Chromosoma (2004) 113: 223–232
DOI 10.1007/s00412-004-0312-6

RESEARCH ARTICLE

Michael B. Evgen'ev · Olga G. Zatsepina · David Garbuz · Daniel N. Lerman · Vera Velikodvorskaya · Elena Zelentsova · Martin E. Feder

Evolution and arrangement of the *hsp70* gene cluster in two closely related species of the *virilis* group of *Drosophila*

Received: 29 June 2004 / Revised: 15 July 2004 / Accepted: 11 August 2004 / Published online: 5 October 2004
© Springer-Verlag 2004

Abstract To investigate the genetic basis of differing thermotolerance in the closely related species *Drosophila virilis* and *Drosophila lummei*, which replace one another along a latitudinal cline, we characterized the *hsp70* gene cluster in multiple strains of both species. In both species, all *hsp70* copies cluster in a single chromosomal locus, 29C1, and each cluster includes two *hsp70* genes arranged as an inverted pair, the ancestral condition. The total number of *hsp70* copies is maximally seven in the more thermotolerant *D. virilis* and five in the less tolerant *D. lummei*, with some strains of each species exhibiting lower copy numbers. Thus, maximum *hsp70* copy number corresponds to *hsp70* mRNA and Hsp70 protein levels reported previously and the size of heat-induced puffs at 29C1. The nucleotide sequence and spacing of the *hsp70* copies are consistent with tandem duplication of the *hsp70* genes in a common ancestor of *D. virilis* and *D. lummei* followed by loss of *hsp70* genes in *D. lummei*. These and other data for *hsp70* in *Drosophila* suggest that evolutionary adaptation has repeatedly modified *hsp70* copy number by several different genetic mechanisms.

Communicated by R. Paro

M. B. Evgen'ev · O. G. Zatsepina · D. Garbuz · E. Zelentsova
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences,
Vavilov str. 32,
117984 Moscow, Russia

M. B. Evgen'ev · V. Velikodvorskaya
Institute of Cell Biophysics,
Puschino, Russia

D. N. Lerman · M. E. Feder
The Committee on Evolutionary Biology, The University of Chicago,
1027 E. 57th Street,
Chicago, IL, 60637, USA

M. E. Feder (✉)
Department of Organismal Biology and Anatomy, The Committees on Genetics and on Evolutionary Biology, The University of Chicago,
1027 E. 57th Street,
Chicago, IL, 60637, USA
e-mail: m-feder@uchicago.edu

Introduction

Invasion and exploitation of novel habitats by species often require the co-evolution of genes encoding the molecular mechanisms that enable the invading organisms to withstand the novel environment. Although this process is increasingly well understood in real time for prokaryotes (Elena and Lenski 2003), for complex eukaryotes its occurrence is typically inferred from patterns of genomic variation remaining from genetic events of long ago (Li 1997). In complex eukaryotes, documenting molecular adaptation as it happens is still comparatively rare (but see for example Huey et al. 2000; Lerman et al. 2003). Here we describe the changes in complex loci containing various numbers of *hsp70* gene copies that have accompanied the evolution of two closely related species, *Drosophila lummei* and *Drosophila virilis*. Both are members of the *virilis* group of *Drosophila*, which comprises 12 species divisible into two phylads: *virilis* and *montana* (Patterson and Stone 1952; Throckmorton 1982). According to many investigators (Patterson and Stone 1952; Evgen'ev et al. 1982; Throckmorton 1982; Spicer 1991, 1992; Nurminsky et al. 1996) *D. virilis* is the most primitive species of the *virilis* phylad and is probably ancestral to it if not to the entire *virilis* group. Its distribution throughout the Northern Hemisphere is primarily below 40°N latitude. *D. lummei*, considered the closest relative of *D. virilis*, occurs from just above 40° to just above 65°N latitude and from Sweden west to the Pacific coast of Asia. The most parsimonious interpretation of the above information is that *D. lummei* descended from *D. virilis* or a common ancestor approximately 5 MYA (Spicer 1992; Spicer and Bell 2002) and invaded higher latitudes than its ancestor.

Given the correlation of latitude and temperature, our prior work (Garbuz et al. 2002, 2003) focused on the

phenotypic aspects of temperature tolerance, which show countergradient variation with the latitudinal ranges of the two species. In the cold, *D. lummei* is far more tolerant than *D. virilis*, and also undergoes a diapause absent in all other *virilis* group species (Lumme 1982). By contrast, *D. virilis* has greater basal and inducible high temperature tolerance than does *D. lummei* (Mitrofanov and Blanter 1975; Garbuz et al. 2002, 2003). In *Drosophila*, the DnaK-Hsp70 superfamily of molecular chaperones provides a key molecular mechanism underlying high temperature tolerance (Feder and Krebs 1998; Zatsepina et al. 2001). Corresponding to their differing temperature tolerances, *D. virilis* expresses both more diverse electromorphs of Hsp70-family members and several-fold more Hsp70-family protein and mRNA than does *D. lummei*, especially at high temperatures (Garbuz et al. 2002, 2003). Different strains or populations of *D. virilis* also vary in Hsp70 levels (Garbuz et al. 2002, 2003).

Here we examine the genetic basis for these inter-specific and intra-specific differences in thermotolerance and Hsp70-family proteins. All Diptera that have been studied share an inverted pair of Hsp70-encoding genes at a single chromosomal locus; in mosquitoes (Benedict et al. 1993) and the relatively basal *Drosophila* species *Drosophila pseudoobscura*, (B. Bettencourt, pers. commun.) these are the only known *hsp70* genes. The *hsp70* genes have undergone various duplications in more derived species, however. In the *montium* subgroup of the *melanogaster* species group of *Drosophila*, for example, one of these paired genes has undergone a tandem duplication to yield three *hsp70* genes (Konstantopoulou et al. 1998). In some other subgroups of this species group, the entire locus has undergone duplication to yield four *hsp70* genes; in *Drosophila melanogaster*, the derived locus (at 87C1) has undergone tandem duplication to yield five (in various wild or wild-type strains) or six (in a single laboratory strain) *hsp70* genes total (Leigh-Brown and Ish-Horowicz 1981; Bettencourt and Feder 2001; Maside et al. 2002). Given that the *virilis* group is nearly basal to the *Drosophila* lineage, our expectation was that *D. virilis* and *D. lummei* should have only the inverted pair of *hsp70* genes at a single locus, which would be consistent with earlier cytogenetic localizations of *hsp70* to a single locus (29C) in *D. virilis* (Evgen'ev et al. 1978; Peters et al. 1980). To the contrary, here we report that while the *hsp70* locus is singular, the *hsp70* genes have undergone multiple tandem duplications to yield differing *hsp70* copy numbers both within and between *D. virilis* and *D. lummei*. These differences may explain in part the differing thermotolerances of these populations and species.

Materials and methods

Drosophila strains and collection dates

Drosophila virilis strains 160 (b, gp, cd, pe, gl), 9 (Batumi, Caucasus), 1433 (Leeds, England), T53 and T40 (collected in Tashkent 30 years apart) and *D. lummei* strains 200 (Moscow region, ca. 1970), 202 (Krasnodar, Southern Russia) and 1102 (Finland) were obtained from the stock center of the Institute of Developmental Biology, Moscow. The following *D. virilis* strains were obtained from Dr. Anneli Hoikkala, Oulu University, Finland: A11 (Matsuyama, Ehime, Japan, 1973), V-ZZP-01 (Zeziping, Hunan Province, China, 2001), V-WW-08 (Wuwei, Gansu Province, China, 2002), SBB (Sapporo, Hokkaido, Japan, 1986), V-Hunan (Hunan Province, China), and V-Nanjing (Nanjing, Jiangsu Province, China). The latter strains were collected within the past 2 years. All flies were reared on a yeast, cornmeal, molasses, and agar medium at 25°C.

Cytological analysis

Larvae were grown at 18°C on medium supplemented with live yeast solution for 2 days before dissection. Salivary glands from third instar larvae were dissected in 45% acetic acid and squashed (Lim 1993). Procedures and labeling of DNA probes for in situ hybridization were as described (Lim 1993).

DNA manipulations and Southern analysis

Southern blot analysis of *D. virilis* and *D. lummei* genomic DNA was performed (Evgen'ev et al. 2000a). Five micrograms of each DNA sample was digested with restriction endonucleases. After agarose gel electrophoresis, each gel was treated for 15 min in 0.25 M HCl and then incubated twice in denaturing buffer (1.5 M NaCl, 0.5 M NaOH) for 30 min. After 30 min incubation in neutralization buffer, gels were capillary-blotted onto nylon membranes and fixed by UV cross-linking using the UV Stratalinker 2400 (Stratagene) protocol. Standard high-stringency hybridization and wash conditions were used for Southern blot analysis. To detect Hsp70 sequences in Southern blots and genomic libraries, fragments of cloned *D. lummei*, *D. virilis*, or *D. melanogaster hsp70* genes were labeled by random priming and used as probes. The original screen of all libraries and Southern blots was performed with 5′ specific and 3′ specific probes generated from the *Cla*I–*Sal*I fragment containing an *hsp70* gene of *D. melanogaster* (McGarry and Lindquist 1985). The 5′-specific probe represents the *Cla*I–*Bam*HI portion of the clone indicated, while the 3′-specific probe represents the *Bam*HI–*Sal*I portion.

Genomic libraries from *D. virilis* strain 160 and *D. lummei* strain 200 were prepared by partial *Sau*3A digestion with subsequent ligation into the *Bam*HI site of lambda Dash phage arms (Stratagene). Isolated phages containing *hsp70* genes were identified by restriction analysis and hybridization (see above), and subcloned into KS-Bluescript for subsequent sequencing (with one exception, see below). Clones were sequenced with Sequenase (Amersham) and ABI 377 sequencers. Se-

quences were assembled manually and aligned using CLUSTAL X (Jeanmougin et al. 1998). Relevant sequence information has been deposited in GenBank.

When the above procedure failed to yield reliable sequence for *hsp70c* and the *hsp70c–d* intergenic region in *D. lummei*, genomic DNA isolated from the *D. lummei* 200 strain was digested with *Xba*I, and fragments varying between 2.5 kb and 3.5 kb ligated into pBluescriptSK. Presumptive positive clones were detected by colony hybridization to the *Xba*I–*Sac*I fragment of *D. lummei hsp70d*, and sequenced. Primers were designed complementary to the 3′ end of this clone (5′-AATAATAAGAGCTAGAGC-3′) and to the 5′ end of clone 17 (5′-ATACCAGAACGTGATCAGAA-3′), which contains *hsp70d*. These were used to amplify the *hsp70c–d* intergenic region from genomic DNA with the following conditions: 2.0 mM $MgCl_2$; 10 pg of each primer; 5 min at 95°C, 30 cycles of 94°C for 1 min denaturation, 58°C for 30 s, and 70°C for 1.5 min, and finally 70°C for 7 min. The resultant 2.8 kb product was cloned into pGEM T-Easy Vector (Promega) and sequenced from both ends as described above.

For densitometric estimation of *hsp70* gene copy numbers, X-ray film images of genomic Southern blots were analyzed with a Molecular Dynamics 300A Computing Densitometer and Origin 6.0.

For densitometric estimation of *hsp70* gene copy numbers, genomic DNA of selected *D. virilis* and *D. lummei* strains was digested with either *Xba*I or *Xba*I + *Acc*I before Southern blotting. Blots were hybridized with the *Xba*I–*Sac*I fragment of the lambda phage 17 clone, which contains the corresponding sequence of the *D. virilis hsp70g* gene (Accession no. AY445084). X-ray film images of blots were analyzed with a Molecular Dynamics 300A Computing Densitometer and Origin 6.0. *Xba*I and *Xba*I + *Acc*I digestion yielded an 11 kb and a 5.5 kb fragment, respectively, containing a single *hsp70* gene; for each strain, the density of other fragments relative to this fragment was calculated. Five such independent estimates of density for each selected strain were averaged to calculate the putative *hsp70* copy number.

Results

Chromosomal localization and puffing of *hsp70* gene loci in the genomes of *D. virilis* and *D. lummei*

Putative *hsp70*-containing phages isolated from genomic libraries of *D. virilis* and *D. lummei*, identified by hybridization to a *D. melanogaster hsp70* probe, hybridized strongly with a single region of Chromosome *II* (Fig. 1a). The *D. melanogaster hsp70* probe also strongly hybridized to this region (data not shown). Previously the *hsp70* locus was localized cytogenetically to 29C in Chromosome *II* in *D. virilis* (see Introduction).

As expected, the 29C region forms large puffs after temperature elevation in both species and in inter-specific hybrids. Characteristically, puffs in the 29C region in

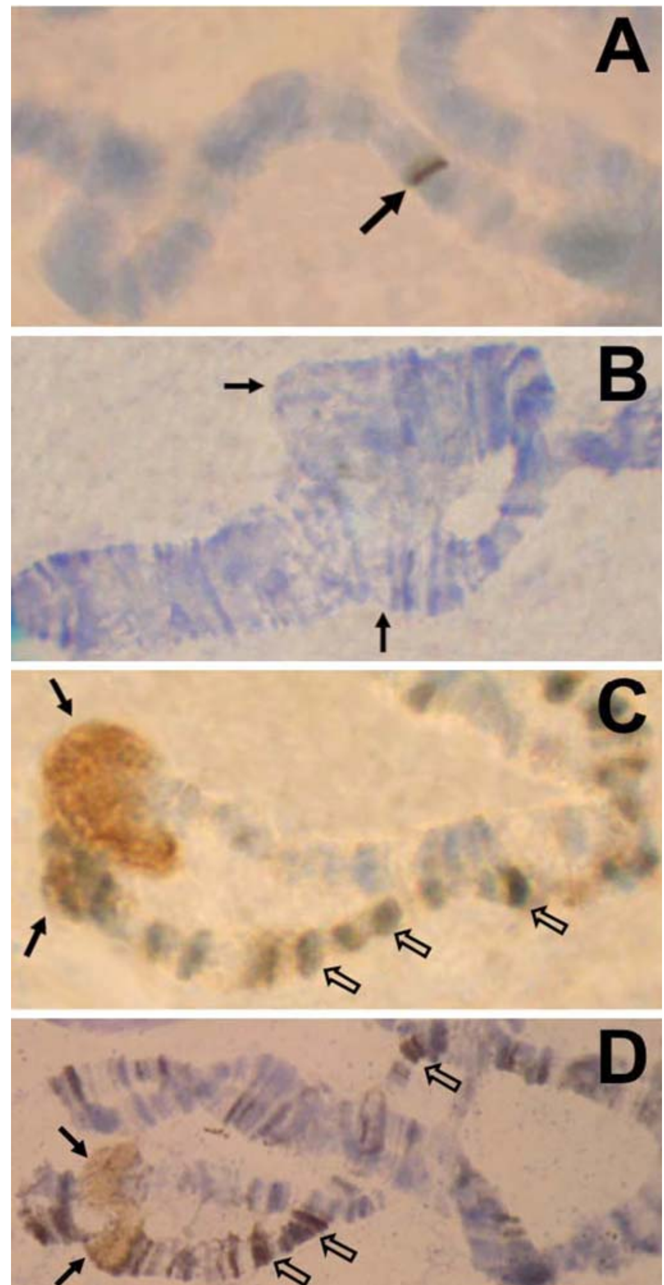

Fig. 1a–d In situ hybridization to specific loci and heat-shock puffs in chromosomes of *Drosophila virilis* and *D. virilis* × *D. lummei* strain 200 (with 3 complete *hsp70* copies) hybrids. **a** A *D. virilis hsp70* probe hybridizes to a single locus in *D. virilis* strain 160. **b** In *D. virilis* strain 160 (with at least 7 *hsp70* copies) × *D. lummei* hybrids, the heat-induced puff at 29C is assymetric, corresponding to the differing *hsp70* copy number in the two parent species. Specific probes were not used to label these chromosomes. **c**,**d** In situ hybridization of a *D. lummei* probe to chromosomes of *D. virilis* × *D. lummei* hybrids. The probe, lambda phage 17, includes both a complete copy of *hsp70* (*hsp70d*) and flanking sequence, which also contains a repetitive element specific to *D. lummei*. Thus this probe labels both a heat-induced puff in the 29C region of each homeolog (*solid arrows*) and the repetitive element throughout the *D. lummei* homeolog (e.g., *open arrows*). The *D. virilis* parent was strain 160 in **c** and strain ZZP (with 4 putative *hsp70* copies) in **d**. Note the differing size of labeled puffs, consistent with the differing *hsp70* copy numbers

inter-specific hybrids were often asynapsed, with the homeologous regions differing in size (Fig. 1b–d). Several features unambiguously assign each homeologous region to its parental species. One probe used in in situ labeling, *D. lummei* phage 17 DNA, contains an unidentified transposable element present in *D. lummei* but absent in *D. virilis* (unpublished data) in addition to the *hsp70d* gene. Conversely, other hybridizations (data not shown) were with probes that contained *Penelope* and *Ulysses*, transposable elements present in *D. virilis* but absent in *D. lummei* (Zelentsova et al. 1999) (data not shown). In all cases (e.g., Fig. 1c,d) the *D. lummei* homeolog was the smaller component in heat-induced puffs at 29C from *D. virilis* × *D. lummei* hybrids.

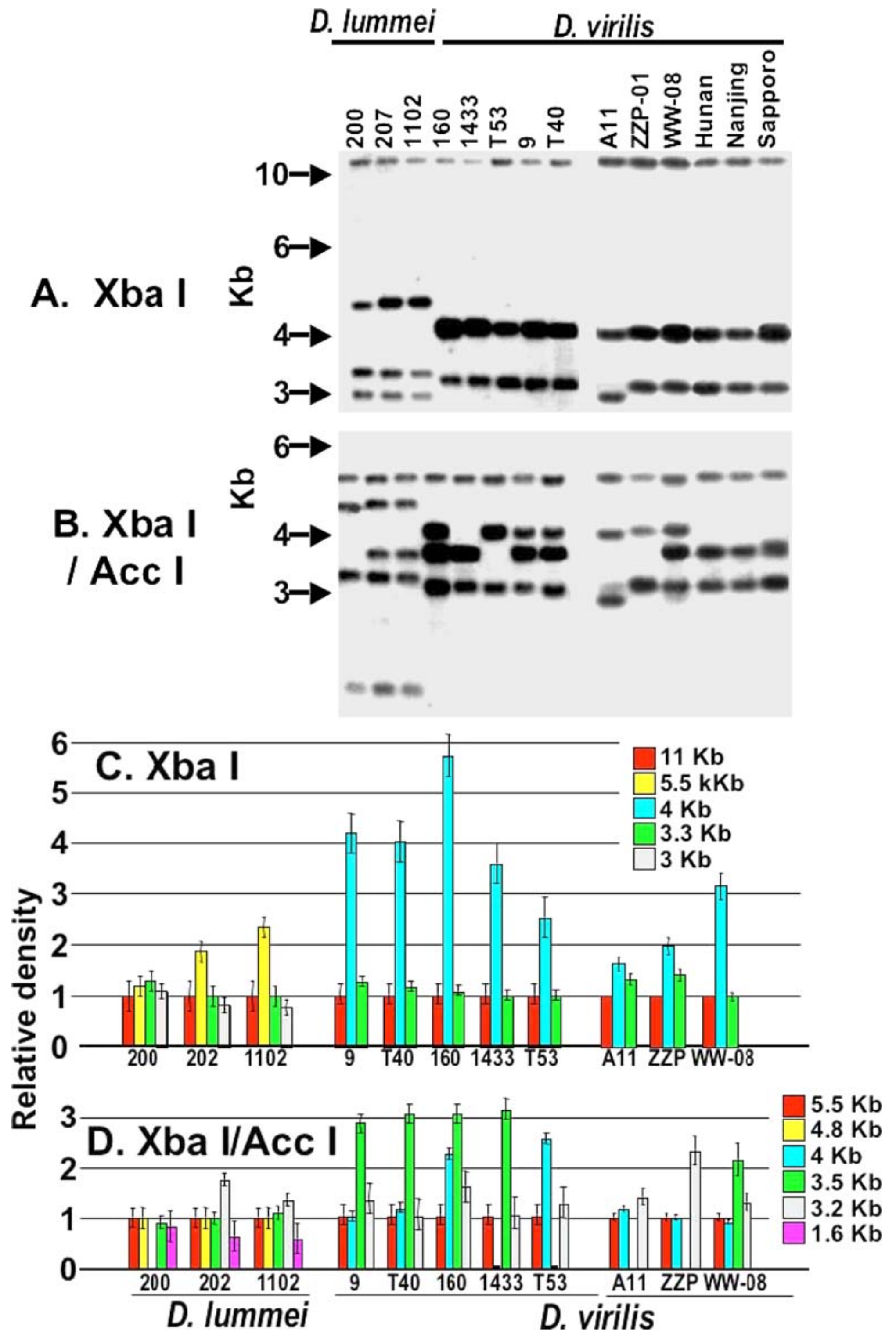

Fig. 2a–d Restriction fragment variation in the *hsp70* gene clusters of *Drosophila virilis* and *D. lummei*. **a**,**b** Variation in genomic DNA from *D. virilis* and *D. lummei* strains hybridized to a *D. virilis hsp70* probe: **a** *Xba*I digest; **b***Xba*I + *Acc*I digest. Note that the first eight strains and the second six strains were on separate blots. **c**,**d** Relative density of *hsp70*-hybridizing restriction fragments in selected *D. virilis* and *D. lummei* strains. Means and standard errors are plotted, and represent summary data from multiple hybridization experiments. *Xba*I digestion yielded an 11 kb fragment containing a single *hsp70* gene; for each strain, the density of other fragments is standardized according to this fragment in **c**. *Xba*I/*Acc*I digestion yielded a 5.5 kb fragment containing a single *hsp70* gene; for each strain, the density of other fragments is standardized according to this fragment in **d**

Inter-specific and intra-specific variation in *hsp70* gene copy number

Southern blot hybridizations indicate that *D. virilis* and *D. lummei* differ strikingly in the size of genomic DNA fragments generated by restriction enzyme digestion (Fig. 2a). Although such differences are consistent with variation in restriction sites, *hsp70* copy number, or both, we present evidence below that these differences reflect intra-specific and inter-specific variation in *hsp70* copy number, at least in part. Densitometry of genomic Southern hybridizations for *D. virilis* (Fig. 2c,d) is consistent with seven copies of *hsp70* genes in strain 160, six copies in strains 9, 1433 and T40, and four to five copies in strains T53, V-ZZP-01, V-WW-08, V-Hunan (data not shown), V-Nanjing (data not shown), and SBB (data not shown), and only three copies in strain A11. Two different restriction digests yield nearly identical data. Densitometry for the *D. lummei* strains studied is likewise consistent with intra-specific variation in copy number (four to five copies of the *hsp70* gene depending on strain; quantitative data not shown), although the maximum copy number (five) is less than in *D. virilis*.

*Xba*I and *Xba*I/*Acc*I restriction digests of *D. lummei* strain 200 genomic DNA were hybridized to both a complete *D. lummei hsp70* sequence and a probe (the *Bam*HI–*Sal*I portion of *D. melanogaster hsp70*) specific for the 3′ region of *hsp70* coding sequence (Fig. 3). Whereas the former probe hybridized to all fragments, the latter probe failed to hybridize to the 2.8 kb *Xba*I fragment and the 1.6 kb *Xba*I/*Acc*I fragment, suggesting that this sequence (which below we name *hsp70c*) does not encode a complete Hsp70 protein. This truncation is present in all *D. lummei* strains studied so far (data not shown).

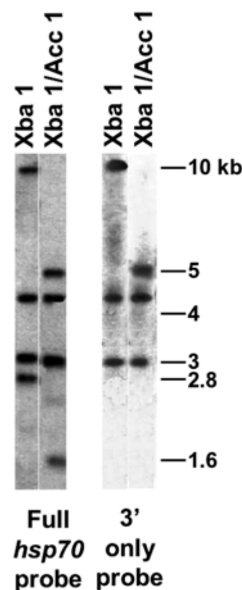

Fig. 3 Hybridization of *Drosophila lummei* DNA restriction fragments with full-length *D. lummei hsp70* sequence and the 3′ portion (*Bam*HI–*Sal*I fragment) of *D. melanogaster hsp70*

Structure and sequence of the *hsp70* gene cluster in *D. virilis* and *D. lummei*

The detailed organization of the *hsp70* genes in *D. virilis* strain 160 and *D. lummei* strain 200 emerges from the subcloning and sequencing of overlapping lambda clones isolated from the corresponding genomic libraries (Fig. 4). In *D. virilis* strain 160 there are six copies in tandem orientation approximately 4.8 kb apart (*hsp70b–g*), and a seventh copy (*hsp70a*) in inverse orientation. Although genes *c–f* are identically spaced, they are distinguishable by the location of restriction sites (e.g., an *Acc*I site adjacent to the 3′ end of gene *f*, *Hin*dIII sites in gene *e* and *d*) and 3′ flanking sequence. By contrast, in *D. lummei* strain 200 there are a pair of copies (*hsp70a* and *hsp70b*) in inverse orientation as in *D. virilis*, a third copy (*hsp70c*) 5.5 kb from *hsp70b*, and a fourth copy (*hsp70d*). Failure of a 3′ region-specific probe to hybridize to its corresponding restriction fragment (see above) suggests that *hsp70c* of *D. lummei* is a pseudogene. In both species, the *hsp70a–hsp70b* intergenic region is approximately 0.8 kb. As described above, the lambda clones yielding these data all hybridize to the 29C region of Chromosome *II*.

The complete coding sequence was obtained for six of the seven *hsp70* genes in *D. virilis* strain 160 (Table 1, Genbank Accession numbers AY445084–7 and AY445090–1). The six genes are highly conserved, varying at only five of 1923 sites. Of these varying sites, three are silent and two result in changes in amino acid sequence (998, GGG Glycine in one of six genes for GTG Valine; 1472, GTA Valine in one of six genes for GCA Alanine). In *D. lummei* strain 200, sequence was obtained for the coding region of the *hsp70* genes and pseudogene (Table 1, Genbank Accession numbers AY4450888–9 and AY445092–3). Compared with the *D. virilis* genes, the complete *D. lummei hsp70* genes exhibit a 9 bp deletion beginning at the counterpart of site 1456 in the *D. virilis hsp70* genes. With these exceptions, the complete *D. lummei hsp70* genes consistently differ from *D. virilis hsp70* genes *a–c* and *e–g* at only 21 of 1923 sites, of which 19 are silent differences and two result in a single replacement (1666–8, TCC Serine in *D. lummei* and CCT Phenylalanine in *D. virilis*). The three complete *D. lummei* genes vary among themselves at ten sites (five silent substitutions and five replacements), and *hsp70b* lacks an entire codon (at 1873–5) present in the other two genes and in the *D. virilis hsp70* genes.

For all of the above genes, at least 408 bp of sequence was obtained 5′ to the start codons. These *D. lummei* and *D. virilis* flanking sequences differ consistently at 17 sites, and all six *D. virilis hsp70s* exhibit a 15 bp deletion beginning 86 bp 5′ to their start codons. With one exception (371 nucleotides 5′ to coding sequence), these differences are at positions whose functional significance has not been established experimentally or is negligible; position −371 is in a putative GAGA element. Within each species, the *hsp70* genes are highly conserved, varying at three to four sites in each species. The 3′ flanking sequence is likewise highly conserved among the *D.*

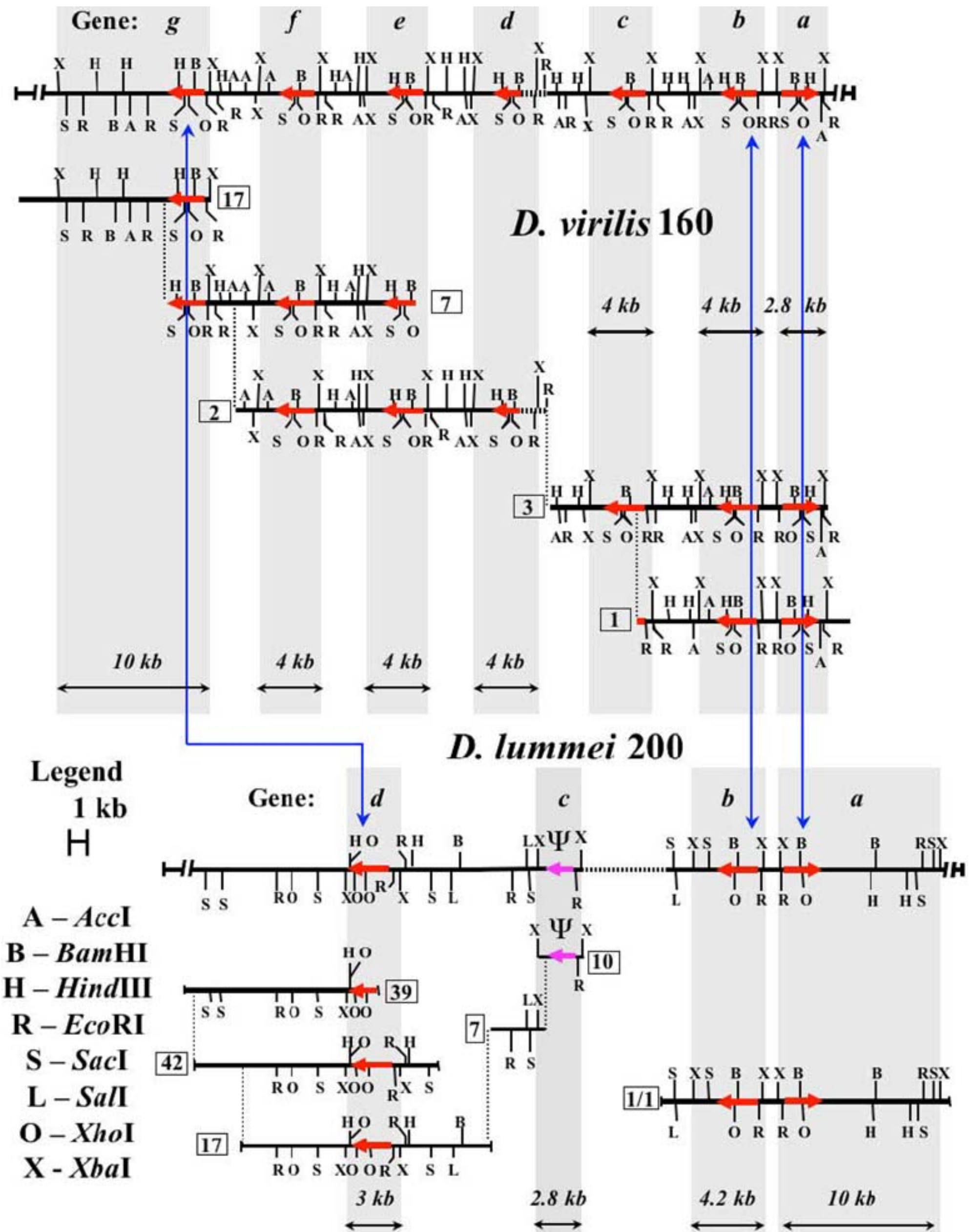

Fig. 4 Arrangement of *hsp70* copies at chromosomal locus 29C in *Drosophila virilis* (*top*) and *D. lummei* (*bottom*). The *purple arrow* indicates a putative pseudogene. Note scale in legend. *Grey bars* indicate size of *Xba*I fragments in kilobases. *Numbers in boxes* indicate specific lambda clones from which the arrangement (*top row* for each species) was deduced; in these arrangements (*top rows*), *broken horizontal lines* indicate hypothesized regions for which no contiguous lambda clones or sequences were obtained. Restriction sites include those represented in Fig. 2 and others (see key)

lummei and *D. virilis hsp70* genes. In particular, the 3′ flanking sequence of *D. virilis hsp70b*, *hsp70e*, and *hsp70f* is near-identical for more than 1600 bp.

The *D. lummei* pseudogene *hsp70c* lacks the first 300 nucleotides of coding sequence and the 13 nucleotides immediately 5′ to those. It then contains the next 809 nucleotides of *D. lummei hsp70* genes, but lacks the remaining 813 nucleotides of *D. lummei hsp70* coding sequence.

Discussion

From the foregoing data, we infer that multiple tandem duplications of *hsp70* genes accompanied the evolution of the *D. virilis* group, and that changes in *hsp70* copy number accompanied changes in thermotolerance both among *D. virilis* populations and between *D. virilis* and *D. lummei*. Multiple lines of evidence are all consistent with this inference: in situ hybridization indicating that the *hsp70* locus itself is singular: chromosome puff size corresponding to putative *hsp70* copy number; genomic Southern hybridizations and quantitative densitometry of restriction fragment polymorphisms; and the coding and flanking sequences of the genes themselves. Moreover, the data are consistent with earlier chromosomal localizations of these genes (Evgen’ev et al. 1978; Peters et al. 1980).

Several distinctive features of *D. lummei* and *D. virilis* have facilitated this diversity of approaches. First, decades of study uniformly indicate that the *virilis* subgroup is a (if not the) basal member of the *Drosophila* clade, that *D. virilis* is itself a (if not the) basal member of the group, and

Table 1 Nucleotide divergence in the *hsp70* coding and 5′ flanking sequences of *Drosophila lummei* and *D. virilis*. Each column provides the site number (e.g., −0408 in the first column for 5′ flanking sequence) and the corresponding nucleotide for each gene included. The first base of the start codon is numbered + 1; the flanking nucleotide just 5′ to that is numbered − 1. For coding sequence, synonymous and replacement differences are noted and labeled S and R, respectively. For replacement differences, the corresponding amino acids are also listed. Nucleotides identical to the first sequence are indicated by dots; deletions by dashes. Blank entries indicate lack of sequence for the indicated nucleotide position. This analysis does not include *hsp70d* of *D. virilis*, for which only partial sequence was obtained, and *hsp70c* of *D. lummei*, which is a putative pseudogene

A. 5' flanking sequence (out of 408 sites in vir_*hsp70a*)

```
            -------------------------------------
            0000000000000000000000000000000000000
            4333332211111111000000000000000000000
            0987559065531111099999999998888764443
  Position: 8311108727336542098765432109876478511

vir_hsp70a  TGTCCGTCGATGATAC---------------ACCATG
vir_hsp70b  ......A..............................
vir_hsp70c  ..........G..........................
vir_hsp70e  ..........G..........................
vir_hsp70f  ..........G..........................
vir_hsp70g  ...T.A....G..........................
lum_hsp70a  .AA.T..AATGACGTAATTAAGAAAGATCAATTAGCC
lum_hsp70b  .AA.T...ATGA.GTAATTAAGAAAGATCAATTAGCC
lum_hsp70d  AAA.T...ATGA.GTAATTAAGAAAGATCAATTAGCC
```

B. Coding sequence (out of 1923 sites in vir_*hsp70a*)

```
            000000000000000111111111111111111111111111111111
            012244555677899900223334444444444466666677788888
            247030232263349484505715555666668356677889017778
  Position: 713251214911088997189236789012342536859281303457

vir_hsp70a  CCTACGGACTAACGTACCGCGGAGTGACGGCCCCGCTCAACCTAGGCC
vir_hsp70b  ................................................
vir_hsp70c  ..G...........G.................T...............
vir_hsp70e  ..GG............................................
vir_hsp70f  ..GG............................................
vir_hsp70g  ..G..................A..........................
lum_hsp70a  AGGG.TCT.CTG.A..G..TA.G..........TCTC.GGTTCT....
lum_hsp70b  AGGG.TCT.C.G.A..G..TA.G..........TCTC.GGTTC.---T
lum_hsp70d  AGGGATCTTC.GTA.GGGATA.G---------.TCTCGGGTTC....T

Syn/Repl    SSSSSSSSSSRSSSRRSRSSSRSRRRRRRRRRSSSRRRSSSSSRRRRS
Variant #1            N   VQ S   D             PPL
Variant #2            I   GR C   N             SSV
```

C. 3' flanking (out of 434 sites in vir_*hsp70a* or vir_*hsp70b*)

Note that sequence for *D. lummei* was determined only through position 128, and that for *D. virilis hsp70a* only through position 274.

```
            0000000000011111111111111111111111111111111122222222222222222224
            022333457880000001111122567888888888999999900000000111111117783
  Position: 712089591294567891235689602013456789012578901234789012345674594
vir_hsp70a  GGACCAGATCCATTATTTCAAGGAACGCG--------CTGCT-------------AGCCA
vir_hsp70b  ............................................................TCA
vir_hsp70c  .TTA.........................TGTTGATATATACCACTAA.......TATA..T.
vir_hsp70e  ...............................................................
vir_hsp70f  .TT.................G...TGTGAGTATGATATATACCACTAATATATTC......T.
vir_hsp70g  .......................
lum_hsp70a  ATT.A...AG.-----.A----.
lum_hsp70b  .TT.ATAGC......--T--.T.
lum_hsp70d  .TT.ATAGC.T.......TT...
```

that *D. lummei* is recently (<5 MYA) derived from *D. virilis* or a common ancestor (see references cited in Introduction). Second, unlike many species pairs, these can interbreed to yield viable and partially fertile hybrids. Indeed, this partial fertility and multiple markers on each chromosome enable inter-specific genetics (Garbuz et al. 2002, 2003). Third, transposable elements present in each but not both species (Zelentsova et al. 1999) permit unambiguous identification of the parentage of any homeolog in hybrids.

According to diverse genetic and biochemical evidence, the inducible Hsp70 proteins play key roles in inducible tolerance of extreme temperatures, and both the magnitude and threshold of Hsp70 expression are correlated with the natural thermal regime in many species in many taxa (Feder and Hofmann 1999). Indeed, as we have shown (Garbuz et al. 2002, 2003), the low-latitude species *D. virilis* has both greater thermotolerance and greater Hsp70 levels than the high-latitude species *D. lummei*; in *D. virilis* populations, thermotolerance is usually correlated with Hsp70 level; and *D. virilis* × *D. lummei* hybrids are intermediate to the parental species in these respects. The coding and flanking sequence for the *hsp70* genes, in those rare cases where it differs within and among species, provides little explanation for the inter-specific differences in thermotolerance and Hsp70 protein levels (Table 1). By contrast, as noted, *hsp70* copy number differs dramatically. Although evolutionary loss or duplication of *hsp70* genes and loci is not the only mechanism of achieving such evolutionary variation in Hsp70 level and thermotolerance, it is an important mechanism that has occurred repeatedly in evolution. The present consensus is that a single locus bearing two copies in inverted orientation is the primitive state of *hsp70* in Diptera (see Introduction). Most likely *hsp70a* and *hsp70b* are orthologs of these two copies. Because *D. pseudoobscura*, intermediate in derivation to the *virilis* group and other *Drosophila*, also has a single locus bearing two copies in inverted orientation (B. Bettencourt, pers. commun.), parsimony suggests that an ancestor of the *virilis* group (rather than a common ancestor of all *Drosophila*) gained additional copies by tandem duplications. *D. virilis* strain A11 is from Southeast Asia, where the species probably originated (Throckmorton 1982), and has at most three *hsp70* copies. By contrast, all other strains examined have four to seven copies, suggesting that tandem duplication of *hsp70* accompanied the geographic range expansion of this species. The maximal copy number (seven), however, is in a strain (160) with multiple marker mutations and numerous transposable elements (Evgen'ev et al. 1997), and may thus be unrepresentative of natural variation. In *D. virilis*, these tandem duplicates are approximately 4.8 kb apart from one another and the ancestral pair member *hsp70b* (Fig. 4a), whereas the first tandem duplicate in *D. lummei* (*hsp70c*) is approximately 5 kb from *hsp70b*. We suggest, therefore, that one or more of the internal tandem duplicates were lost during or after the divergence of *D. virilis* and *D. lummei* to result in the lesser copy number but greater *hsp70b*–*hsp70c* spacing in *D. lummei*. Indeed, *hsp70c* in *D. lummei* is a putative pseudogene. Additional support for this scenario is that the sequences of both *hsp70g* in *D. virilis* and *hsp70d* in *D. lummei* diverge from that for the other *D. virilis* genes 202 bp 3′ of the end of the coding sequence. In any event, *D. virilis* clearly represents a remarkable example of gene proliferation by rampant gene duplication. *D. melanogaster* has evolved comparable *hsp70* copy numbers, but apparently by a dissimilar route: duplication of the primitive *hsp70* gene cluster and tandem duplication of one gene in the derived gene cluster, with the primitive gene cluster retaining its original copy number (Leigh-Brown and Ish-Horowicz 1981; Bettencourt and Feder 2001; Maside et al. 2002). Evidently in *Drosophila* the *hsp70* copy number has been remarkably malleable during evolution (Fig. 5).

One possible explanation for this malleability is the abundance of transposable elements in the *Drosophila* genome; these facilitate retrotransposition, ectopic recombination, and other modes of gene duplication or

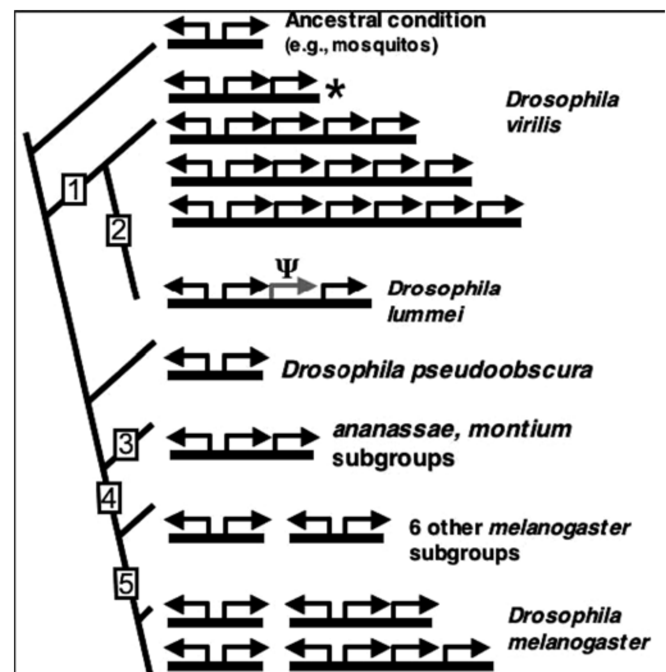

Fig. 5 Mapping of *hsp70* clusters and their arrangement onto a phylogeny of selected *Drosophila* and other Diptera (Ashburner 1989; Bettencourt and Feder 2001). A *heavy bar* underlies each *hsp70* cluster. Five events are inferred by parsimony or from the cited references: (1) tandem duplication of a member of the ancestral inverted pair (present study); (2) loss of one or more tandem duplicates (present study); (3) tandem duplication of a member of the ancestral inverted pair (Konstantopoulou et al. 1998); (4) duplication of the ancestral inverted pair via retroposition (Leigh-Brown and Ish-Horowicz 1981; Bettencourt and Feder 2001; Maside et al. 2002); (5) tandem duplication of a member of the duplicated ancestral pair (Leigh-Brown and Ish-Horowicz 1981; Bettencourt and Feder 2001; Maside et al. 2002). The cluster marked by an *asterisk* contains at most three copies, and is for *D. virilis* recently collected from Southeast Asia. Ψ in *D. lummei* indicates a pseudogene lacking 3′ sequence of *hsp70*. Note that characterizations of *hsp70* clusters for subgroups are based on one or a few species in each subgroup

rearrangement (Bushman 2002). Moreover, the *hsp70* genes may be especially susceptible to the insertion of transposable elements owing to their constitutively open chromatin conformation (Zatsepina et al. 2001; Lerman et al. 2003). Indeed, not only do the sequences flanking the *D. virilis hsp70* genes bear micro-satellites and PDV sequences (Zelentsova et al. 1986), the 3′ flanking sequence includes a portion of an ancient, now non-functional LTR-containing retrotransposon, JEM (J. Vieira, pers. commun.). These sequences collectively are abundant throughout the *virilis* group genomes (Biessmann et al. 2000; Evgen'ev et al. 2000b), and may explain the hybridization of *hsp70*-containing lambda clones to diverse chromosomal loci.

Once having duplicated, the *hsp70* genes have remained remarkably similar in both *D. virilis* and *D. lummei*, and in both coding and flanking sequence. This similarity is not due to inadvertent multiple cloning of the same gene because the various genes in both *D. virilis* and *D. lummei* are distinguishable by restriction sites, fragment sizes, and 3′ flanking regions (Fig. 3). Potential explanations for this similarity include strong positive selection on the *hsp70* loci, selective sweeps of linked loci, and concerted evolution via gene conversion. Because we have sequence for only a single line or population of each species, we cannot differentiate among these explanations directly. Nonetheless, as in *D. melanogaster*, the *hsp70a* and *hsp70b* genes in each species have likely been distinct for >100 MY (Bettencourt and Feder 2002). Despite this time, these paralogs are identical in both coding and flanking sequence in *D. virilis*, even at silent sites. We therefore hypothesize that, as in *D. melanogaster* and other members of its subgroup (Bettencourt and Feder 2002), gene conversion has been especially effective in homogenizing this sequence. The discovery of the ancient transposable element JEM in the 3′ flanking region of most *hsp70* copies of *D. virilis* and *D. lummei* favors this hypothesis because transposable elements have been implicated in the concerted evolution of tandemly repetitious DNA (Thompson-Stewart et al. 1994). Interestingly, the *Adh* gene in three *virilis* group species, as Nurminsky et al. (1996) describe, has many parallels with the present study, including multiple tandem duplications, repeated elements in flanking sequence, and extensive similarity of both coding and flanking sequence between the ancestral and duplicated genes (Nurminsky et al. 1996). As do we, they attribute this similarity to extensive gene conversion.

Acknowledgements In Russia, research was supported by Russian Grants for Basic Science 0304-48-918 and 0204-49121. In Chicago, research was supported by an International Supplement to NSF grant IBN99-86158 and a Howard Hughes Medical Institute Predoctoral Fellowship. We thank Dr. G.T. Lyozin for technical assistance, Dr. Anneli Hoikkala for providing *D. virilis* strains, and Dr. B.R. Bettencourt for help in interpreting *Drosophila pseudoobscura* genomic data and comments on the manuscript.

References

Ashburner M (1989) *Drosophila*. Cold Spring Harbor Laboratory, Cold Spring Harbor

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

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 (2002) Rapid concerted evolution via gene conversion at the *Drosophila hsp70* genes. J Mol Evol 54:569–586

Biessmann H, Zurovcova M, Yao J, Lozovskaya E, Walter M (2000) A telomeric satellite in *Drosophila virilis* and its sibling species. Chromosoma 109:372–380

Bushman F (2002) Lateral DNA transfer: mechanisms and consequences. Cold Spring Harbor Laboratory, Cold Spring Harbor

Elena SF, Lenski RE (2003) Evolution experiments with micro-organisms: the dynamics and genetic bases of adaptation. Nat Rev Genet 4:457–469

Evgen'ev MB, Kolchinski A, Levin A, Preobrazhenskaya AL, Sarkisova E (1978) Heat-shock DNA homology in distantly related species of *Drosophila*. Chromosoma 68:357–365

Evgen'ev MB, Yenikolopov GN, Peunova NI, Ilyin YV (1982) Transposition of mobile genetic elements in interspecific hybrids of *Drosophila*. Chromosoma 85:375–386

Evgenev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, Corces VG (1997) Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*.Proc Natl Acad Sci USA 94:196-201

Evgen'ev M, Zelentsova H, Mnjoian L, Poluectova H, Kidwell MG (2000a) Invasion of *Drosophila virilis* by the Penelope transposable element. Chromosoma 109:350–357

Evgen'ev MB, Zelentsova H, Poluectova H, Lyozin GT, Veleikod-vorskaja V, Pyatkov KI, Zhivotovsky LA, Kidwell MG (2000b) Mobile elements and chromosomal evolution in the *virilis* group of *Drosophila*. Proc Natl Acad Sci USA 97:11337–11342

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

Garbuz DG, Molodtsov VB, Velikodvorskaia VV, Evgen'ev MB, Zatsepina OG (2002) Evolution of the response to heat shock in genus *Drosophila*. Russ J Genet 38:925–936

Garbuz D, Evgen'ev MB, Feder ME, Zatsepina OG (2003) Evolution of thermotolerance and the heat-shock response: evidence from inter/intraspecific comparison and interspecific hybridization in the *virilis* species group of *Drosophila*. I. Thermal phenotype. J Exp Biol 206:2399–2408

Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. Science 287:308–309

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

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

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

Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME (2003) Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. Mol Biol Evol 20:135–144

Li W-H (1997) Molecular evolution. Sinauer, Sunderland

Lim JK (1993) In situ hybridization with biotinylated DNA. Dros Inf Serv 72:73–77

Lumme J (1982) The genetic basis of the photoperiodic timing of the onset of winter dormancy in *Drosophila littoralis*. Acta Univ Oulu 129:1–42

Maside X, Bartolome C, Charlesworth B (2002) S-element insertions are associated with the evolution of the *hsp70* genes in *Drosophila melanogaster*. Curr Biol 12:1686–1691

McGarry TJ, Lindquist S (1985) The preferential translation of *Drosophila hsp70* mRNA requires sequences in the untranslated leader. Cell 42:903–911

Mitrofanov VG, Blanter TV (1975) A study of temperature-sensitive mutations in the virilis group of *Drosophila*. Ontogenesis 6:386–391

Nurminsky DI, Moriyama EN, Lozovskaya ER, Hartl DL (1996) Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the *alcohol dehydrogenase* gene. Mol Biol Evol 13:132–149

Patterson JT, Stone WS (1952) Evolution in the genus *Drosophila*. Macmillan, New York

Peters FP, Lubsen NH, Sondermeijer PJ (1980) Rapid sequence divergence in a heat shock locus of *Drosophila*. Chromosoma 81:271–280

Spicer GS (1991) Molecular evolution and phylogeny of the *Drosophila virilis* species group as inferred by 2-dimensional electrophoresis. J Mol Evol 33:379–394

Spicer GS (1992) Reevaluation of the phylogeny of the *Drosophila virilis* species group (Diptera, Drosophilidae). Ann Entomol Soc Am 85:11–25

Spicer GS, Bell CD (2002) Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA genes. Ann Entomol Soc Am 95:156–161

Thompson-Stewart D, Karpen G, Spradling A (1994) A transposable element can drive the concerted evolution of tandemly repetitious DNA. Proc Natl Acad Sci USA 91:9042–9046

Throckmorton LH (1982) The *virilis* species group. In: Ashburner M, Carson HL, Thompson JN (eds) The genetics and biology of drosophila, vol 3d. Academic, London, pp 227–296

Zatsepina OG et al (2001) A *Drosophila melanogaster* strain from sub-equatorial Africa has exceptional thermotolerance but decreased Hsp70 expression. J Exp Biol 204:1869–1881

Zelentsova ES, Vashakidze RP, Krayev AS, Evgen'ev MB (1986) Dispersed repeats in *Drosophila virilis*: elements mobilized by interspecific hybridization. Chromosoma 93:469–476

Zelentsova H et al (1999) Distribution and evolution of mobile elements in the *virilis* species group of *Drosophila*. Chromosoma 108:443–456