homogenization of 87A7/87C1 divergence by intercluster conversion also extends to the *hsp70* flanking regions (see below) (Figs. 3 and 4).

Intracluster Conversion Explains Shared Polymorphisms. hsp70Aa and *hsp70Ab* share 16 polymorphisms (4 flanking, 12 coding), for which intracluster gene conversion is the most likely explanation. The probability that these polymorphisms arose by parallel mutation is low ($p < 0.000001$; see Table 1). For details on the calculation of this probability, which is based upon the hypergeometrical distribution, see Rozas and Aguade (1994), Rozas et al. (1999), and Bettencourt (2001). We can also reject parallel mutation for shared polymorphisms in all three pairwise comparisons of 87C1 paralogues (Table 1). Alternatively, polymorphisms shared among clusters could have arisen from unequal crossing-over. This, however, would also affect the *hsp70* copy number, which does not occur in nature (Krebs and Bettencourt 1999). We identify putative intracluster conversion tracts by analysis of linkage disequilibrium at shared sites (see below).

Conversion Homogenizes and Diversifies 5′ and 3′ Flanking Regions. Both in the entire sample and within the PA population, inter- and intracluster conversions contribute to patterns of flanking sequence variation that differ between 87A7 and 87C1: diversification in some cases, homogenization in others. For example, nucleotide variability is sometimes much higher in the 87A7 flanking regions than in the CDSs as a consequence of conversion-mediated shared polymorphisms (Table 2, Fig. 4). In the PA sample, the *hsp70Ab* 5′-UTR displays a 40-fold increase in π relative to the CDS and a significantly negative Tajima's *D* value, due largely to the intercluster-converted $hsp70^{PA47}$ allele.

The 87A7 3′-UTRs, while also diversified, display a pattern of variation slightly different from that of the 5′-UTRs (Fig. 4). Intercluster conversion tracts create high variation among allelic classes but homogeneity within each class. The >240-bp intercluster conversion tract results in 46% divergence between the three converted and the three nonconverted *hsp70Aa* alleles in the worldwide sample. Otherwise, the 3′ flanks are homogeneous, without the shared single nucleotide polymorphisms (SNPs) characteristic of the 5′ flanks. The intercluster-converted region varies at only one site. Similarly, the 3′ flanks of the remaining *hsp70Aa* alleles and all the *hsp70Ab* alleles are virtually identical for 198 bp (Fig. 4). The greater within-class homogeneity of the 3′ flanking sequences in this region is consistent with more stringent constraint, higher-frequency conversion, or both. Three of the *hsp70Aa* PA alleles share a 13-bp region that is divergent at seven sites. This strongly affects *D,* which is now positive (0.89181). In this case, escaping conversion causes the strong change in the site frequency spectrum: the other five alleles are homosequential in this region and diverge from *hsp70Ab* at only two sites.

In contrast to 87A7, the 5′-UTRs of all three 87C1 *hsp70* genes are virtually identical, displaying the high degree of homogeneity and shared polymorphism consistent with intracluster conversion (Fig. 5). All three 87C1 *hsp70* genes share a 7-bp indel and linked SNPs (black bars in Fig. 5). In the PA sample, only $hsp70Ba^{PA126}$ bears the indel and a linked SNP, while all others differ only at one singleton SNP; predictably, *D* is negative (−1.44751; n.s.). Otherwise, both the 5′- and the 3′-UTRs of all three 87C1 *hsp70* genes are highly homogenized in the PA sample.

The levels of both homogeneity and shared polymorphism in the above cases suggest frequent intracluster conversion that extends beyond the CDSs well into both the 5′ and the 3′ flanking sequences. Both the nature and the frequency of indel polymorphisms in these flanking sequences are indicative of conversion event boundaries (see Discussion).

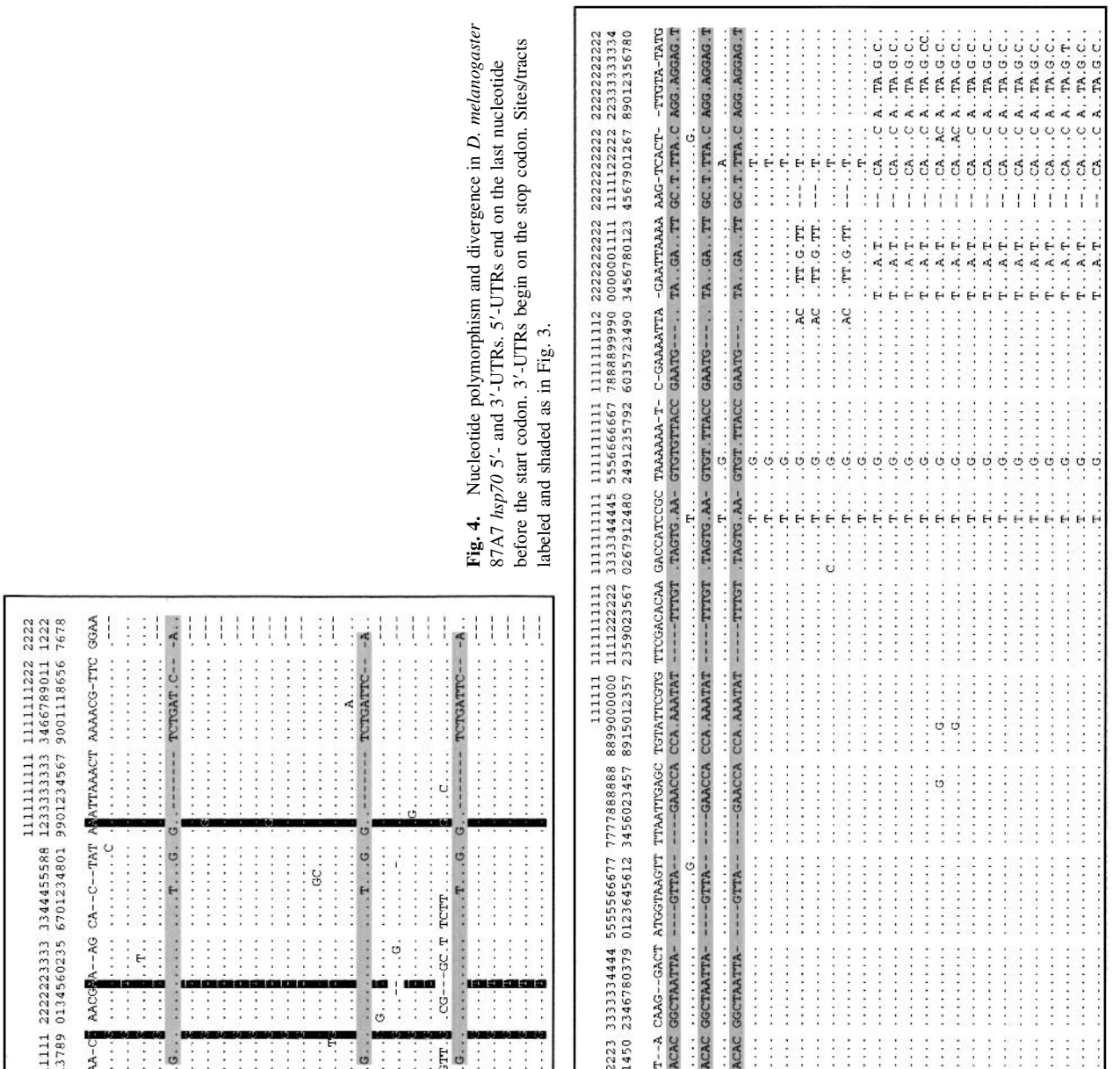


Fig. 4. Nucleotide polymorphism and divergence in *D. melanogaster* 87A7 *hsp70 5′*- and *3′*-UTRs. *5′*-UTRs end on the last nucleotide before the start codon. *3′*-UTRs begin on the stop codon. Sites/tracts labeled and shaded as in Fig. 3.

The Interaction of Selection and Conversion Shapes hsp70 *Molecular Evolution*

Hypotheses Regarding Intracluster Conversion and Selection. In conversion between the virtually identical paralogues at either cluster (as they exist today), tracts should be evident only from shared polymorphic sites, as fixed differences are completely absent and gene-specific polymorphism low. Furthermore, as conversion rate is negatively correlated with heterozygosity (Stephan and Langley 1992; Nassif and Engels 1993), conversion should be frequent between *hsp70* paralogues. Finally, as most shared SNPs are silent (e.g., at 87A7, 15 shared silent SNPs, 1 shared replacement SNP available for transfer; see Table 3), recombination or conversion events should be neutral and drift should govern the change in frequency of converted vs nonconverted alleles. Thus in a population at equilibrium intracluster conversion should both accelerate the elimination of deleterious mutations and maintain neutral shared polymorphisms [an extension of the well-documented Hill–Robertson effect of recombination rate on selection

Table 1. Intracluster shared single-nucleotide polymorphisms[a]

	hsp70Aa and *hsp70Ab*	*hsp70Ba* and *hsp70Bb*	*hsp70Ba* and *hsp70Bc*	*hsp70Bb* and *hsp70Bc*
N	1836.15	2087.46	2087.46	2091.2
η_S	78	72	72	108
Shared SNPs	16	11	2	6
p value[b]	<0.000001	<0.000001	0.016415	0.000048

[a] *N,* total number of silent (noncoding) and synonymous (coding) sites. η_S, total number of silent (noncoding) and synonymous (coding) mutations. Alleles bearing intercluster conversion tracts excluded.
[b] Based on the hypergeometric distribution.

Table 2. Nucleotide diversity in Pennsylvanian *D. melanogaster hsp70* genes[a]

Gene	Region	*n*	*S*	η	π	π_S	π_{NS}	θ	*D*
hsp70Aa	5′	232	5	5	0.00539			0.00831	−1.59524
	CDS	1926	15	15	0.00204	0.00800	0.00017	0.00300	−1.64202
	3′	225	10	11	0.02222			0.01886	0.89181
hsp70Ab	5′	225	41	43	0.04984			0.07371	−1.73567[b]
	CDS	1926	11	11	0.00143	0.00598	0.00000	0.00220	−1.75686[b]
	3′	235	1	1	0.00108			0.00167	−1.05482
hsp70Ba	5′	207	3	3	0.00362			0.00559	−1.44751
	CDS	1923	7	7	0.00136	0.00274	0.00093	0.00140	−0.16319
	3′	385	3	3	0.00195			0.00301	−1.44751
hsp70Bb	5′	214	0	0	0.00000			0.00000	n/a
	CDS	1923	7	7	0.00106	0.00337	0.00034	0.00140	−1.12898
	3′	386	3	3	0.00241			0.00300	−0.81245
hsp70Bc	5′	215	1	1	0.00117			0.00180	−1.05482
	CDS	1923	18	18	0.00321	0.00838	0.00161	0.00361	−0.56930
	3′	64	0	0	0.00000			0.00000	n/a

[a] *n,* number of sites; *S,* number of segregating sites; η, total number of mutations; π, nucleotide diversity; π_S, silent nucleotide diversity; π_{NS}, nonsilent (replacement) nucleotide diversity; θ, nucleotide diversity; *D,* Tajima's *D* statistic. CDS, coding sequence.
[b] Significant at $p < 0.05$.

efficacy (see Begun and Aquadro 1992; Aguade and Langley 1994; Aquadro et al. 1994)]. Importantly, in an autosomal region of normal recombination rate, linkage between shared polymorphic sites should be low. The data are consistent with some, but not all, of these expectations.

Linkage Disequilibrium is High Among High-Frequency Shared SNPs at 87A7. At both *hsp70* clusters, paralogues possess no fixed differences, gene-specific (unique) polymorphisms are rare, and most haplotypes are shared; in essence, paralogues behave as a single gene. Accordingly, to examine linkage at shared SNPs, paralogous sequences were paired (*hsp70Aa/Ab, hsp70Ba/Bb, hsp70Bb/Bc*) and the analysis restricted to shared SNPs that display statistically significant linkage. At 87A7, *hsp70Aa* and *hsp70Ab* share 12 SNPs in their CDSs (black bars in Fig. 3). In contrast to expectation, significant linkage disequilibrium exists in the Pennsylvanian population at several of these shared sites (Fig. 6). Two large linkage groups can explain all the observed disequilibrium, suggesting the existence of two conversion tracts (numbers 1 and 2 in Fig. 3). In group 1, linkage disequilibrium is complete and one haplotype is preserved at both *hsp70Aa* and *hsp70Ab* on seven of eight chromosomes. Recombination has slightly disrupted group two in four alleles, but again, one haplotype is at high frequency (six of eight chromosomes). As expected, intracluster conversion is maintaining shared polymorphism. Disequilibrium is high and particular haplotypes occur at a high frequency, however, two results which are inconsistent with the neutral expectation of shared and unique silent SNPs (above).

In contrast to 87A7, linkage is weak between shared sites at 87C1. *hsp70Ba* and *hsp70Bb* share eight SNPs in their CDSs, but only one linkage group is recovered (number 3 in Fig. 3). *hsp70Bb* and *hsp70Bc* share four SNPs in their CDSs, three of which are contained in the single recovered linkage group (number 4 in Fig. 3). Unlike at 87A7, numerous shared-site recombinant alleles of all three 87C1 *hsp70* genes segregate in the PA sample (Fig. 3). Reduced levels of conversion at 87C1 could explain the low level of linkage at shared sites in 87C1, compared to 87A7 (see below).

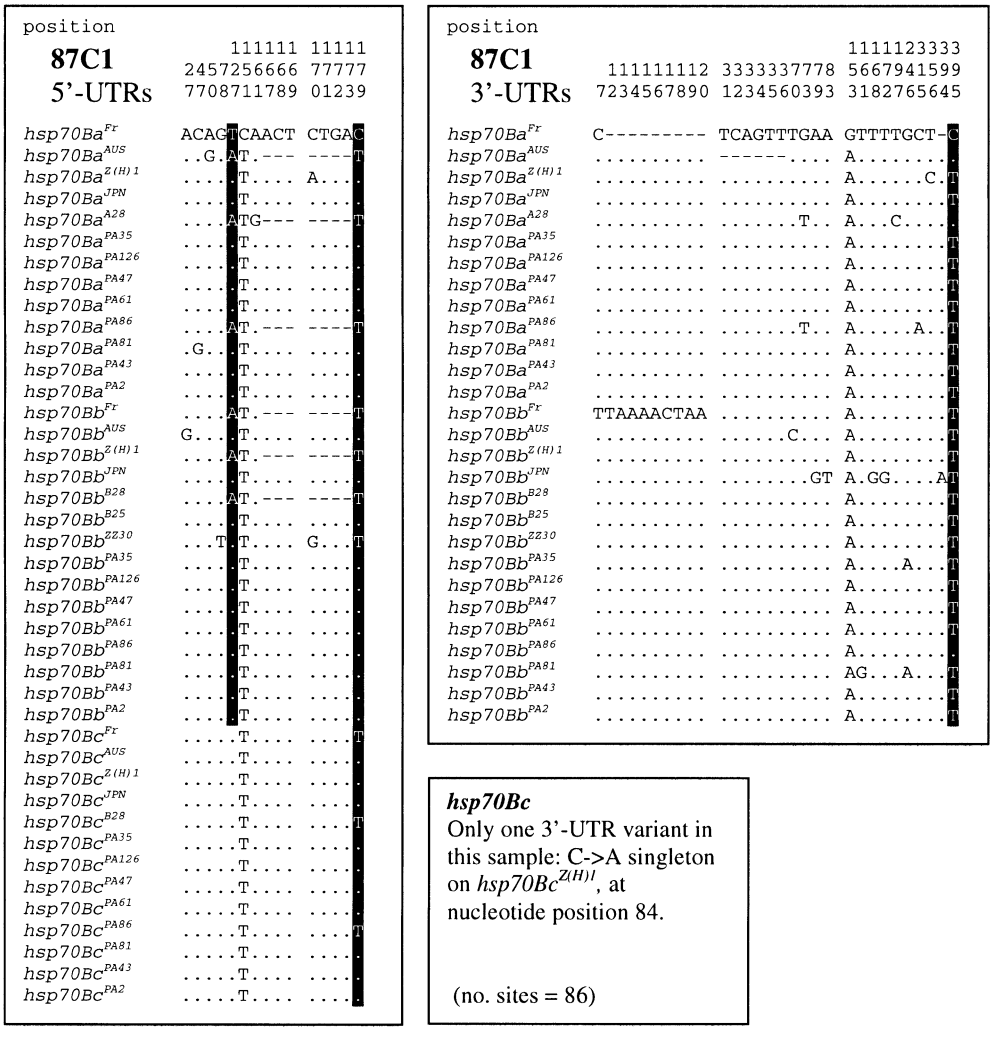


Fig. 5. Nucleotide polymorphism and divergence in *D. melanogaster* 87C1 *hsp70* 5′- and 3′-UTRs. 5′-UTRs end on the last nucleotide before the start codon. 3′-UTRs begin on the stop codon. Sites/tracts labeled and shaded as in Fig. 3.

Table 3. Constraint on shared polymorphisms and fixed differences[a]

	Unique polymorphisms		Conversion-mediated polymorphisms				Fixed differences	
			Intercluster		Intracluster			
Gene	Syn	Repl	Syn	Repl	Syn	Repl	Syn	Repl
hsp70Aa	1	2	*15*	*1*	*11*	*1*	*53*	*9*
hsp70Ab	4	7	0	0	**11**	**1**	**53**	**9**
hsp70Ba	3	4	**23**	**4**	5	4	**53**	**9**
hsp70Bb	3	2	21	5	8	4	*53*	*9*
hsp70Bc	8	6	0	0	4	0	*54*	*9*

[a] Syn, synonymous mutations; repl, replacement mutations. The typeface indicates the significance level of *G* tests of independence: italics, $p < 0.05$; boldface, $p < 0.001$. A *G* test could not be performed on *hsp70Bc* intracluster shared polymorphisms.

Reduction of Nucleotide Variation at Non-Conversion-Mediated Sites in the Pennsylvania Population Suggests Selection on hsp70Ab *and* hsp70Bb. A strong reduction in nucleotide variability in regions of normal recombination and rarity of variant haplotypes can be caused by background selection or hitchhiking under positive selection (Andolfatto and Przeworski 2000; Begun and Whitley 2000). To determine whether either process affects the *hsp70* genes, we examined the pattern and frequency of *hsp70* nucleotide variation

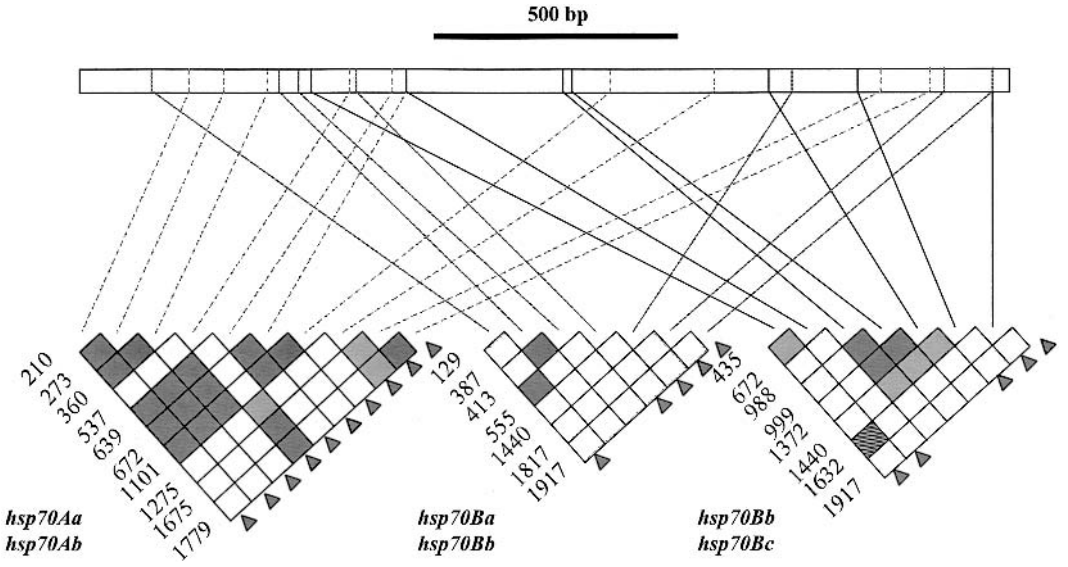


Fig. 6. Polymorphism and linkage disequilibria at conversion-mediated sites in the Pennsylvanian sample. The *horizontal white bar* represents *hsp70* CDS, drawn roughly to the size indicated by the *black scale bar. Ticks* mark positions of sites involved in intracluster conversion, with the line color/style indicating the gene pair examined: *light gray/dashed, hsp70Aa/hsp70Ab; gray, hsp70Ba/hsp70Bb; black, hsp70Bb/hsp70Bc.* Only informative sites were considered. *Triangular matrices* represent the statistical significance determined by Fisher's exact test (uncorrected for multiple tests); *white,* $p \geq 0.05$; *light gray,* $p < 0.05$; *dark gray,* $p < 0.01$. *Dashed lines* connect polymorphic sites to corresponding matrix columns. *Gray triangles* to the right of matrices indicate shared polymorphic sites participating in intracluster conversion.

within the PA population. Nucleotide variability is dramatically reduced at the PA *hsp70Ab* and *hsp70Bb* genes with the exception of conversion-mediated shared polymorphisms. The PA *hsp70Ab* sample contains only one nonshared silent SNP in the CDS, and the 3′-UTR is invariant. Interestingly, the 5′-UTR possesses significant variation on two chromosomes, including an intercluster conversion tract and a small tract of divergent sequence. *hsp70Bb* is strikingly homogenized in the PA sample: the only nonshared varying site is a singleton SNP in the 3′-UTR. Clearly, both genes participate in conversion, as they vary at conversion-mediated sites. However, in general, and especially for *hsp70Ab,* one haplotype dominates at these sites (discussed above; also note the significantly low *D* in Table 2). If conversion rates are heterozygosity dependent, perhaps these two genes have experienced a cyclical reduction of variability and increased rates of conversion in their history. This seems unlikely, however; their paralogues do not display the same pattern despite extremely close physical linkage.

In contrast to *hsp70Ab* and *hsp70Bb,* their nearby paralogues vary. *hsp70Aa* bears three unique SNPs and moderate divergence in the 3′-UTR; *hsp70Bc* bears 11 unique SNPs. Why these gene copies are more variable than their neighbors is not obviously due to a difference in the degree of intragenic crossing-over. If conversion rates are extremely sensitive to these single-nucleotide changes, then perhaps these genes experience less conversion. Furthermore, while *hsp70Bb* and *hsp70Bc* are only ~1 kb apart from one another, their tandem arrangement may be less conducive to conversion than the inverted pair arrangement at 87A7. Some models of gene conversion in *Drosophila* posit that an inverted pair arrangement facilitates the process (Leigh Brown and Ish-Horowicz 1981; Wang et al. 1999). Regardless, the reduction of variability at non-conversion-mediated sites at both the *hsp70Ab* and the *hsp70Bb* PA alleles relative to their nearby paralogues is striking. In contrast to the pattern at nonshared sites, both *hsp70Ab* and *hsp70Bb* in the PA sample vary to no greater or lesser extent than their paralogues at conversion-mediated sites; gene conversion is clearly participating in the strong reduction of nucleotide variability at *hsp70Ab* and *hsp70Bb*. The question remains, What reduced the "standing crop of variation" at these genes in the first place? Selection and/or hitchhiking are intriguing candidate explanations (see Discussion).

Conversion and Selection Act Differently on Unique vs Multigenic Substitutions, in Accordance with Theory. If gene conversion is frequent, any variant that rises in frequency at one gene due to drift or selection should soon either become shared among genes or be eliminated (Ohta and Dover 1984; Ohta 1989). Additionally, conversion ought to accelerate selection against deleterious mutations (see Hypotheses Regarding Intracluster Conversion and Selection, above). Thus, variants unique to any one gene should be rare. The *hsp70* genes at both 87A7 and 87C1 display such a bias. At 87A7, *hsp70Aa* and *hsp70Ab* share all 12 nonsingleton SNPs (excluding intercluster tracts on $hsp70Aa^{122}$) involved in intracluster conversion (Fig. 3). In contrast, 15 singleton SNPs are in

our 87A7 sample. At 87C1, polymorphisms not involved in conversion are again usually singletons: the 87C1 *hsp70* CDSs exhibit 13 shared (conversion-mediated) and 26 unique SNPs (Fig. 3).

Selection against unique replacements may be reduced in redundant gene families, facilitating adaptive divergence (Basten and Ohta 1992; Nowak et al. 1997; Wagner 2000). Any *deleterious* replacement polymorphisms, even those kept rare by conversion-accelerated selection (see above), that spread among genes via conversion should face even stronger selection and be eliminated. Thus, compared to polymorphism unique to any one gene and/or fixed differences between intercluster paralogues, conversion-mediated polymorphism should be skewed toward silent variation. In 4 of 10 cases of conversion-mediated polymorphism, the *hsp70* genes display such a bias (Table 3). This result is according to *G* tests of independence of the synonymous:replacement ratios of unique polymorphisms to shared polymorphisms, with intracluster and intercluster conversion-mediated sites considered separately. In all five comparisons, synonymous changes predominate among the intercluster fixed differences (Table 3).

The hsp70 *Genes of* D. simulans *Resemble Those of* D. melanogaster *in Patterns of Concerted Evolution and Reveal Recent Concerted Change in Both Species*

The *hsp70* gene sequences of *D. simulans* may reveal the generality of and provide comparison to results from *D. melanogaster.* The *hsp70* 3′ flanking regions, which diverge between genes/clusters in *D. melanogaster* [see above and Torok et al. (1982)], display parallel orthology in *D. simulans* (Fig. 7). On average, the 3′ flank of each *D. simulans hsp70* gene is 10% divergent from its putative *D. melanogaster* orthologue. According to the homology of their 3′-UTRs, *D. simulans hsp70Bb* is orthologous to *D. melanogaster hsp70Bc.* The remaining three *D. simulans hsp70* genes are orthologous with their similarly named genes in *D. melanogaster,* in agreement with Leigh Brown and Ish-Horowicz (1981).

Relative to the 3′ flanking sequences, which diverge between the two species, the four *hsp70* CDSs of *D. simulans* are more similar to their orthologues in *D. melanogaster* (Fig. 8). The mean CDS nucleotide identity between orthologues is 98.4%. Because the 87A7/87C1 duplication predates the *melanogaster/simulans* split (Konstantopoulou et al. 1998), the two species share seven 87A7/87C1 informative sites (excluding intercluster-converted alleles).

Orthologues are not identical; both 87A7 and 87C1 display concerted divergence between the two species. Note that the chromosomal positions of the *D. simulans hsp70* clusters have been mapped only to minor-band resolution [i.e., 87A and 87C (see Leigh Brown and Ish-Horowicz 1981)] but, for simplicity, are reported here as

```
position
                1111 1112222222 2
           1156891344 5590011112 3
           3455988923 1463702357 0

DsimAa1    CAGC-GAGC- GGTTGCCATT T
DsimAa2    .......... .......... .
DmelAaPA35 GTATACCAAA TA--AGTTCA G

position
                11111 1122222222 2
            168913445 5900001133 3
           5459889231 4704560902 5

DsimAb1    AAC-GAGC-G GTTTAAATTT -
DsimAb2    .......... .G.-...-.. .
DmelAbPA35 GTTACCAAAT AGA-----GC G

position
                 111 1111122222 22222
           1223399011 2367900000 11112
           1022334704 7471246789 01236

DsimBa1    AGGTAGTTAA CCATAT---- ----G
DsimBa2    .......... .......... .....
DmelBaPA35 TATGCACATT AGTATGCAGG GAGTA

position

           1222223356 66667
           0345790355 67890

DsimBc1    ACAGTCTAA- -----
DsimBc2    .......... .....
DmelBcPA35 T------CTG TGCCA
```

Fig. 7. Nucleotide polymorphism in *D. simulans hsp70* 3′-UTRs and divergence from *D. melanogaster.* Sites labeled as in Fig. 3. Gene names abbreviated according to species, gene, and allele, i.e., $DsimAb^{1}$ = *D. simulans* $hsp70Ab^{1}$ and $DmelBc^{PA35}$ = *D. melanogaster* $hsp70Bc^{PA35}$.

87A7 and 87C1. The *D. simulans* 87A7 *hsp70* genes differ from the *D. melanogaster* 87A7 genes at 51 sites and two indels (excluding the converted $Dsimhsp70Ab^{4}$). At 87C1, the orthologues are divergent at 59 sites and one indel (excluding the converted $Dsimhsp70Bb^{3}$). These unique cluster-specific changes, and 11 substitutions found in all four *D. simulans* genes but absent in the *D. melanogaster* genes, are consistent with concerted evolution after speciation. The rate and nature of this change are described under Discussion.

Intercluster Conversion Reduces D. simulans *87A7/87C1 Divergence.* Large intercluster conversion tracts in the *D. simulans* $hsp70Ab^{4}$ and $hsp70Bb^{3}$ almost completely eliminate the 87A7/87C1 sequence divergence (gray bars in Fig. 9). Including these alleles, one 87A7/87C1 difference is fixed and five polymorphisms are shared; without them, 57 differences are fixed and zero polymorphisms are shared. While DNASP identifies two large contiguous conversion tracts on $Dsimhsp70Bb^{3}$, the number of 87C1-type changes contained within the tracts

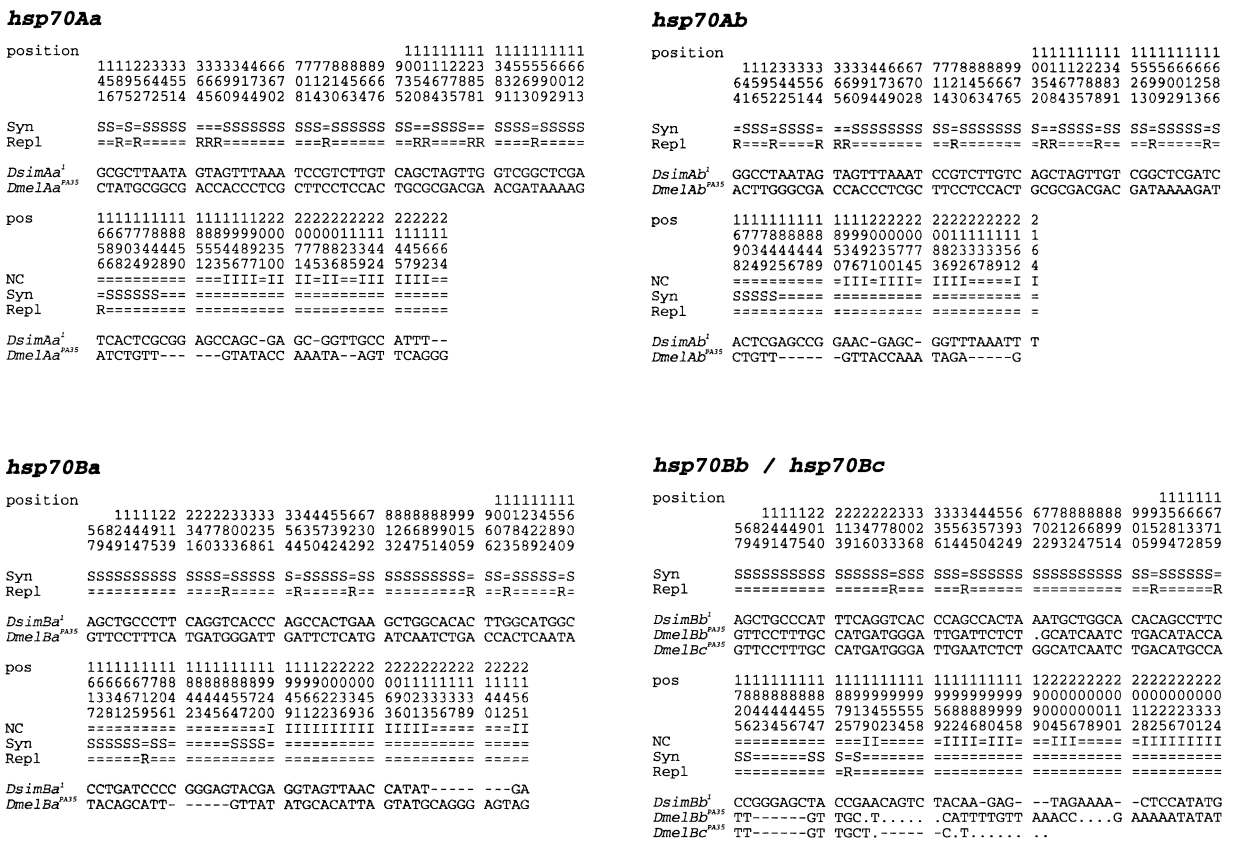


Fig. 8. Nucleotide divergence between *D. simulans* and *D. melanogaster hsp70* coding sequences. Sites labeled as in Fig. 3.

is high (40). Alternative interpretations of this allele are multiple smaller tracts and/or recombination and conversion with unconverted alleles.

Same-Cluster Paralogues Are Homogenized. There are no fixed nucleotide differences between the *Dsimhsp70Aa* and *Dsimhsp70Ab* CDSs, and one between *Dsimhsp70Ba* and *Dsimhsp70Bb* (Fig. 9). The *D. simulans hsp70* genes possess numerous unique replacements, more than in *D. melanogaster* (Fig. 9). As all *D. simulans* alleles were amplified from one laboratory strain, however, this information is of limited value. We thus measure population genetic parameters only in the natural *D. melanogaster* PA population.

In contrast to the *D. melanogaster* genes, *D. simulans* same-cluster paralogues share no polymorphisms (excluding those in *Dsimhsp70Ab*[4] and *Dsimhsp70Bb*[3]), implying less frequent intracluster conversion, Additionally, the *D. simulans hsp70* 3′-UTRs are nearly invariant (only two polymorphic sites, in *Dsimhsp70Ab;* data not shown). The failure to observe shared polymorphisms among same-cluster paralogues may thus reflect these sites' fixation or near invariance in the DSR strain. Indeed, in *D. melanogaster,* only interstrain comparison reveals many clusters of shared polymorphisms, and in the PA sample most variant haplotypes are rare (e.g., the intracluster conversion tracts at 87A7; see Fig. 3).

Discussion

Gene Conversion Plays Dual Roles: Homogenization and Diversification

Homogenization. On a broad scale, the net effect of conversion on the *hsp70* genes is homogenization, especially in coding regions. Examination of the 3′ flanking sequences of *D. simulans* confirms that gene conversion, rather than selection alone, drives homogenization: at several sites where the two species diverge, same-cluster paralogues are identical (Fig. 7). The preservation of redundant and functional copies, rather than their degeneration or divergence as in *D. mauritiana,* suggests that the *hsp70* genes are of critical functional importance (Bettencourt and Feder 2001).

Whether the bias toward low-frequency variation at the *hsp70* genes results from a mechanism (gene conversion) or a population-level process (drift, selection) and/or departs from neutrality remains untestable. Empirical detection and analysis of gene conversion—in single genes—have received a fair amount of recent attention, and some investigators now assign conversion a major role in generating sequence diversity (Berry and Barbadilla 2000). However, molecular evolutionary theory of gene conversion is at a nascent stage: modeling of the

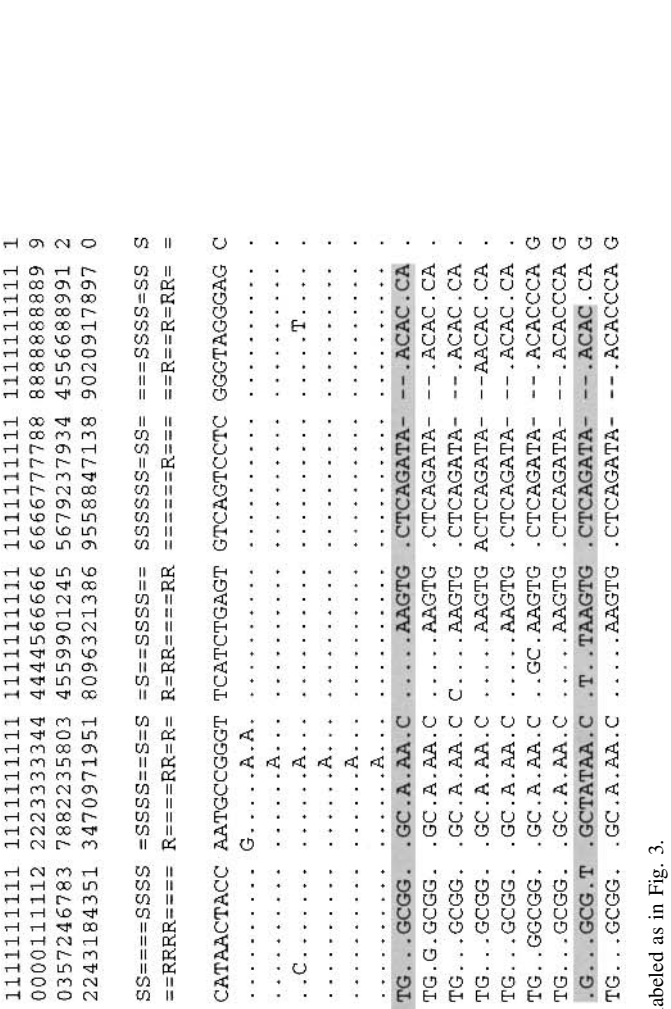


Fig. 9. Nucleotide polymorphism and divergence in the *hsp70* coding regions of *D. simulans*. Sites/tracts labeled as in Fig. 3.

Table 4. McDonald–Kreitman tests of neutrality at *D. melanogaster* and *D. simulans hsp70* genes[a]

Gene	Type of change	Fixed	Polymorphic	p (G)[b]
hsp70Aa	Syn	42 (38)	11 (28)	
	Rep	9 (9)	13 (14)	** (ns)
hsp70Ab	Syn	41 (24)	16 (63)	
	Rep	9 (5)	8 (17)	ns (ns)
hsp70Ba	Syn	55 (45)	8 (31)	
	Rep	5 (4)	14 (20)	*** (***)
hsp70Bb	Syn	58 (40)	10 (43)	
	Rep	6 (4)	11 (22)	*** (**)
hsp70Bc	Syn	58 (50)	13 (26)	
	Rep	6 (4)	18 (24)	*** (***)

[a] Syn, synonymous; rep, replacement. Values when alleles bearing intercluster conversion tracts are excluded are listed first. Values for all alleles follow, in parentheses.

[b] Asterisks indicate the significance level of G tests of independence: **, $p < 0.01$; ***, $p < 0.001$.

coalescent with conversion remains restricted to single genes and has not yet been extended to molecular models or data (Wiuf and Hein 2000). Clearly, in this multigene/cluster family, predictions concerning the frequency distribution of neutral variants with gene conversion versus without gene conversion cannot be formalized and/or tested until the molecular evolution of duplicate genes undergoing conversion is better understood. Furthermore, our broad sampling scheme may overestimate the number of rare variants in *D. melanogaster*. Yet almost all unique polymorphisms are singletons, while almost all polymorphisms shared among genes occur at a high frequency, consistent with frequent gene conversion. Within the PA population, most unique polymorphisms are similarly rare. Moreover, polymorphism in *D. simulans,* sampled in one strain, displays a similarly strong bias toward rare variants.

Although the frequency of polymorphisms may be of limited informative value, their nature (replacement or silent substitutions) is more revealing. Purifying selection appears to act on the *hsp70* genes collectively; while unique replacement polymorphisms are relatively common, few replacements are either shared among genes or fixed between clusters. A serious problem arises when conducting McDonald–Kreitman-type tests on paralogues, however (Table 3). These types of tests require that each locus under comparison have an independent evolutionary history, something which the *hsp70* genes clearly lack. Inasmuch as point mutation and gene conversion are independent processes, partitioning the variation as in Table 3 is a partial fix. Another approach is to compare the divergence between *D. melanogaster* and *D. simulans* orthologues and exclude intercluster conversion tracts (Table 4). These tests reveal a strong difference between largely silent divergence and an excess of replacement polymorphisms. This excess may be interpreted as the presence of deleterious alleles in the DSR strain (Nachman 1998); further sampling beyond this

strain we analyzed will determine the frequencies of such alleles in other *D. simulans* populations. Nonetheless, both of the above results are consistent with paralogous buffering of unique replacements, with very few replacements becoming multicopy shared polymorphisms or fixed differences. Thus, each *hsp70* gene may be relatively free to evolve individually, but only neutral or beneficial changes may become shared among genes via conversion or duplication.

Diversification. While conversion clearly homogenizes the *hsp70* genes, in so doing it also creates broad allelic classes and a mosaic of variability that is shared across genes and gene clusters (Fig. 10). This mosaic structure is especially evident in recent tracts, where variation within the converted region is nonexistent (e.g., *hsp70Aa* 122 type 3′-UTR). In the coding regions, most conversion tracts transmit 87A7/87C1 differences and thus contribute largely silent polymorphism. Intracluster conversion also mediates mainly silent shared polymorphism, with few shared replacements (discussed above, also see *hsp70Ba* and *hsp70Bb* from the 28°C populations in Fig. 3). Overall, conversion in the CDSs is generally restricted to a homogenizing role; however, the process can dramatically diversify the flanking regions.

Diversifying conversion in the *hsp70* CDSs yields byproducts with potential functional consequences: indels in the 5′ and 3′ flanking sequences. Small (3- to 15-bp) indels, both shared and unique, make up a significant component of variation in the 5′- and 3′-UTRs of all five *hsp70* genes (see above and Fig. 10). These indels likely demarcate the boundaries of conversion events between *hsp70* CDSs. Gene conversion is an imperfect process; nonhomologous sequences such as indels often flank conversion tracts (Davies 1992; Liang et al. 1998). The indels occur at a high frequency (with conserved lengths) at different genes, supporting their role in conversion (see Figs. 4 and 6). Also, the abundance of shared SNPs indicates that the flanks themselves participate in conversion (see Results). The shared SNPs and indels only occasionally display linkage as perfect markers of conversion tracts, unlike in the CDS. This is not surprising if, as described earlier, recombination outpaces selection or drift; as converted alleles rise in frequency they will have numerous opportunities to recombine with unconverted alleles. This makes the observed linkage among CDS shared SNPs in the Pennsylvanian population all the more interesting (discussed below). Furthermore, apart from the conversion-mediated indels and SNPs, the sequences of the 5′- and 3′-UTRs are conserved at both 87A7 and 87C1 (see Figs. 4 and 6). This suggests that conversion, rather than relaxation of selective constraint and/or deletion bias (Petrov and Hartl 1998), creates the pattern of variation.

In a more extreme example of the effects of conversion on flanking sequence variability, several *hsp70* alleles share intercluster conversion tracts throughout their entire 5′- and/or 3′-UTRs (Fig. 10). Petersen and Lindquist (1989) found that the 3′-UTR is critical for *hsp70* mRNA degradation following heat shock. Interestingly, the divergent 87A7 and 87C1 3′-UTRs drive different tissue- and time-specific patterns of *hsp70* expression (Prasanth and Lakhotia 1999). Thus, the complete replacement of some *hsp70Aa* 3′-UTRs with 87C1-derived sequences may have organismal consequences.

Natural Selection Acts on hsp70 *Genes*

Selection on Individual Genes. As expected given their participation in conversion, nucleotide variability in the *hsp70* CDSs is low, whether comparing across or within the single PA population. Clearly, purifying selection also plays some role in reducing variation (see above); this can be seen by the strong skew toward synonymous changes between both *hsp70* paralogues and *hsp70* orthologues. However, the observation of strong linkage disequilibrium and biased representation of haplotypes at the largely synonymous shared polymorphic sites is unusual because the expectation under frequent conversion and recombination would be low linkage, even under purifying selection, and there is no evidence that particular haplotypes are favorable. Founder effects and drift could cause for the low variation and strong linkage, if, for example, the PA lines do not represent a good sample of a natural population, but ongoing analyses of other genes in these lines suggests that they are (A.G. Clark, personal communication; also see Lazzaro and Clark 2001). Could other forms of selection cause the strong reduction in nucleotide and haplotype diversity? Since selection on Hsp70 expression in nature is strong (Feder et al. 1997, Roberts and Feder 1999), one explanation for the reduction in nucleotide and haplotype diversity in the *hsp70Ab* and *hsp70Bb* genes could be a selective sweep. Recently, Langley and co-workers (2000) found a pattern of variation similar to that described in this study in two X-linked *Drosophila* genes, *su*(*s*) and *su*(*w*(*a*)), and raised the same hypothesis. A key distinction is that unlike *su*(*s*) and *su*(*w*(*a*)), which are on the tip of the X chromosome, the *hsp70*s are located in the central region of the right arm of the third chromosome and should thus not experience a strong reduction in recombination rates (Hudson and Kaplan 1995). Therefore reduced recombination can be eliminated as a cause of reducing variation prior to or after the history of conversion began. This study echoes the same question raised by Langley et al. (2000): Assuming that conversion is ongoing and increases in frequency with homogeneity, what reduced the standing crop of variation to begin with? The main impediment to answering the question in both studies re-

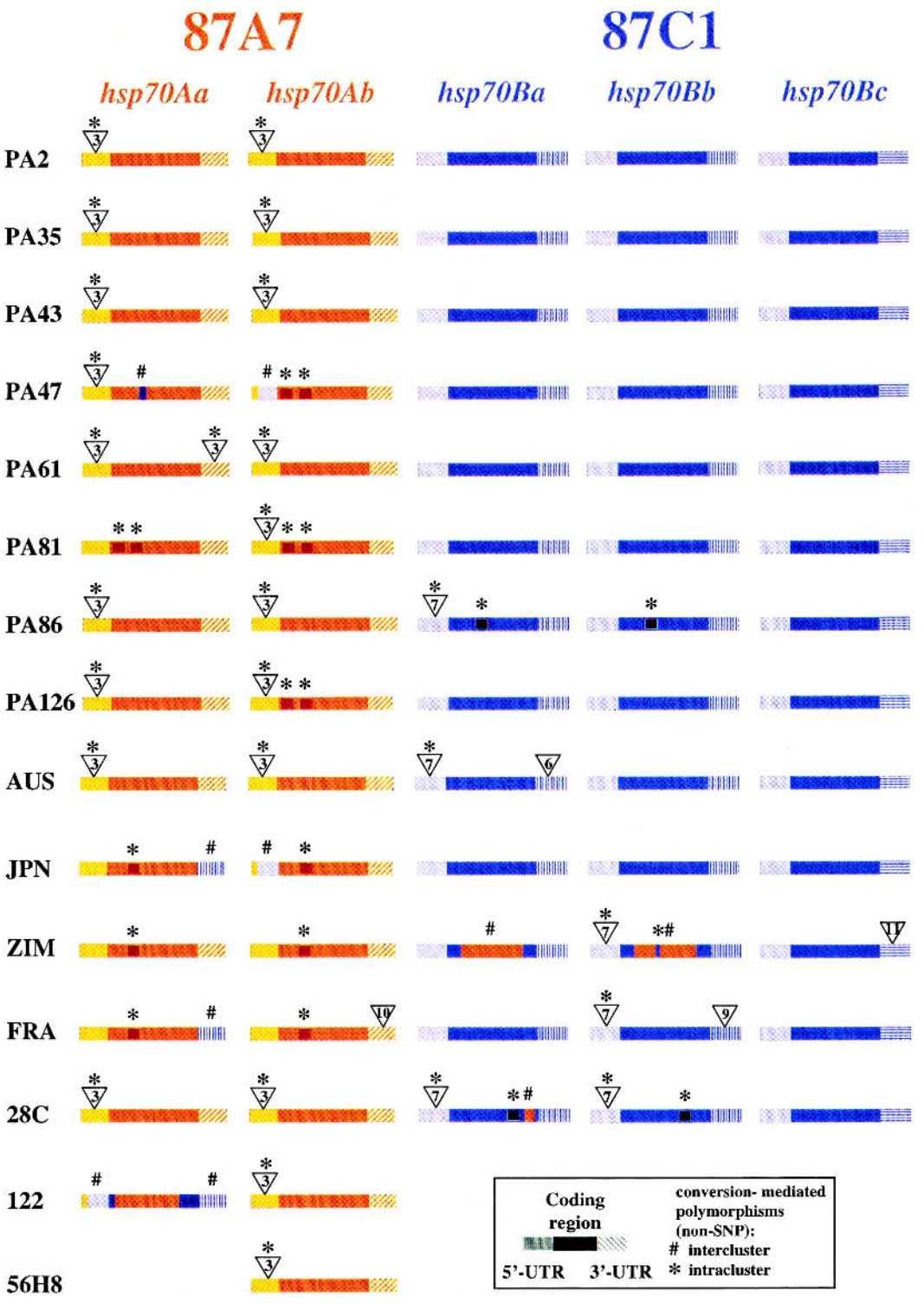


Fig. 10. Schematic of conversion-mediated-polymorphism in *D. melanogaster hsp70* genes. Rows labeled according to source of sequences: PA###, Pennsylvanian strains; AUS, Australian; JPN, Japanese (QD18); ZIM, Zimbabwean [Z(H)1]; FRA, French (FrV_3-1); 28C, A28/B28; 122 and 56H8, 122 and 56H8 genomic clones. Indels indicated by *triangles* (numbers give size in base pairs). Intracluster conversion tracts in coding regions are noted only when all internal informative sites are linked.

mains—one cannot disentangle the relative contributions of conversion vs selection via statistical means without a theoretical model of multigene coalescents with conversion (see above). An alternative approach is to examine whether silent (presumably neutral) sites in nearby genes, which would not participate in conversion between the *hsp70*s, also show a reduction in variability due to hitchhiking (ongoing).

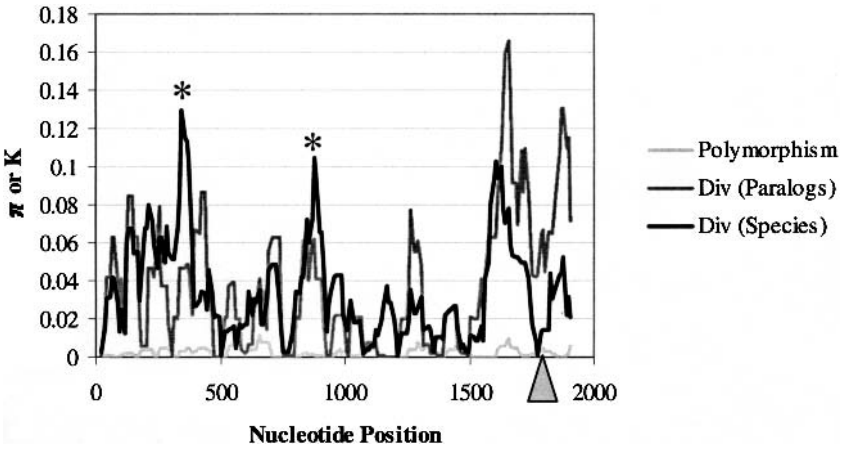


Fig. 11. Sliding-window diagram of nucleotide divergence, *K,* across the *hsp70* coding region. *Light gray line,* mean polymorphism in the five *D. melanogaster hsp70* genes; *dark gray line,* mean divergence between *hsp70* paralogues in *D. melanogaster; heavy black line,* mean divergence between orthologous *D. melanogaster* and *D. simulans hsp70* genes. *Asterisks* mark positions of elevated between-species divergence described under Discussion. The *triangle* marks the indel-rich region (positions 1842–1851).

Selection on Multiple Genes. At the level of the whole gene family, could conversion's homogenizing role fuel "molecular drive" (Ohta and Dover 1984)? Selection against multicopy (presumably deleterious) replacements is effective (see above). Do advantageous mutations, on the other hand, become shared features of the entire gene family?

Comparison of the *D. melanogaster hsp70* genes with those of *D. simulans* reveals the nature of whole-family evolution. Since the 87A7/87C1 duplication predates *D. melanogaster/simulans* speciation, orthologues ought to be more similar to one another than paralogues if the *hsp70* genes were evolving in the typical pattern of duplicate genes diverging with time. In several regions, however, paralogous CDSs are more similar to one another than are orthologous CDSs, reflecting lineage-differential concerted change of all the *hsp70* genes. Comparing 87A7/87C1 divergence within *D. melanogaster* and *D. simulans* vs divergence of the entire gene family between the two species in a "sliding window" diagram reveals regions of the gene/protein displaying this rapid concerted evolution (Fig. 11). Many substitutions are fixed between the two species' *hsp70* CDSs (see Results); interspecies divergence often equals or exceeds intercluster divergence, despite the duplication having predated speciation. Furthermore, a few complex mutations (replaced or indel codons) are also fixed between species. Notably, multiple concerted changes mark the C-terminal region, including indels and both silent and replacement substitutions (see Fig. 11). This poorly defined "regulatory domain" is dispensable for Hsp70 function, except for the conserved C-terminal "EEVD" motif (Freeman et al. 1995); these concerted changes may be neutral. However, species-specific concerted change is also evident in *hsp70*'s ATPase domain, which is highly conserved (Hughes 1993). For example, two silent substitutions are fixed between all *D. melanogaster* and all *D. simulans hsp70* sequences. And, strikingly, positions 364–367 read ACC (threonine) in all *D. melanogaster* sequences, while the same codon differs greatly in *D. simulans:* GTA (valine) in the 87A7 genes and GCC (alanine) in the 87C1 genes (excluding intercluster-converted alleles). Additionally, 69 substitutions distinguish gene clusters between species (e.g., *melanogaster* 87A7 vs *simulans* 87A7), reflecting concerted evolution within clusters.

Clearly, concerted evolution shapes each of the 87A7 and 87C1 gene clusters and all five *hsp70* genes collectively. Diversification fuels selection, and homogenization under purifying selection transmits neutral or beneficial changes among the genes. Whether or not the *hsp70* genes' evolution is truly adaptive remains to be tested. A functional approach that directly examines the expression of naturally occurring *hsp70* variants will be necessary (Feder and Hofmann 1999), as few statistical approaches are yet applicable to multigene families. Moreover, recent sequencing of entire genomes reveals gene families to be very common [e.g. ~40% of *C. elegans* genes are duplicated (see Semple and Wolfe 1999)]. Furthering our understanding of adaptive molecular evolution in general will thus require both the extension of statistical approaches beyond single-gene models and the increased utilization of functional techniques in evolutionary contexts (e.g., transgenes, laboratory selection, expression arrays).

Conclusions

The evolutionary dynamics of the *hsp70* genes are novel in two ways. First, these ancient genes undergo frequent conversion, with no a priori functional reason to suspect so (see the Introduction). Second, gene conversion's normally contrasting roles of diversification and homogenization both interact with selection to shape the gene family's concerted evolution. This mode of evolution may in fact be common if/when gene copies are redundant, and it certainly appears to be powerful: duplicate copies allow greater evolutionary "exploration" of mutational space while buffering against deleterious changes, and the interaction of selection and conversion ensures that only neutral or beneficial mutations spread

among genes and/or fix throughout the whole family. The duplicate *hsp70*s do not follow the roles that would be assigned them by the "birth–death" model of duplicate gene evolution (Nei and Hughes 1992). They neither degenerate into pseudogenes nor diverge from one another. However, the *Drosophila hsp70* genes are clearly not "steady state" (Ohta 1990); they have evolved rapidly and concertedly in a lineage- and cluster-specific fashion. Models of duplicate gene evolution may need expansion to account for dynamic concerted evolution via both diversifying and homogenizing conversion.

Acknowledgments. We thank M. Nordborg and two anonymous reviewers for their comments on an early version of the manuscript and C.M. Bergman, J.P. Comeron, M.A. Daley, J. Fay, M.E. Jensen, N.M. Pearson, K.R. Thornton, and C.-T. Ting, as well as other members of the Feder, Kreitman, Long, and Wu laboratories (University of Chicago), for invaluable discussion and assistance. This work was supported by National Science Foundation Grant IBN-9723298 to M.E.F. and Howard Hughes Medical Institution Predoctoral Fellowship and National Science Foundation Doctoral Dissertation Improvement Grant IBN-9972678 to B.R.B.

References

Agashe VR, Hartl FU (2000) Roles of molecular chaperones in cytoplasmic protein folding. Semin Cell Dev Biol 11:15–25

Aguade M, Langley CH (1994) Polymorphism and divergence in regions of low recombination in *Drosophila.* In: Golding GB (ed) Non-neutral evolution: Theories and molecular data. Chapman and Hall, New York, pp 67–76

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

Aquadro CF, Begun DJ, Kindahl EC (1994) Selection, recombination and the levels of DNA polymorphism in natural populations of *Drosophila.* In: Golding GB (ed) Non-neutral evolution: Theories and molecular data. Chapman and Hall, New York, pp 45–56

Basten CJ, Ohta T (1992) Simulation study of a multigene family, with special reference to the evolution of compensatory advantageous mutations. Genetics 132:247–252

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila.* Nature 356:519–520

Begun DJ, Whitley P (2000) Reduced X-linked nucleotide polymorphism in *Drosophila simulans.* Proc Natl Acad Sci USA 97:5960–5965

Benedict MQ, Cockburn AF, Seawright JA (1993) The Hsp70 heat-shock gene family of the mosquito *Anopheles albimanus.* Insect Mol Biol 2:93–102

Berry A, Barbadilla A (2000) Gene conversion is a major determinant of genetic diversity at the DNA level. In: Singh RS, Krimbas CB (eds) Evolutionary genetics: From molecules to morphology. Cambridge University Press, New York, pp 102–123

Bettencourt BR (2001) Molecular and phenotypic adaptation of Hsp70 and thermotolerance in *Drosophila,* Doctoral dissertation. University of Chicago

Bettencourt BR, Feder ME (2001) *Hsp70* duplication in the *Drosophila melanogaster* species group: How and when did two become five? Mol Biol Evol 18:1272–1282

Bettencourt BR, Feder ME, Cavicchi S (1999) Experimental evolution of Hsp70 expression and thermotolerance in *Drosophila melanogaster.* Evolution 53:484–492

Davies PL (1992) Conservation of antifreeze protein-encoding genes in tandem repeats. Gene 112:163–170

Edwards SV, Gasper J, March M (1998) Genomics and polymorphism of *Agph-DAB1,* an Mhc class II B gene in red-winged blackbirds (*Agelaius phoeniceus*). Mol Biol Evol 15:236–250

Feder ME, Hofmann GE (1999) Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. Annu Rev Physiol 61:243–282

Feder ME, Krebs RA (1998) Natural and genetic engineering of thermotolerance in *Drosophila melanogaster.* Am Zool 38:503–517

Feder JH, Rossi JM, Solomon J, Solomon N, Lindquist SL (1992) The consequences of expressing Hsp70 in *Drosophila* cells at normal temperatures. Genes Dev 6:1402–1413

Feder ME, Blair N, Figueras H (1997) Natural thermal stress and heat-shock protein expression in *Drosophila* larvae and pupae. Funct Ecol 11:90–100

Freeman BC, Myers MP, Schumacher R, Morimoto RI (1995) Identification of a regulatory motif in Hsp70 that affects ATPase activity, substrate binding and interaction with HDJ-1. EMBO J 14:2281–2292

Gloor GB, Preston CR, Johnson-Schlitz DM, et al. (1993) Type I repressors of P element mobility. Genetics 135:81–95

Goldschmidt-Clermont M (1980) Two genes for the major heat-shock protein of *Drosophila melanogaster* arranged as an inverted repeat. Nucleic Acids Res 8:235–252

Gupta RS, Golding GB (1996) The origin of the eukaryotic cell. Trends Biochem Sci 21:166–171

Haber JE (1998) Mating-type gene switching in *Saccharomyces cerevisiae.* Annu Rev Genet 32:561–599

Hudson RR, Kaplan NL (1995) Deleterious background selection with recombination. Genetics 141:1605–1617

Hughes AL (1993) Nonlinear relationships among evolutionary rates identify regions of functional divergence in heat-shock protein 70 genes. Mol Biol Evol 10:243–255

Inomata N, Shibata H, Okuyama E, Yamazaki T (1995) Evolutionary relationships and sequence variation of alpha-amylase variants encoded by duplicated genes in the *Amy* locus of *Drosophila melanogaster.* Genetics 141:237–244

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. Trends Biochem Sci 23:403–405

Karlin S, Brocchieri L (1998) Heat shock protein 70 family: Multiple sequence comparisons, function, and evolution. J Mol Evol 47: 565–577

Konstantopoulou I, Nikolaidis N, Scouras ZG (1998) The *hsp70* locus of *Drosophila auraria* (montium subgroup) is single and contains copies in a conserved arrangement. Chromosoma 107:577–586

Krebs RA, Bettencourt BR (1999) Evolution of thermotolerance and variation in the heat shock protein, Hsp70. Am Zool 39:910–919

Krebs RA, Feder ME (1997) Natural variation in the expression of the heat-shock protein Hsp70 in a population of *Drosophila melanogaster,* and its correlation with tolerance of ecologically relevant thermal stress. Evolution 51:173–179

Langley CH, Lazzaro BP, Phillips W, Heikkinen E, Braverman JM (2000) Linkage disequilibria and the site frequency spectra in the *su*(*s*) and *su*(*w*(*a*)) regions of the *Drosophila melanogaster* X chromosome. Genetics 156:1837–1852

Lazzaro BP, Clark AG (2001) Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the Attacin genes of *Drosophila melanogaster.* Genetics 159:659–671

Leigh Brown AJ, Ish-Horowicz D (1981) Evolution of the 87A and 87C heat-shock loci in *Drosophila.* Nature 290:677–682

Li WH (1997) Molecular evolution. Sinauer, Sunderland, MA

Liang F, Han M, Romanienko PJ, Jasin M (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci USA 95:5172–5177

Lis JT, Prestidge L, Hogness DS (1978) A novel arrangement of tan-

demly repeated genes at a major heat shock site in *D. melanogaster.* Cell 14:901–919

Liu TJ, Liu L, Marzluff WF (1987) Mouse histone H2A and H2B genes: four functional genes and a pseudogene undergoing gene conversion with a closely linked functional gene. Nucleic Acids Res 15:3023–3039

Mayer MP, Rudiger S, Bukau B (2000) Molecular basis for interactions of the DnaK chaperone with substrates. Biol Chem 381:877–885

Moran L, Mirault ME, Tissieres A, et al. (1979) Physical map of two *D. melanogaster* DNA segments containing sequences coding for the 70,000 dalton heat shock protein. Cell 17:1–8

Nachman MW (1998) Deleterious mutations in animal mitochondrial DNA. Genetica 102–103:61–69

Nassif N, Engels W (1993) DNA homology requirements for mitotic gap repair in *Drosophila.* Proc Natl Acad Sci USA 90:1262–1266

Nei M, Hughes AL (1992) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: Tsuji K, Aizawa M, Sasazuki T (eds) 11th Histocompatibility Workshop and Conference, Vol 2. Oxford University Press, Oxford, pp 27–38

Nowak MA, Boerlijst MC, Cooke J, Smith JM (1997) Evolution of genetic redundancy. Nature 388:167–171

Nurminsky DI, Nurminskaya MV, De Aguiar D, Hard DL (1998) Selective sweep of a newly evolved sperm-specific gene in *Drosophila.* Nature 396:572–575

Ohta T (1983) On the evolution of multigene families. Theor Popul Biol 23:216–240

Ohta T (1989) Role of gene duplication in evolution. Genome 31:304–310

Ohta T (1990) How gene families evolve. Theor Popul Biol 37:213–219

Ohta T, Dover GA (1984) The cohesive population genetics of molecular drive. Genetics 108:501–521

Petersen RB, Lindquist SL (1989) Regulation of HSP70 synthesis by messenger RNA degradation. Cell Regul 1:135–149

Petrov DA, Hartl DL (1998) High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol Biol Evol 15:293–302

Polanco C, Gonzalez AI, De La Fuente F, Dover GA (1998) Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions. A possible mechanism to resolve this paradox. Genetics 149:243–256

Polanco C, Gonzalez AI, Dover GA (2000) Patterns of variation in the intergenic spacers of ribosomal DNA in *Drosophila melanogaster* support a model for genetic exchanges during X-Y pairing. Genetics 155:1221–1229

Popadic A, Norman RA, Doanet WW, Anderson WW (1996) The evolutionary history of the amylase multigene family in *Drosophila pseudoobscura.* Mol Biol Evol 13:883–888

Prasanth KV, Lakhotia SC (1999) Tissue and developmental stage specificity and differential stability of the heat shock induced transcripts of the different *hsp70* genes in *Drosophila melanogaster.* 2nd International Workshop on Molecular Biology of Stress Response, Wuhan, China, October

Roberts SP, Feder ME (1999) Natural hyperthermia and expression of the heat-shock protein Hsp70 affect development in *Drosophila melanogaster.* Oecologia 121:323–329

Rozas J, Aguade M (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, Segarra C, Ribo G, Aguade M (1999) Molecular population genetics of therp49 gene region in different chromosomal inversions of *Drosophila subobscura.* Genetics 151:189–202

Schlotterer C, Tautz D (1994) Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution [see comments]. Curr Biol 4: 777–783

Semple C, Wolfe KH (1999) Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. J Mol Evol 48:555–564

Sharma A, Lakhotia SC (1995) In situ quantification of *hsp70* and alpha-beta transcripts at 87A and 87C loci in relation to *hsr-omega* gene activity in polytene cells of *Drosophila melanogaster.* Chrom Res 3:386–392

Sharon D, Glusman G, Pilpel Y, Khen M, Gruetzner F et al. (1999) Primate evolution of an olfactory receptor cluster: Diversification by gene conversion and recent emergence of pseudogenes. Genomics 61:24–36.

Shyue SK, Hewett-Emmett D, Sperling HG, et al. (1995) Adaptive evolution of color vision genes in higher primates. Science 269: 1265–1267

Simon JA, Sutton CA, Lobell RB, Glaser RL, Lis JT (1985) Determinants of heat shock-induced chromosome puffing. Cell 40:805–817

Stephan W, Langley CH (1992) Evolutionary consequences of DNA mismatch inhibited repair opportunity. Genetics 132:567–574

Takahata N, Satta Y (1998) Selection, convergence, and intragenic recombination in HLA diversity. Genetica 102–103:157–169

Torok I, Mason PJ, Karch F, Kiss I, Udvardy A (1982) Extensive regions of homology associated with heat-induced genes at loci 87A7 and 87C1 in *Drosophila melanogaster.* In: Schlesinger MJ, Ashburner M, Tissieres A (eds) Heat shock from bacteria to man. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 19–25

Wagner A (2000) The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. Genetics 154:1389–1401

Walsh JB (1987) Sequence-dependent gene conversion: Can duplicated genes diverge fast enough to escape conversion? Genetics 117:543–557

Wang S, Magoulas C, Hickey D (1999) Concerted evolution within a trypsin gene cluster in *Drosophila.* Mol Biol Evol 16:1117–1124

Wiuf C, Hein J (2000) The coalescent with gene conversion. Genetics 155:451–462

Zhao Z, Hewett-Emmett D, Li WH (1998) Frequent gene conversion between human red and green opsin genes. J Mol Evol 46:494–496