Cell Stress and Chaperones (2010) 15:271–280
DOI 10.1007/s12192-009-0140-y

ORIGINAL PAPER

Characterization of the *shsp* genes in *Drosophila buzzatii* and association between the frequency of Valine mutations in *hsp23* and climatic variables along a longitudinal gradient in Australia

Jane Frydenberg · J. Stuart F. Barker · Volker Loeschcke

Received: 15 June 2009 / Revised: 20 August 2009 / Accepted: 20 August 2009 / Published online: 6 October 2009

Abstract The small heat shock gene (*shsp*) cluster of *Drosophila buzzatii* was sequenced and the gene order and DNA sequence were compared with those of the *shsps* in *Drosophila melanogaster*. The *D. buzzatii shsp* cluster contains an inversion and a duplication of *hsp26*. A phylogenetic tree was constructed based on *hsp26* genes from several *Drosophila* species of the *Sophophora* and *Drosophila* subgenera. The tree shows first a separation of the *Sophophora* and the *Drosophila* subgenera and then the *Drosophila* subgenus is divided into the Hawaiian *Drosophila* and the *repleta/virilis* groups. Only the latter contain a duplicated *hsp26*. Comparing the gene organisation of the *shsp* cluster shows that all the *Drosophila* subgenus species contain the inversion. Putative heat shock elements (HSE) were found in the promoters of all the *shsp* and putative regulator elements for tissue specific expression were found in the promoter of *hsp23*, *hsp27* and one of the *hsp26* genes. *hsp23* was found to be polymorphic for four non-synonymous changes that all lead to exchange of a Valine. The duplicated *hsp26* gene in *D. buzzatii* (*phsp26)* was polymorphic for two non-synonymous changes. The allele frequencies of these variants were determined in nine *D. buzzatii* populations covering most of its distribution in Australia using high-resolution melting curves. The allele frequencies of one of the *hsp23* variants showed a significant linear regression with longitude and the pooled frequency of the four Valine changes of *hsp23* in the nine populations showed a significant linear regression with longitude and with a composite measure of climatic variables.

Keywords Heat shock protein · Gene duplication · Inversion · Conserved noncoding sequences · Geographical clines

Abbreviations

shsp	small heat shock protein gene
HSE	heat shock element
EcRE	ecdysone response element
CTAB	hexadecyltrimethylammonium bromide
CNS	conserved non-coding sequence

Introduction

The cactophilic species *Drosophila buzzatii* has a wide geographical distribution, and it breeds and feeds on decaying tissues of *Opuntia* cacti. *D. buzzatii* originates from South America, from where it has colonised and spread to the Old World and Australia. *D. buzzatii* belongs to the mulleri complex of the *repleta* group within the subgenus *Drosophila* and due to its rather broad geographic distribution, it has been a model for evolutionary adaptation to different environmental conditions. Several studies on

Electronic supplementary material The online version of this article (doi:10.1007/s12192-009-0140-y) contains supplementary material, which is available to authorized users.

J. Frydenberg (✉) · V. Loeschcke
Department of Biological Sciences, Ecology and Genetics, University of Aarhus,
Ny Munkegade, Bldg. 1540,
8000 Aarhus C, Denmark
e-mail: jane.frydenberg@biology.au.dk

J. S. F. Barker
School of Biological Sciences, University of Queensland,
St. Lucia, Brisbane, Queensland 4072, Australia

this species have described latitudinal and altitudinal variation in different traits (longevity, fecundity, senescence, stress resistance and chromosome polymorphism; Hasson et al. 1985; Sorensen et al. 2005; Norry et al. 2006) and genetic variation in allozyme markers (Barker et al. 1985) and microsatellites (Frydenberg et al. 2002; Barker et al. 2009). However, variation of candidate genes for traits associated with climatic adaptation has only rarely been studied. In *Drosophila melanogaster*, allele frequencies in candidate genes have been successfully correlated to climatic variables and latitude, including *shsps* (Frydenberg et al. 2003), *hsr-omega* (Collinge et al. 2008) and *Adh* (Fry et al. 2008). Heat shock genes are highly conserved, which makes these genes also potential candidate genes in other *Drosophila* species, especially in species such as *D. buzzatii*, where several life stages occur inside rotting cladodes that are often exposed to the sun.

In *D. melanogaster*, there are seven *shsp* genes located closely in a linked cluster on chromosome three in the order *hsp27*, *hsp23*, *hsp67Ba*, *hsp26*, *hsp22*, *hsp67Bb* and *hsp67Bc* (Ayme and Tissières 1985). The sHsps have many different functions, both under stress and under non-stress conditions (Mestril et al. 1986; Michaud et al. 1997; Michaud and Tanguay 2003). Under stress conditions they have chaperone functions, protect the cytoskeleton and modulate the apoptotic process. The sHsps have a cell-specific pattern of expression, likely reflecting different chaperone activities (Morrow et al. 2006). In the absence of stress, the cell-specific expression of the *shsps* is tightly regulated, suggesting specific functions during development.

In *D. melanogaster*, the regulation of each of the *shsps* has been thoroughly investigated, either using promoter deletions to locate important regulator regions (Cohen and Meselson 1985; Riddihough and Pelham 1986; Frank et al. 1992; Michaud et al. 1997), or regulation by known tissue specific transcription factors (Michaud and Tanguay 2003) or search for consensus transcription factor binding sites. The expression after heat shock is regulated by the heat shock factor (HSF) binding to the heat shock element (HSE) and HSE have been found in the promoter of most *hsp* genes (Mestril et al. 1986; Sandaltzopoulos et al. 1995; Riddihough and Pelham 1986).

In order to characterise the *shsp* genes in *D. buzzatti* and analyse their genetic variation, we have obtained sequences of the gene region of *D. buzzatii* which contains the genes coding for the sHsps. The gene organisation in this region is compared with the homologous region of *D. melanogaster* and an inversion in the region is described. Comparative genomic regions of the *D. buzzatii* and the *D. melanogaster* sequences are used to define phylogenetic footprinting that may identify the tissue-specific transcription factor binding sites and the HSEs.

The sequence variation of the *D. buzzatii hsp23*, *phsp26* and *hsp26* is analysed. The allele frequencies of four non-synonymous SNP in *hsp23* are determined in nine Australian populations using a high resolution melting (HRM) method and the variation is assessed for association with geographical location and climatic variables.

Material and methods

DNA extraction and genomic walking

Genomic DNA was isolated from *D. buzzatii* collected near Chumbicha in Argentina (Sorensen et al. 2005) using the CTAB method (Doyle and Doyle 1987).

The sequences of the *shsps* from *D. melanogaster* were found in Fly Base and aligned. We designed two primers that are located in the α-crystallin domain, conshspF GTGGGCAAGGATGGCTTCCAG and conshspR TTGCTGGATCCGGACGATGCGCTC, in order to get the α-crystallin domain from *shsp* in *D. buzzatii*. A 300 bp fragment was amplified under standard polymerase chain reaction (PCR) conditions. The fragment was sequenced (Macrogen Genomics, Korea) and found to contain part of *hsp67Ba*. From this sequence two forward and two reverse specific primers were designed. Using respectively the two reverse primers and the two forward primers as specific primers in the DNA Walking Speedup™ kit (Seegene, USA), we walked out in the genome in both directions. This procedure was repeated and combined with primers designed from published sequences. The coding sequences of *D. melanogaster hsp23*, *hsp26* and *hsp27* were used for a search for related sequences in the *Drosophila mojavensis* gene library using the sequences available http://rana.lbl.gov/drosophila/. The *D. mojavensis* genome assembly was produced by Agencourt Bioscience Corporation. Identified *D. mojavensis hsp23*, *hsp27* and *hsp26* genes were then used to design PCR primers for *D. buzzatii* DNA. The fragment was isolated and sequenced (Macrogen Genomics, Korea) with the PCR primers and internal primers.

Molecular phylogenetic analysis

The chromosome regions between CG4461-RA and CG4080-RA from *D. melanogaster*, *Drosophila ananassae*, *Drosophila pseudoobscura*, *Drosophila willistoni*, *Drosophila grimshawi*, *Drosophila virilis* and *D. mojavensis* were available at http://www.ncbi.nlm.nih.gov/genome/guide/fly/ (Adams et al. 2000; Richards et al. 2005; Flybase Consortium 2002).

A phylogenetic tree of the *hsp26* genes was constructed using the MEGA version 3.1 (Kumar et al. 2004) by the

Minimum Evolution method with the distance calculated by the Tamura and Nei method (Tamura and Nei 1993). Bootstrap values were calculated from 1,000 replications.

Screening for homologous promoter sequences

The promoter region of the *shsp* genes from *D. melanogaster*, *D. pseudoobscura*, *D. virilis* and *D. mojavensis* were available at http://www.ncbi.nlm.nih.gov/genome/guide/fly/ (Adams et al. 2000; Flybase Consortium 2002).

Local blocks of homologous sequences were found with the use of Chaos/DiAlign (Brudno et al. 2004; http://dialign.gobics.de/chaos-dialign-submission), DNA block aligner (DBA; Birney and Durbin 1997; http://www.ebi.ac.uk/Wise2/dbaform.html) and the VISTA programme (Mayor et al. 2000; http://genome.lbl.gov/vista/index.shtml) with a 10 bp 70% window.

DNA sequence variation analysis

The DNA sequence variation of the coding region of *hsp26*, *phsp26* and *hsp23* was examined in *D. buzzatii* by sequencing *hsp26*, *phsp26* and *hsp23* in a total of 40 flies from four different populations (Cafayate, Trancas, Chumbicha and Santiago de Estero) from Argentina (Sorensen et al. 2005). *hsp26* was amplified by primers Dbuz26vaF (ATTCATCAAGCGAACAACTTC) and DbuzvaR (TGGAATGAGGATGACACAGG) and DbuzvaR was used as sequence primer. *phsp26* was amplified using primers Dbp26pro (CAAAGAACTTCGCTGAGCGA) and Dbp26RS (GGAATGCGATTGACTTTGGTT). The PCR products were sequenced directly using the primer Dbp26pro1 (AAAACTACATAACCTTAAATC) (Macrogen Genomics, Korea). *hsp23* was amplified by the primers Db23F1 (TACTGTTTGAAAAAACCAAG) and Db23R (TAGATATTGACTTCAATAATT). The PCR products were sequenced by Db23R (Macrogen Genomics, Korea).

Population studies

The allele frequencies of non-synonymous mutations in *hsp26*, *phsp26* and *hsp23* were analysed in nine populations (30 individuals from each) from Australia (Fig. 1), Mulambin Beach (Mub; 23°19′ S, 150°78′ E, 0 km distance from coast (further on called dist)), Isla Gorge (Isg; 25°25′ S, 149°94′ E, 230 km dist), Grandchester (Grd; 27°68′ S, 152°45′ E, 75 km dist), Hickey Island (His; 29°43′ S, 153°36′ E, 0 km dist), Baradine (Bar; 30°95′ S, 149°07′ E, 345 km dist), Tambar Springs (Tam; 31°31′ S, 149°94′ E, 250 km dist), Gerongar Point (Ger; 34°56′ S, 150°81′ E, 0 km dist), Maldon (Mal; 36°98′ S, 143°77′ E, 125 km dist), Bulla (Bul; 37°65′ S, 144°77′ E, 25 km dist; Barker et al. 2009).

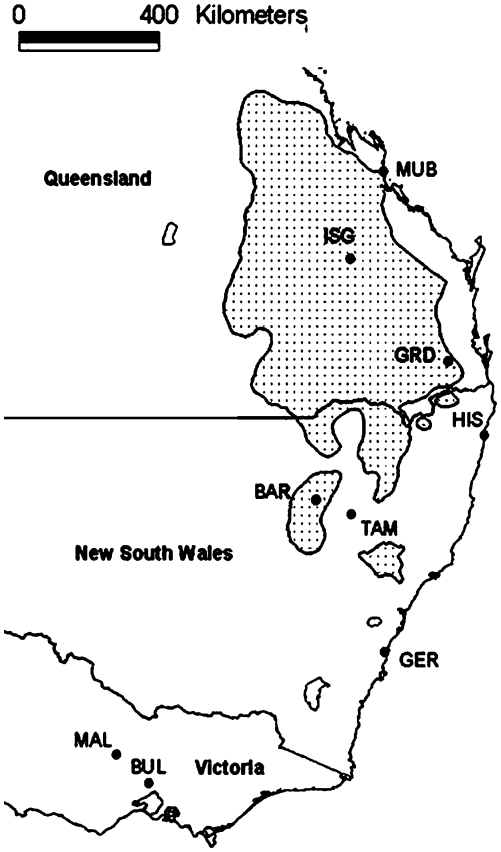

Fig. 1 Map of eastern Australia showing the locations where *D. buzzatii* were collected. The *hatched area* shows the main distribution of *Opuntia* in 1920 (modified from Barker et al. 2009)

High resolution melting analysis and scoring of indels

A 100-bp PCR fragment including both SNP601 and SNP610 in *phsp26* was amplified with the primers Dbp26HRM1 (TGGTTCAAATACAGCAAGTG) and Dbp26R (CTACTTGGCGTTGTTTCCATT) using the LightCycler 480 High Resolution Melting Master kit (Roche). The reactions were run in 10-μl reaction volume with 3.0 mM $MgCl_2$ and annealing temperature 55°C. A 146-bp fragment including SNP58 and SNP127 in *hsp23* was amplified by the primers Db23F (ATGGCAAATCTTC CACTTC) and Db23HRMR (CAGCTGCTGGTCAAGAG GAA) and a 99-bp fragment including SNP166 and SNP184 was amplified with the primers Db23HRMR1 (ATGCA C A C C T G G A A T C C G T C) a n d D b 2 3 H R M F (TTCCTCTTGACCAGCAGCTG). Both the *hsp23* fragments were amplified under the same conditions as the *phsp26* fragment. The genotypes which were known from the sequenced *hsp23* and *phsp26* genes were used as standard in the melting curves. Within each group of melting curves four flies were sequenced to see if the melting curve groupings correlate to the expected sequences.

The HRM curves were analysed using the LC480 software. In some cases the melting curves of normal homozygote and a mutant homozygote were identical, so in order to identify mutant homozygotes, the homozygote samples were mixed with a sample known to be a normal homozygote. The mutant homozygote samples give then a melting curve as heterozygote.

The deletion of Val18 and insertion of Asp-Lys in *hsp26* were scored as microsatellites using primers located around the indels. The PCR products were analysed on an ALF express DNA sequencer together with internal and external size standards.

Statistical analysis

Nucleotide diversities were calculated using DNASP version 4 (Rozas et al. 2003). The associations between allele frequencies and geographical location and climatic variables were analysed using JMP v.7 (2007). Geographical location was defined using latitude, longitude and distance to the coast. Distance to the coast was included because climatological zones in eastern Australia tend to run parallel to the coast (Barker et al. 2005). Principal component analysis has been applied to 35 climatic variables for each locality estimated using the BIOCLIM programme of the ANUCLIM 5.1 package (Barker et al. 2005; Houlder et al. 2000). The climate was characterised by the scores for the first four principal components (PC1, PC2, PC3 and PC4), which account for 93% of the climatic variation. Of this, PC1 accounts for 43.0% and PC2 for 31.8% (Barker et al. 2005).

Results

Chromosomal organisation of the *shsps* in *D. buzzatii*

The structure and the sequences of the *shsps* in *D. buzzatii* were determined by a combination of DNA walking and PCR derived from amplifying genomic DNA with primers designed from *D. mojavensis* sequences. A nucleotide sequence of 12,545 bp was determined (Genbank accession no. EU196734). By comparison of the sequence with those of the *shsp* genes in *D. melanogaster* and using the same nomenclature for the homologous genes, we found that the region contains *hsp67Ba*, *hsp23*, *hsp27* and *hsp26* in two copies: we call one *phsp26* and the other *hsp26* (Fig. 2). In *D. melanogaster hsp26*, *hsp67Ba*, *hsp23* and *hsp27* are located between the genes CG4461-RA and CG4080-RA with *hsp27* located closest to CG4080-RA (Fig. 2). In *D. buzzatii hsp27* is located closest to CG4461-RA and the *hsp26* gene is located between *hsp27* and CG4461-RA. This suggests that the region containing *phsp26*, *hsp67Ba*, *hsp23* and *hsp27* is inverted in the *D. buzzatii* genome.

The most likely location of the *shsp* locus on the *D. buzzatii* chromosomes was found by comparing the locations of mapped *D. buzzatii* RAPD fragments from Laayouni et al. (2000) and *D. buzzatii* microsatellites (Barker et al. 2009) with the corresponding location in the *D. mojavensis* genome scaffolds. The *D. mojavensis shsp* region which is homologous to the *D. buzzatii shsp* region is located around *D. mojavensis* scaffold 6680: 7.190.000, whereas the *D. buzzatii* RAPD 70.09.6sts is located at *D. mojavensis* scaffold 6680: 7.007.768 and the *D. buzzatii* microsatellite M142 (now Db142) is located at *D. mojavensis* scaffold 6680: 7.131.400 (Frydenberg et al. 2002). As the *D. buzzatii* RAPD 70.09.6sts and Db142 are located on *D. buzzatii* chromosome 4 outside inversion 4s this is probably also the case for the *D. buzzatii shsp* locus (Laayouni et al. 2000; Barker et al. 2009).

Drosophila species are classified into the *Sophophora* subgenus and the *Drosophila* subgenus. The *Sophophora* subgenus is then divided into the *willistoni*, the *obscura* and the *melanogaster* species groups, whereas the *Drosophila* subgenus is divided into the Hawaiian *Drosophila*, the *repleta* group and the *virilis* group. In order to date the inversion and the duplication, we analysed the gene order of the homologous genes of CG4080-RA, the *shsps* and CG4461-RA in seven *Drosophila* species: *D. melanogaster* and *D. ananassae* from the *melanogaster* group, *D. pseudoobscura* from the *obscura* group, *D. willistoni* from the *willistoni* group, *D. grimshawi* from the Hawaiian *Drosophila*, *D. virilis* from the *virilis* group and *D. mojavensis* from the *repleta* group, using the Flybase consortium database. The results show that all the *Sophophora* species and *D. grimshawi* contain one copy of *hsp26*, while *D. virilis* and *D. mojavensis* both contain two *hsp26* genes as in *D. buzzatii* and that all species from the *Drosophila* group have the inversion. A molecular phylogenetic tree of the *hsp26* genes was constructed by the ME method (Fig. 3). The tree is identical to the tree described above constructed from the sequence of other *Drosophila* genes (Drosophila 12 Genomes Consortium 2007).

Identification of putative response element by comparative sequence analysis

In order to localise putative response elements in the promoter region of the *shsp* genes of *D. buzzatii*, we analysed the conserved sequences in the promoter region between the *shsp* genes of *D. buzzatii* and *D. melanogaster* in more detail using three different alignment methods: DiAlign, DBA and VISTA.

We found two possible HSEs and a TATA-box conserved in the intergenic region between *hsp67Ba*

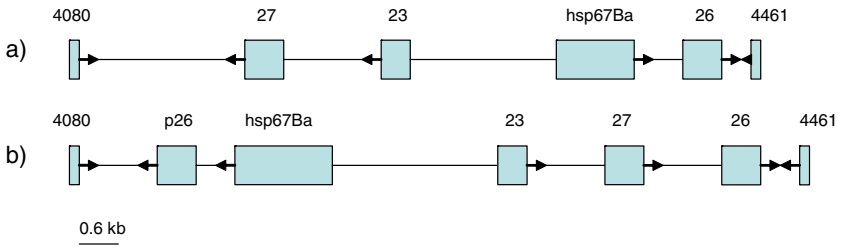

Fig. 2 Chromosomal organisation in *D. melanogaster* (**a**) and *D. buzzatii* (**b**) of GC4080-RA (4080), *phsp26* (p26), *hsp67Ba*, *hsp23* (23), *hsp27* (27), *hsp26* (26) and CG4461-RA (4461). The direction of transcription is indicated by *arrows*

and *hsp23* (Supplemental data Table S1). Two regions are overlapping with regions earlier found to be important for hormone induction of *hsp23* (Mestril et al. 1986; Rogulski and Cartwright 1995; Dubrovsky et al. 2001) and three regions contain a proneural transcription factor binding site CAGCTG sequence (E-box) (Powell et al. 2004).

The *D. buzzatii phsp26* promoter and *D. melanogaster hsp26* promoter homologous regions contain two regions located in the region important for ovarian expression. One region represents an alignment of the *D. melanogaster* distal GAGA element to a possible *D. buzzatii* GAGA element and two regions contain a possible HSE and a TATA-box, respectively (Supplemental data Table S2).

Searching for homologous regions between the *D. buzzatii hsp26* and *D. melanogaster hsp26* promoters only reveals few homologous regions (Supplemental data Table S3). One region is a possible TATA-box, one is a possible *D. buzzatii* HSE and one is an alignment of the first part of proximal *D. melanogaster* GAGA elements to a possible proximal *D. buzzatii* GAGA element.

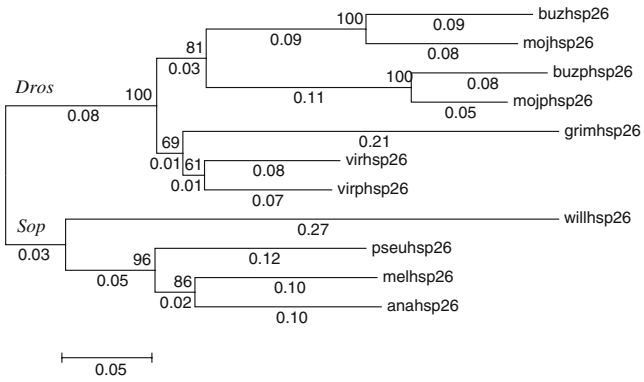

Fig. 3 Phylogenetic tree of *hsp26* and *phsp26* genes (when present) from *D. buzzatii* (*buz*), *D. mojavensis* (*moj*), *D. virilis* (*vir*), *D. grimshawi* (*grim*), *D. willistoni* (*will*), *D. pseudoobscura* (*pseu*), *D. ananassae* (*ana*) and *D. melanogaster* (*mel*). The *Sophophora* (*Sop*) and *Drosophila* (*Dros*) subgenera are indicated. The tree was constructed using MEGA and Minimum Evolution method. The bootstrap values (percent of 1,000 replicates) are indicated at branch points and the Tamura and Nei (1993) genetic distances are given on the branches

When the *D. melanogaster* and *D. buzzatii hsp27* promoter regions were aligned two regions known to be involved in the ecdysone induction of *hsp27* were found. One of them contains a possible *D. buzzatii* ecdysone response element (EcRE). Two regions contain possible HSE and one contains a TATA-box (Supplemental data Table S4).

Intra-specific sequence variation in *hsp23*, *hsp26* and *phsp26*

Polymorphic sites in the coding region of *hsp23*, *hsp26* and *phsp26* were identified by sequencing the genes in 40 flies from four different *D. buzzatii* populations from Argentina. In *hsp23* we found 15 synonymous sites and four non-synonymous sites. Three of the non-synonymous mutations are changing of a Valine. SNP58 was changing Val20 to Met20, SNP127 Val43 to Ileu43 and SNP166 Val56 to Ileu56. Val20 and Val56 are conserved between the *D. melanogaster* Hsp23 and the *D. buzzatii* Hsp23. The last non-synonymous mutation SNP153 is a change of GluN51 to His51 and was found only in one individual. In *hsp26* there were 22 synonymous sites and two indels, a deletion of GCT at 52–54 that leads to deletion of Val18 and insertion of GATAAG at 606 that leads to an Asp-Lys repeat. In *phsp26* there were 44 synonymous sites and three non-synonymous sites. The three non-synonymous mutations in *phsp26* were SNP601, SNP607 and SNP610. SNP601 was changing Pro101 to Ser101, SNP607 Thr103 to Ala103 and SNP610 Thr104 to Ala104.

The levels of nucleotide heterozygosity of *hsp23*, *hsp26* and *phsp26* were estimated by means of π and θ and compared to π and θ values of other *D. buzzatii* genes (Table 1; Gómez and Hasson 2003; Piccinali et al. 2004, 2007). The total π and θ values for *hsp23* and *hsp26* were lower than that for *Xdh*, *α-E5* and *Est-A*, whereas for *phsp26* it was in between the values for the other genes. But looking at the values for synonymous and non-synonymous variation separately, the differences between the *hsp* genes and other genes are more striking as the values of π_{ns} and θ_{ns} for all the *hsp* genes are low as compared with the values for other genes. The same is the case when the percentages of non-synonymous mutations are compared.

Table 1 Polymorphisms of *D. buzzatii shsp*, *Xdh*, *α-E5* and *Est-A*

	Length (bp)	syn	nsyn	% nsyn	Singl	Ind	π_T 10^3	θ_T 10^3	π_s 10^3	θ_s 10^3	π_{ns} 10^3	θ_{ns} 10^3
hsp23	558	15	4	21	7	0	3.9	6.5	11.4	21.9	1.6	1.8
hsp26	666	22	0	0	9	2	4.2	6.4	17.8	27.7	0	0
phsp26	636	44	3	7	21	0	6.4	14.5	27.3	60.0	0.3	1.2
Xdh (Piccinali et al 2004)	1,875	97	53	35	98	0	12.9	21.1	35.7	57.3	5.7	10.1
α-E5 (Piccinali et al 2007)	738	13	14	52	18	0	5.1	8.6	12.5	18.4	3.4	5.8
Est-A (Gómez and Hasson 2003)	1,061	69	65	46	69	0	22.5	30.0	49.0	64.7	14.8	19.6

syn synonymous mutations; *nsyn* non-synonymous mutations; *Singl* singletons; *Ind* indels, π_T, π_s and π_{ns} average number of pairwise differences per site in total and synonymous and non-synonymous sites, respectively; θ_T, θ_s and θ_{ns} Watterson's (1975) nucleotide diversity estimate in total and for synonymous and non-synonymous mutations, respectively

Association between variation in the small heat shock genes and environment variables

The allele frequencies of the non-synonymous mutations in *hsp23* and the two non-synonymous mutations in *phsp26* were determined in nine populations from Australia (Fig. 1). The fragment which was used to determine the SNP166 in *hsp23* was found also to include a non-synonymous mutation at 184 in some of the Australian samples. SNP184 was a change from Val62 to Leu62. Val62 is conserved between *D. buzzatii* and *D. melanogaster*. SNP153 was not found in the Australian populations. The allele frequencies of the four other *hsp23* mutation (SNP58, SNP127, SNP166 and SNP184) and the three *phsp26* mutations (SNP601, SNP607 and SNP610) are shown in Table 2. The presence of the indels in *hsp26* was determined in the same populations using fragment size variation. The two indels in *hsp26* were only found once in the Australia populations and therefore not analysed further.

The allele frequencies of the different SNP in *phsp26* and *hsp23* were compared to geographical location (longitude, latitude and distance from the coast) and the four principal components (PC) which account for 93% of the climate variability and of this PC1 accounts for 43%. None of the polymorphic SNP in *phsp26* show any effect of either geographical or climatic variables. When the *hsp23* mutations were analysed, SNP127 showed a linear association with longitude (R^2=0.40, P<0.038, Fig. 4a). Historical records and microsatellite variation (Barker et al. 2009) indicate that the southern populations represent a secondary colonisation from the north. If the two southern populations Bul and Mal were omitted, the SNP127 showed an even stronger linear association with longitude (R^2=0.93, P<0.0002, Fig. 4a). Since SNP58, SNP127, SNP166 and SNP184 all are changes of a Valine, we analysed the total frequencies of these four SNP (hsp23Val). hsp23Val showed a linear association with longitude (R^2=0.66, P<0.0048, Fig. 4b) and with the climatic PC1 (R^2=0.87, P<0.0002, Fig. 4c). If the two southern population are removed the fit is increased with longitude (R^2=0.84, P<0.002, Fig. 4b) but not with PC1 even though it still is significant (R^2=0.81, P<0.004, Fig. 4c). Finally, the frequencies of hsp23Val show a linear association with distance to coast (R^2=0.80, P<0.004, Fig. 4d).

Table 2 Allele frequencies of SNP in *hsp23* and *phsp26* in Australian populations of *D. buzzatii*

	hsp23 SNP58	SNP127	SNP166	SNP184	hsp23Val	*phsp26* SNP601	SNP610
Mal	0	0.02	0	0	0.02	0.08	0
Bul	0.03	0.02	0	0	0.05	0	0.30
Bar	0	0	0.05	0	0.05	0	0
Isg	0	0	0.11	0	0.11	0.02	0
Tam	0.02	0.05	0.02	0	0.09	0.03	0
Mub	0.21	0.05	0	0	0.26	0.04	0.04
Ger	0	0.08	0.02	0.13	0.23	0.02	0.12
Grd	0.02	0.17	0.07	0	0.26	0.02	0
His	0.05	0.23	0.10	0	0.38	0	0

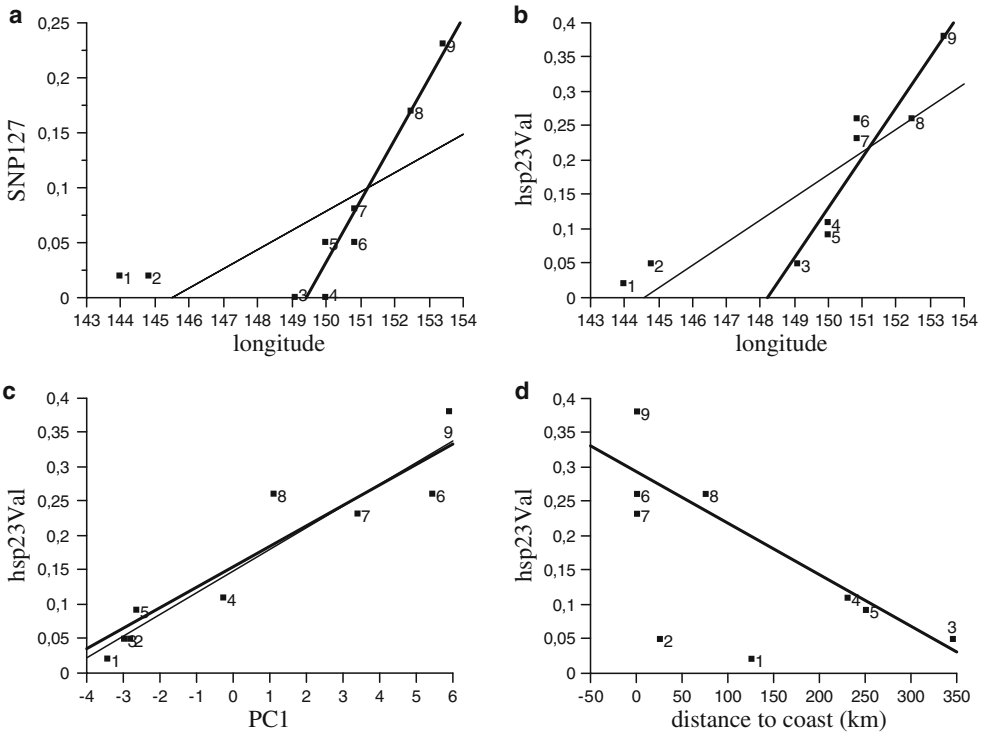

Fig. 4 Linear regression of the allele frequencies of SNP127 or the total frequencies of Valine changes (hsp23Val) in Australian population of *D. buzzatii* Mal (*1*), Bul (*2*), Bar (*3*), Isg (*4*), Tam (*5*), Mub (*6*), Ger (*7*), Grd (*8*) and His (9). **a** SNP127 with longitude, **b** hsp23Val with longitude, **c** hsp23Val with PC1 and **d** hsp23Val with distance to coast in kilometers. *Thin lines* are with all populations included and *thick lines* are where the two southern populations Mal and Bul are excluded

Discussion

Evolution of the *shsp* cluster

In this study we sequenced the chromosome region containing the *phsp26*, *hsp67Ba*, *hsp23*, *hsp27* and *hsp26* genes in *D. buzzatii,* a member of the *repleta* group from the *Drosophila* subgenus. The gene region was inverted as compared with species of the *Sophophora* subgenus, and the *hsp26* gene is duplicated in the *virilis* and the *repleta* groups. Some inversions arise by simple cut and paste events, others arise by more complicated events which may result in duplication of coding sequences at the break points (Matzkin et al. 2005; Sharakhov et al. 2006). Since *D. buzzatii phsp26* and *D. buzzatii hsp26* are located at the borders of the inversion, it seems likely that the *hsp26* duplication originated together with the inversion event, but there could be several explanations of the origin of the inversion/duplication events.

The first explanation includes three inversion events. In the molecular phylogenetic tree the two *D. virilis hsp26* genes are more closely related to each other than either is to the *D. buzzatii* and *D. mojavensis hsp26* genes, whereas the *D. buzzatii* and the *D. mojavensis hsp26* genes are clustered together and the *D. buzzatii* and the *D. mojavensis phsp26* genes are clustered together. This would date the inversion/duplication event after the separation of the *virilis* group and the *repleta* group and include two inversion events. But *D. grimshawi* also has the inversion, so the inversion could also have happened in this lineage but in this case without duplication of *hsp26*. The reoccurrence of inversion breakpoints is common so it is not unlikely that the three inversion events share breakpoints (González et al. 2007).

The second explanation could be that the inversion/duplication happened only once at the *Sophophora*/*Drosophila* separation. In this case *D. grimshawi* must have lost one of the *hsp26* copies. *D. grimshawi* has several rearrangements including the *shsp* loci, which include duplication of *hsp23* and part of *hsp27* and moving of the *hsp23-CG4461* region away from the *hsp67Ba gene*. Therefore, we consider a deletion of the *hsp26* gene in this species a possible event. It is more difficult to explain the clustering of the *D. virilis hsp26* genes in the *hsp26* tree. Here one explanation could be divergent selection on each of the two *hsp26* genes in the *repleta* group after the separation from the *virilis* group. This may also lead to the closer relationship between the *D. grimshawi hsp26* and the two *D. virilis hsp26* genes than to the *repleta* group *hsp26* genes.

A third explanation could be that the inversion happened at the *Sophophora/Drosophila* separation and the duplication of the *hsp26* genes happened twice, once in the *repleta* group and once the *virilis* group. However, we find this latter explanation unlikely due to the location of the duplicated *hsp26* genes at each end of the inversion.

Identification of putative regulator sequences

Transcription factor binding sites were localised in the *shsp* promoter regions on the basis of coincidence between

studies of important regulator regions and CNS found in alignments of the *D. buzzatii* and *D. melanogaster shsp* promoter sequences. The *hsp23* promoter contains three CNS with proneural transcription factor binding site E-boxes, each surrounded with a specific conserved sequence which may indicate that each E-box is specified for a specific proneural transcription factor which then regulates the expression of *hsp23* in different tissues (Powell et al. 2004). The cell-specific and temporary expression of *hsp23* in the imaginal discs has been shown to be regulated by proneural transcription factors (Alonso et al. 2005).

When the gene of a multifunctional protein is duplicated, the two copies sometimes get different functions, either by mutation in the coding region or by separation of regulator sequences (Force et al. 2005). The *D. melanogaster hsp26* is expressed in spermatocytes, nurse cells and the epithelium of larvae, pupae and adults. The sequences which regulate the tissue-specific expression and the heat shock expression of *D. melanogaster hsp26* have been shown to be located in different parts of the promoter. Three regions involved in spermatocyte-specific expression are identified downstream from −352, whereas the ovarian expression is dependent on the region (−522 to −352) (Cohen and Meselson 1985; Frank et al. 1992; Michaud et al. 1997).

In the comparison of the *D. buzzatii hsp26* and *phsp26* promoters to the *D. melanogaster hsp26* promoter, we find that both the *D. buzzatii* genes retain the regulatory sequences for heat shock expression. In the *D. buzzatii phsp26* there are two CNS located in the region important for ovarian expression, so *D. buzzatii phsp26* may have retained the ovary specific expression. On the other hand the *D. buzzatii hsp26* gene seems to have lost its tissue-specific expression as we found no CNS located in regions important for ovarian expression.

The *D. melanogaster hsp27* gene contains an EcRE but three regions are important for ecdysone induction during oogenesis and embryogenesis (Riddihough and Pelham 1986). The *D. buzzatii hsp27* seems to have retained both the heat shock and the tissue-specific expression as the promoter contains homology to both HSE and EcRE binding sites and has a CNS located in a region important for ecdysone induction.

Intra-specific sequence variation and its ecological association

There is a great difference between the percentages of non-synonymous replacements in *D. buzzatii hsp23*, *hsp26* and *phsp26* and all of them are lower than the percentages for *D. buzzatii Xdh*, *α-E5* and αE5 (Table 1; Gómez and Hasson 2003; Piccinali et al. 2004, 2007). But analyses of the DNA variation in different genes within and between *Drosophila* species have shown that there is great variation in the percentage of non-synonymous replacements both between genes within species and between species (Moriyama and Powell 1996; Andolfatto 2001; Piccinali et al. 2004). A low level of non-synonymous mutations is probably connected to high evolutionary constraints in the *hsp* genes. This is in agreement with the high level of conserved sequence in *hsp* genes in general. Both *hsp26* and *phsp26* have a low number of non-synonymous mutations which makes it likely that they both are active genes in *D. buzzatii.*

The frequency of SNP127 shows a linear association with longitude. If we leave out the two southern populations the strength of the association is increased. The total frequency of Valine changes in *hsp23* also shows a linear association with longitude, distance to the coast and the climatic variable PC1 when we leave out the two southern populations. A study of the population differentiation and colonisation of the Australian *D. buzzatii* populations has shown that the three most southern populations have lower genetic variability than the other populations and it was concluded that they probably arose by a secondary colonisation from the northern populations (Barker et al. 2009). Thus it is reasonable to delete these populations from our analysis but Ger, which is the most northern population of the three, has a non-synonymous mutation SNP184 that is exclusively found in this population and which as SNP58, SNP127 and SNP166 leads to a Valine change. Therefore we leave Ger in our Valine change analysis.

The changes in *hsp23* are changes of Valine to either Methonine (SNP58, Val20), Isoleucine (SNP127, Val43 and SNP166, Val56) or Leucine (SNP184, Val62), which are all hydrophobic amino acids. Valine is less hydrophobic than the three others amino acids, so all the mutations make Hsp23 more hydrophobic. Gottler et al. (2008) have shown that changing of the hydrophobicity of antimicrobial peptide protegrin-1 by Valine to Leucine changes the biological activity of the peptide. So it seems reasonable to claim that the total amount of Valine in Hsp23 could be connected to the function of the protein.

The three variables PC1, longitude and distance to the coast are all correlated but we have chosen to make a separate model for each of them to get a broader biological understanding of the result. SNP127 was significant only with longitude. PC1 and distance to the coast each have some distinctive characteristics. hsp23Val shows a significant association with PC1 both with and without the two southern populations, whereas it only shows a significant association with the distance to the coast when the two southern populations were deleted, so the climatic information seems important.

Much of the difference between the *D. buzzatii hsp26* and *phsp26* genes is located in the 3′end where *hsp26* has

a repeated sequence which codes for Asp-Lys-(Glu-Lys)$_4$. The insertion in *hsp26* leads to an increase of the repeat by Asp-Lys. The non-mutated *phsp26* does not have a repeat in the 3`end but SNP601 leads to a repeat of 6 bp since TCGAAGCCGAAG changes to TCGAAGTCGAAG which gives a repeat of Ser-Lys in the protein. We have earlier found *D. melanogaster hsp26* to be polymorphic for an insertion of 9 bp (GACGGCAAG, Asp-Gly-Lys) in the 3′end where 8 bp are part of a repeat. This variation showed a latitudinal cline in Australian *D. melanogaster* populations (Frydenberg et al. 2003). The *D. buzzatii hsp26* and *phsp26* repeats found in this study were very rare in the Australian populations and there seems to be no correlation with environmental variation in the frequencies of these SNP. But *D. buzzatii* and *D. melanogaster* are living in very different niches, so parallel latitudinal clines could not necessarily be expected. However, it is notable that all the *hsp26* genes are polymorphic for repeats in the 3′end and all the repeats involve a Lysine.

In summary, we show that the region which includes the *D. buzzatii shsp* genes is inverted as compared with the *D. melanogaster* genome structure and the *hsp26* gene is duplicated. The *D. buzzatii hsp23* has four SNP that all lead to a replacement of a Valine. Summation of the frequencies of these SNPs shows a linear association with longitude, distance to the coast and climatic variables in *D. buzzatii* populations from Australia. In future work, the frequency of these *hsp23* SNPs should be investigated in different parts of the world to look for similar associations with longitude, distance to the coast or climatic variables.

Acknowledgments The authors thank Jesper G. Sørensen for help catching some of the Australian fly samples and Fabian Norry for some of the Argentinean fly samples, Jürgen Tomiuk and Cino Pertoldi for critical reading of the manuscript and the Novo Nordisk Foundation for financial support.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG et al (2000) The genome sequence of *Drosophila melanogaster*. Science 287:2185–2195

Alonso J, Rodriguez JM, Baena-López LA, Alonso MT, Santarén JF (2005) Constitutive expression of heat shock protein p23 correlates with proneural territories in imaginal discs of *Drosophila melanogaster*. Proteomics 5:1–10

Andolfatto P (2001) Contrasting pattern of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol Biol Evol 18:279–290

Ayme A, Tissières A (1985) Locus 67B of *Drosophila melanogaster* contains seven, not four, closely related heat shock genes. EMBO J 4:2949–2954

Barker JSF, Sene FD, East PD, Pereira MAQR (1985) Allozyme and chromosomal polymorphism of *Drosophila buzzatii* in Brazil and Argentina. Genetica 67:161–170

Barker JSF, Krebs RA, Davies HI (2005) Geographical distributions, relative abundance and coexistence of *Drosophila aldrichi* and *Drosophila buzzatii* in Australia. Austral Ecol 30:546–557

Barker JSF, Frydenberg J, González J, Davies HI, Ruiz A, Sørensen JG, Loeschcke V (2009) Bottlenecks, population differentiation and apparent selection at microsatellite loci in Australian *Drosophila buzzatii*. Heredity 102:389–401

Birney E, Durbin R (1997) Dynamite: a flexible code generating language for dynamic programming methods used in sequence comparison. Proc Int Conf Intell Syst Mol Biol 5: 56–64

Brudno M, Steinkamp R, Morgenstern B (2004) The CHAOS/DIALIGN WWW server for multiple alignment of genomic sequences. Nucleic Acids Res 32:W41–W44

Cohen RS, Meselson M (1985) Separate regulatory elements for the heat-inducible and ovarian expression of the *Drosophila hsp26* gene. Cell 43:737–746

Collinge JE, Anderson AR, Weeks AR, Johnson TK, McKechnies (2008) Latitudinal and cold-tolerance variation associate with DNA repeat-number variation in the *hsr-omega* RNA gene of *Drosophila melanogaster*. Heredity 101:260–270

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450:203–218

Dubrovsky EB, Dubrovskaya VA, Berger EM (2001) Selective binding of *Drosophila* BR-C isoforms to a distal regulatory element in the *hsp23* promoter. Insect Biochem Mol Biol 31:1231–1239

Flybase Consortium (2002) The flybase database of the *Drosophila* genome projects and community literature. Nucleic Acids Res. 30:106–108

Force A, Cresko WA, Pickett FB, Proulx SR, Amemiya C, Lynch M (2005) The origin of subfunctions and modular gene regulation. Genetics 170:433–446

Frank LH, Cheung HK, Cohen RS (1992) Identification and characterization of *Drosophila* female germ line transcriptional control elements. Development 114:481–491

Fry JD, Donlon K, Saweikis M (2008) A worldwide polymorphism in aldehyde dehydrogenase in *Drosophila melanogaster*: evidence for selection mediated by dietary ethanol. Evolution 62:66–75

Frydenberg J, Pertoldi C, Dahlgaard J, Loeschcke V (2002) Genetic variation in original and colonizing *Drosophila buzzatii* populations analysed by microsatellite loci isolated with a new PCR screening method. Mol Ecol 11:181–190

Frydenberg J, Hoffmann AA, Loeschcke V (2003) DNA sequence variation and latitude associations in *hsp23*, *hsp26* and *hsp27* from populations of *Drosophila melanogaster*. Mol Ecol 12:2025–2032

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

González J, Casals F, Ruiz A (2007) Testing chromosomal phylogenies and inversion breakpoint reuse in Drosophila. Genetics 175:167–177

Gottler LM, Salud Bea R, Shelburne CE, Ramamoorthy A, Marsh ENG (2008) Using fluorous amino acids to probe the effects of changing hydrophobicity on the physical and biological properties of β-Hairpin antimicrobial peptide protegrin-1. Biochem 47:9243–9250

Hasson E, Rodriguez C, Fanara JJ, Naveira H, Reig OA, Fontdevila A (1985) The evolutionary history of *Drosophila buzzatii* XXVI. Macrogeographic patterns in the inversion polymorphism in New World populations. J Evol Biol 8:369–384

Houlder DJ, Hutchinson MF, Nix HA, McMahon JP (2000) *Anuclim User Guide, Version 5.1*. Centre for Ressource and Environmental Studies, Australian National University, Canberra

JMP (2007) SAS Institut. Inc. Cary, NC. www.JMP.com

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Laayouni H, Santos M, Fontdevila A (2000) Toward a physical map of *Drosophila buzzatii*: use of randomly amplified polymorphic DNA polymorphisms and sequence-tagged site landmarks. Genetics 156:1797–1816

Matzkin LM, Merritt TJS, Zhu C-T, Eanes WF (2005) The structure and population genetics of the breakpoints associated with the cosmopolitan chromosomal inversion *In(3R)Payne* in *Drosophila melanogaster*. Genetics 170:1143–1152

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I (2000) VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16:1046–1047

Mestril R, Schiller P, Amin J, Klapper H, Ananthan J, Voellmy R (1986) Heat shock and ecdysterone activation of *Drosophila melanogaster hsp23* gene; a sequence element implied in developmental regulation. EMBO J 5:1667–1673

Michaud S, Tanguay RM (2003) Expression of the Hsp23 chaperone during *Drosophila* embryogenesis: association to distinct neural and glial lineages. BMC Dev Biol 3:9–20

Michaud S, Marin R, Tanguay RM (1997) Regulation of heat shock gene induction and expression during *Drosophila* development. Cell Mol Life Sci 53:104–113

Moriyama EN, Powell JR (1996) Intraspecific nuclear DNA variation in *Drosophila*. Mol Biol Evol 13:261–277

Morrow G, Heikkile JJ, Tanguay RM (2006) Differences in the chaperone-like activities of the four main small heat shock proteins of *Drosophila melanogaster*. Cell Stress Chaperones 11:51–60

Norry FM, Sambucetti P, Scannapieco AC, Loeschcke V (2006) Altitudinal patterns for longevity, fecundity and senescence in *Drosophila buzzatii*. Genetica 128:81–93

Piccinali R, Aguadé M, Hasson E (2004) Comparative molecular population genetics of the *Xdh* locus in the cactophilic sibling species *Drosophila buzzatii* and *D. koepfeae*. Mol Biol Evol 21:141–152

Piccinali RV, Mascord LJ, Barker JSF, Oakeshott JG, Hasson E (2007) Molecular population genetics of the α-*Esterase5* gene locus in original and colonized populations of *Drosophila buzzatii* and its sibling *Drosophila koepferae*. J Mol Evol 64:158–170

Powell LM, zur Lage PI, Prentice DRA, Senthinathan B, Jarman AP (2004) The proneutral proteins Atonal and Scute regulate neural target genes through different E-box binding sites. Mol Cell Biol 24:9517–9526

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R et al (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. Genome Res 15:1–18

Riddihough G, Pelham HRB (1986) Activation of the *Drosophila hsp27* promoter by heat shock and by ecdysone involves independent and remote regulatory sequences. EMBO J 5:1653–1658

Rogulski KR, Cartwright IL (1995) Multiple interacting elements delineate an Ecdysone-dependent regulatory region with secondary responsive character. J Mol Biol 249:298–318

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Sandaltzopoulos R, Mitchelmore C, Bonte E, Wall G, Becker PB (1995) Dual regulation of the *Drosophila hsp26* promoter *in vitro*. Nucleic Acids Res 23:2479–2487

Sharakhov IV, White BJ, Sharakhova MV, Kayondo J, Lobo NF, Santolamazza F, Torre AD, Simard F, Collins FH, Besanaky NJ (2006) Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the *Anopheles gambiae* complex. Proc Natl Acad Sci USA 103:6258–6262

Sorensen JG, Norry FM, Scannapieco AC, Loeschcke V (2005) Altitude variation for stress resistance traits and thermal adaptation in adult *Drosophila buzzatii* from the New World. J Evol Biol 18:829–837

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526