DR. FRANCISCO RODRÍGUEZ-TRELLES (Orcid ID : 0000-0001-6279-3630)



Article type : Special Issue

# Chromosomal inversions promote genomic islands of concerted evolution of *Hsp70* genes in the *Drosophila subobscura* species subgroup.

**Authors**: Marta Puig Giribets[1], María Pilar García Guerreiro[1], Mauro Santos[1], Francisco J. Ayala[2], Rosa Tarrío[1], & Francisco Rodríguez-Trelles[1]*.

1: Departament de Genètica i de Microbiologia, Grup de Genòmica, Bioinformàtica i Biologia Evolutiva (GGBE), Universitat Autonòma de Barcelona, 08193 Bellaterra (Barcelona), Spain

2: Department of Ecology and Evolutionary Biology, University of California, Irvine, California, United States of America

*: Corresponding Author: Francisco Rodríguez-Trelles

Departament de Genètica i de Microbiologia

Grup de Biologia Evolutiva (GBE)

Universitat Autonòma de Barcelona

08193 Bellaterra (Barcelona), Spain

Telephone: +34 93 581 2725

Email: franciscojose.rodrigueztrelles@uab.cat

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.14511

**Running Title**: Inversions as Islands of Concerted Evolution

**Abstract**

Heat-shock (HS) assays to understand the connection between standing inversion variation and evolutionary response to climate change in *D. subobscura* found that "warm-climate" inversion $O_{3+4}$ exhibits non-HS levels of Hsp70 protein like those of "cold-climate" $O_{ST}$ after HS induction. This was unexpected, as overexpression of *Hsp70* can incur multiple fitness costs. To understand the genetic basis of this finding, we have determined the genomic sequence organization of the *Hsp70* family in four different inversions, including $O_{ST}$, $O_{3+4}$, $O_{3+4+8}$ and $O_{3+4+16}$, using as outgroups the remainder of the *subobscura* species subgroup, namely *D. madeirensis* and *D. guanche*. We found: i) In all the assayed lines, the *Hsp70* family resides in cytological locus 94A, and consists of only two genes, each with four HS elements (HSEs) and three GAGA sites on its promoter. Yet in $O_{ST}$ the family is comparatively more compact. ii) The two *Hsp70* copies evolve in concert through gene conversion, except in *D. guanche*. iii) Within *D. subobscura*, the rate of concerted evolution is strongly structured by inversion, being higher in $O_{ST}$ than in $O_{3+4}$. And iv) In *D. guanche* the two copies accumulated multiple differences, including a newly evolved "gap-type" HSE2. The absence of concerted evolution in this species may be related to a long-gone-unnoticed observation that it lacks *Hsp70* HS response (HSR), perhaps because it has evolved within a narrow thermal range in an oceanic island. Our results point to a previously unrealized link between inversions and concerted evolution, with potentially major implications for understanding genome evolution.

**Key Words**: chromosomal inversion polymorphism, concerted evolution, climate change, *Hsp70*, evolution on islands, *Drosophila subobscura* subgroup.

## Introduction

Chromosomal inversions are a type of genomic rearrangement that is ubiquitous in nature (Hoffmann *et al*. 2008; Hoffmann & Rieseberg 2008). They consist in a breakage of a chromosome segment, and the reinsertion of the segment in the reversed orientation. Of the various consequences of chromosomal inversions, perhaps the one of most general evolutionary significance is suppression of recombination when in heterozygous combination, especially around the breakpoints (Kirkpatrick 2010). Through their linkage generation effects, inversions contribute to maintain favorable combinations of alleles in the face of gene flow, and are key to local adaptation (Dobzhansky 1947; Kirkpatrick & Barton 2006).

The evolution of chromosome inversions has been mainly investigated in dipterans, including *Drosophila* and anopheline mosquitoes, because of the technical advantages offered by the presence of giant polytene chromosomes. In many instances, inversion frequencies were found to exhibit systematic spatiotemporal variation patterns (e.g., Dobzhansky 1970; Coluzzi *et al*. 2002; Levitan & Hedges 2005; Umina *et al*. 2005; Ayala *et al*. 2011; Kapun *et al*. 2016), with those from *D. subobscura* standing out as some of the clearest ones (see below). This species is originary from the temperate Palearctic region, where it is broadly distributed with intense gene flow. Together with the insular endemics *D. madeirensis* and *D. guanche*, it forms the *subobscura* three-species subgroup (Krimbas 1992).

Extensive field data on inversion frequencies from *D. subobscura* revealed regular latitudinal (Krimbas 1992), seasonal (Rodríguez-Trelles *et al*. 1996) and long-term directional trends (Rodríguez-Trelles & Rodríguez 1998, 2010; Balanyà *et al*. 2006), which are overall consistent with expectations assuming temperature is the causative agent (Rodríguez-Trelles & Rodríguez 1998; 2010; Rezende *et al*. 2010). Twenty years ago, the inversion polymorphisms of the species were found to be evolving in association with contemporary climate warming (Rodríguez-Trelles & Rodríguez 1998; Hughes 2000; IPCC 2001, Parmesan & Yohe 2003; Bradshaw and Holzapfel 2006; Balanyà *et al*. 2006). Furthermore, the standing inversion variation maintained by the spatiotemporally fluctuating thermal environment enabled a rapid genome-wide evolutionary response of the species to a recent heat wave (Rodríguez-Trelles *et al*. 2013).

Of the five major acrocentric chromosomes of *D. subobscura*, the O chromosome (Muller element E) has received most attention. Although it has multiple gene rearrangements, the two common cosmopolitan $O_{ST}$ and $O_{3+4}$ are particularly interesting. They exhibit antagonistic spatiotemporal patterns, with cold-climate $O_{ST}$ increasing in frequency with laltitude and during the winter, and warm-climate $O_{3+4}$ increasing in frequency towards the equator and during the summer (Menozzi & Krimbas 1992; Rodríguez-Trelles *et al*. 1996; 2013). These arrangements differ by two overlapping inversions (denoted by lines below subscripts) originated independently on separate O chromosomes with the ancestral gene order $O_3$: $O_{ST}$ by reversal of $O_3$, and $O_{3+4}$ by superposition of inversion 4 on $O_3$ (Figure 1). The origin of $O_{3+4}$ was followed by several independent inversions which gave rise to the $O_{3+4}$ phylad, including, among others, the relatively rare $O_{3+4+8}$ and $O_{3+4+16}$ arrangements that show less clear spatiotemporal patterns. Nucleotide variation analyses found that $O_{ST}$ and $O_{3+4}$ segregate as linked blocks of loci that are recombinationally isolated from each other (Munté *et al*. 2005).

*D. madeirensis* and *D. guanche* originated allopatrically from continental Palearctic $O_3$ ancestors (González *et al*. 1983; Khadem *et al*. 2012) that dispersed into the Madeira and Canary Islands volcanic archipelagos of the Macaronesia 0.6-1.0 and 1.8-2.8 Mya, respectively (Ramos-Onsins *et al*. 1998). *D. madeirensis* remained monomorphic for $O_3$, whereas *D. guanche* became fixed for the Canary endemic $O_{3+g}$. Currently, both species coexist with *D. subobscura* in their respective islands owing to independent secondary contacts after continental *D. subobscura* propagules nearly fixed for $O_{3+4}$ re-entered the archipelagos. The three species are completely isolated reproductively from each other, except for *D. madeirensis* and *D. subobscura* (Krimbas & Loukas 1984; Rego *et al*. 2006) which are capable of limited gene exchange in collinear genomic regions not affected by inversions (Herrig *et al*. 2013). Compared to the thermal generalist *D. subobscura*, *D. madeirensis* and *D. guanche* have evolved within the narrower thermal range typical of the Tertiary relictual forests of the small oceanic islands they live on.

Laboratory assays with *D. subobscura* found that warm-climate $O_{3+4}$ shows higher adult thermal preference and heat tolerance than cold-climate $O_{ST}$ (Rego *et al*. 2010). Subsequent experiments aimed to elucidate the underlying physiological differences focused on the stress-inducible Hsp70 protein, because it is the major protein involved in thermal stress in *Drosophila* (Parsell & Lindquist 1993) and its gene locus has been mapped inside the region covered by $O_{3+4}$ (Moltó *et al*. 1992; Cuenca *et al*. 1998). $O_{3+4}$ has associated non-HS levels of Hsp70 protein like those of $O_{ST}$ after HS induction (Calabria *et al*. 2012), which was unexpected, because excessive expression of *Hsp70* can incur multiple fitness costs (Hoekstra & Montooth 2013), and it is in contrast with what is typical in short-lived Diptera like *Drosophila* (Garbuz & Evgen'ev 2017).

The observed difference in non-HS levels of Hsp70 protein between $O_{ST}$ and $O_{3+4}$ could be explained by a change in cis-regulatory sequence, which could (but not necessarily) be accompanied with a change in gene family size. In *Drosophila*, *Hsp70* promoters are among the simplest promoters (Tian *et al*. 2010). In the proximal promoter region, typically within 400 bp of the transcription start site (TSS), they contain regulatory HSEs that are binding sites for HS transcription factors (HSFs). HSFs are trimeric protein complexes encoded by a single-copy gene (CG5748; Jedlicka *et al*. 1997) which is highly conserved, thereby variation in the interaction HSF-HSE is expected to result mainly from variation in the HSEs. HSEs consist of contiguous inverted repeats of the pentamer 5'-nGAAn-3', where "n" can be any nucleotide. HSEs usually contain a minimum of three pentanucleotide units, that can be arranged in either head-to-head (HtH; nGAAnnTTCn) or tail-to-tail (TtT; nTTCnnGAAn) orientation (Perisic *et al*. 1989). The affinity with which HSFs bind HSEs is influenced by the degree of conservation of the canonical pentanucleotide motif. In a nGAAn unit the 2nd position is clearly the most conserved, in agreement with its critical role for binding, followed by the 3rd and 4th, the 1st and, in last place, the 5th position which varies freely (Tian *et al*. 2010). The number of consecutive pentanucleotide units in an HSE relates to the strength of the HS response. HSEs tolerate insertions, even if they bear no sequence similarity to the canonical motif, as long as they do not alter the spacing and phase of the pentanucleotide units. The resulting HSEs are classified as either "gap-type" or "step-type" HSEs (Hashikawa *et al*. 2006). Gap-type HSEs consist of two (or more) inverted units separated by a 5 bp gap [*e.g.*, nTTCnnGAAn(5bp)nGAAn], and step-type HSEs consist of direct repeats of nGAAn or nTTCn units separated by 5 bp [*e.g.*, nGAAn(5bp)nGAAn(5bp)nGAAn]. Of the different HSEs in a proximal promoter, the one located most downstream is the most important, as it is the one for which HSFs exhibit the most affinity, and binding to this HSE enhances binding to the next upstream HSEs. In

addition to HESs, the promoters of *Drosophila Hsp70* genes also contain binding-sites for the GAGA factor (GAF), which is required to establish an open chromatin conformation state for rapid activation of the HSR (Wilkins & Lis 1997). GAF binding sites typically occur within 150 bp upstream of the TSS and can be arranged in direct (GAGA) or inverse (TCTC) orientation (O'Brien *et al*. 1995).

In addition to having a compact promoter, *Hsp70* genes also lack introns and show codon usage bias towards efficiently translated codons. The fitness benefits of efficient coordinated upregulation of *Hsp70* copies are expected to spark a positive feedback loop, such that strong negative selection against the accumulation of divergence between paralogs would boost concerted evolution of the paralogs via nonreciprocal gene conversion (Bettencourt & Feder 2002; Nikolaidis & Nei 2004), which in turn would increase the efficiency of negative selection to maintain sequence identity of the paralogs (Sugino & Innan 2006). As a result, *Hsp70* copies within a genome are predicted to be more similar to each other than they are to orthologous copies in a related genome, regardless the level of regional functional constraint. In *Drosophila* this phenomenon has been investigated at the species level only (Bettencourt & Feder 2002; Garbuz & Evgen'ev 2017). Owing to their recombination suppression effects, however, chromosomal inversions have the potential to evolve their own patterns of concerted evolution that should vary depending on the level of constraint on *Hsp70* function. The stronger (weaker) the negative selection, the greater the chances that the paralogs evolve in concert (escape the *Hsp70* family) (Walsh 1987).

To understand the connections between chromosomal inversions and molecular evolution of *Hsp70* in *D. subobscura*, we have determined the genomic sequence organization of the *Hsp70* gene family in four different gene arrangements, including $O_{ST}$,

$O_{3+4}$, $O_{3+4+8}$ and $O_{3+4+16}$ (Figure 1), and the remainder of the *subobscura* species subgroup, namely *D. madeirensis* and *D. guanche*. Our results point to a previously unrecognized link between inversions and concerted evolution, with potentially major implications for understanding genome evolution.

## Materials and Methods

### *Drosophila* lines

We used five strains from *D. subobscura* plus one from each of *D. madeirensis* and *D. guanche*. The *D. subobscura* strains were made isogenic for the O arrangements of interest. The O arrangements were first isolated by crossing wild males to virgin females from the *cherry-curled* (*ch-cu*) recessive marker stock, and then isogenized using the *Varicose*/*Bare* (*Va*/*Ba*) balancer stock (Sperlich *et al*., 1977). $O_{ST}$ was isogenized using the *Va* mutant chromosome. The expression of the *Ba* gene is highly variable. Therefore, to prevent potential errors at sorting out phenotypically $O_{3+4}$, $O_{3+4+8}$ and $O_{3+4+16}$ homokaryotypes, the *Va*/*Ba* stock was previously selected for zero macrobristles on the scutum and scutellum. Crossing schemes and the methods for polytene chromosome staining and identification are described elsewhere (Rodríguez-Trelles *et al*. 1996). In the case of *D. madeirenseis* and *D. guanche*, we used inbred line material from our lab. All assayed lines were stored frozen at -20ºC immediately upon obtention.

*D. subobscura* strains were derived from our surveys of natural populations from Spain. Sampling locations and dates were as follows: $O_{ST1}$, $O_{ST2}$ and $O_{3+4}$, Berbikiz, latitude: 43°11’20.31’’N, longitude: 3°5’23.74’’W, date: May 14, November 14, and July 7, 2012, respectively; $O_{3+4+8}$, Vélez de Benaudalla, 36º50’23.66’’N, 3º30’59.32’’W, April 16-17, 2014; and $O_{3+4+16}$, Jérez del Marquesado, 37º11’6.74’’N, 3º10’30.37’’W, August 29-30, 2014. *D. madeirenseis* and *D. guanche* lines were derived from flies collected in Ribeiro Frío

(Madeira Island, Portugal; 32°43'00"N, 16°52'00"W, October 2011), and from the San Diego Stock Center ID 14011-0095.01, respectively

**In situ hybridization**

Polytene chromosome preparations were performed following Labrador *et al*. (1990). The gDNA template region for synthesis of the *in situ* hybridization probe was selected by BLASTN against a *de novo* genome draft assembly of our lab stock of the *ch-cu* strain, using available *Hsp70 cds* information from *D. pseudoobscura*, the closest relative to our species available in public databases. The assembly was generated upon request by Macrogen Inc (Seoul, South Korea) using 54,924,584 300 bp long mate-paired reads from a Miseq run of a 10 kb insert library (referred to as 2×300MP10 assembly). One of the hit contigs included a complete *Hsp70 cds* spanning 1.9 kb, whose termini were used for non-degenerate primer design. The fragments obtained by PCR amplification using gDNA from the *ch-cu* strain were gel-band extracted with PCR Clean-up gel extraction kit (Macherey-Nagel, Düren, Germany) and cloned into a 3kb pGEM-T vector (Promega, Madison, WI). Probes were labeled by random priming with digoxigenin (DIG-11dUTP) of purified PCR amplicons. Post-hybridisation washes were done following Schmidt (2002). Digital images were obtained at 400× magnification using a phase contrast Axio Imager.A1 Zeiss microscope and an AxioCam MRc 5 Zeiss camera. The location of the hybridization signals was determined using the standard cytological map for *D. subobscura* (Kunze-Mühl & Müller 1958).

***Hsp70* family genome sequence reconstruction strategy.**

Previous knowledge from *Drosophila* indicated that the *Hsp70* region in the *subobscura* species could be repetitious, difficult to assemble using short-read sequencing technologies alone (Tian *et al*. 2010). Accordingly, a combined *in silico*-wetlab recursive

approach was adopted, whereby, first, an *Hsp70 cds* from *D. pseudoobscura* was used as a query for BLASTN against our 2×300MP10 assembly. The contig sequence data was then used to screen genomic libraries from $O_{ST}$ and $O_{3+4}$ for positive clone sequence information, which was in turn used for the next rounds of BLASTN against the 2×300MP10 assembly.

**Genomic library preparation and screening**

Total high molecular weight gDNA from 500 mg of frozen adults from the $O_{ST}$ and $O_{3+4}$ lines was extracted using phenol-chloroform (Piñol *et al*. 1988). gDNA was digested with *Sau3A* to yield DNA fragments of ~15 kb. The fragments were ligated into the CIAP-treated Lambda Dash II/BamH 1 vector, and the recombinant DNA was packaged using Gigapack III XL (Agilent Technologies, Santa Clara, CA). Phage P2 infected *E. coli* were plated on NZYM culture medium (Amresco, Solon, OH). Plaque lifts were carried out onto a positively charged Biodyne B Nylon membrane (Pall Corporation, Pensacola, FL). Library screening used the same DIG-labelled 1.9 kb *Hsp70 cds* fragment as that for *in situ* hybridization. Prehybridization and hybridization steps followed García Guerreiro and Fontdevila (2007).

Membranes were incubated with anti-DIG antibody conjugated to alkaline phospatase, and the label developed with chromogenic alkaline phosphatase substrate NBT/BCIP (Roche, Indianapolis, IN). Candidate clones were secondarily screened to avoid false positives. Of all obtained isolates, only one from the $O_{ST}$ library was used in this study.

**DNA isolation, PCR amplification, sequencing and annotation.**

gDNAs were isolated from 5-10 frozen adults using phenol-chloroform and isopropanol precipitation. Oligonucleotides for PCR amplification and sequencing were designed using the software Primer3Plus (Untergasser *et al*. 2007) from sequence information

obtained from the combined *in silico*-wetlab recursive approach. PCR reactions were performed using DFS-Taq DNA polymerase (Bioron, Ludwigshafen, Germany) on a MJ Research PTC-100 thermal cycler (MJ Research Inc., Watertown, MA). PCR products were purified using PCR Clean-up kit (Macherey-Nagel, Düren, Germany), quantified with a NanoDrop-2000 spectrophotometer (Thermo Scientific Nanodrop), and assessed by agarose gel electrophoresis. In the case of *D. madeirensis* and *D. guanche*, PCR products were cloned into 3kb pGEM-T vectors (Promega, Madison, WI) before sequencing. Bidirectional DNA sequencing of DNA products was outsourced to Macrogen Inc. Sequences of the primers used for PCR amplification and sequencing are available from the authors upon request. Obtained sequences were manually annotated for gene structure using the BLAT tool implemented in the UCSC Genome Browser (Kent *et al*., 2002; http://genome.ucsc.edu). Putative cis-regulatory elements, including TSTs and TATA-like boxes, as well as HSEs and GAGA sites were mapped using the eukaryotic Neural Network Promoter Prediction server (http://www.fruitfly.org/seq_tools/promoter.html) and compiled *Drosophila* HSE data (Tian *et al*. 2010), respectively.

**Multiple sequence alignment (MSA), recombination scans and phylogenetic inference.**

MSA of the seven *Hsp70* sequences (hereafter referred to as seven OTUs dataset) was conducted using the progressive guide tree-based MAFFT algorithm (vs7; http://mafft.cbrc.jp/alignment/software/) with the accuracy-oriented method "L-INS-i" (Katoh *et al*. 2005) and default settings. The resulting MSA was set as the base MSA for assessing the reliability of the positional homology inference using Guidance2 (Penn *et al*. 2010; Sela *et al*. 2015). The base MSA obtained a guidance confidence score of 0.985, where a score of 1 indicates 100% robustness of the MSA to 100 bootstrap perturbations in the guide tree. All gaps localized to noncoding regions, particularly to the central segment of the

intergenic region (CIR) between the two *Hsp70* genes. Columns scoring below the guidance default value of 0.935 were removed. The MSA was manually refined using the Impale vs1.28 alignment editor (Martin *et al*. 2015), and checked for presence of stop codons and/or editing problems using MEGA (vs7.0.26) (Kumar *et al*. 2016). The conclusions derived from downstream analyses of this MSA were robust to the use of Gblocks (Castresana 2000) with either less or more stringent settings as an alternative MSA curating method.

Shared nucleotide composition biases among taxa can mislead phylogenetic reconstructions of *Drosophila* (Tarrío *et al*. 2001). Prior to modeling the substitution processes, we conducted exploratory tests of the hypothesis that each sequence conforms to the average character composition of the MSA using the chi-square (chi2) method implemented in the online version of IQ-Tree (W-IQ-Tree; http://iqtree.cibiv.univie.ac.at/; Trifinopoulos *et al*. 2016). Chi2 tests were applied both to unpartitioned and partitioned data, in the last case separately to every hypothesized partition. Besides bearing interest on itself, recombination also may be a source of conflicting phylogenetic signals in MSAs. Caution must be exercised, however, when assessing the impact of recombination as its effects can be mimicked by heterotachy and other similarly non-reticulating evolutionary processes (Sun & Golding 2011). The power to detect recombination in our MSA is limited because of the sparseness of intra-class, chromosomal rearrangement/species, sequence representation. On the other hand, we do not expect inter-class recombination to be a major factor in our data, considering recombination suppression effects of the inversions, and that the species are at least partially isolated reproductively. The MSA was checked for evidence of recombination using indirect tests of recombination, including the substitution distribution methods GeneConv (Sawyer 1989; Padidam *et al*. 1999) and MaxChi (Maynard-Smith 1992; Posada & Crandall 2001), and the phylogenetic method RDP (Martin & Rybicki 2000) implemented in RDP4 vs4.94 with default settings (Martin *et al*. 2015). In addition, we used the single

breakpoint phylogenetic method SBP (Kosakovsky Pond *et al*. 2006) implemented in the Datamonkey webserver (http://http://www.datamonkey.org/; Delport *et al*. 2010).

A maximum-likelihood framework was adopted for tree reconstruction. Model selection and tree inference were conducted using W-IQ-Tree. Unpartitioned substitution models, in which all characters are assumed to evolve under the same substitution process, and partitioned substitution models, in which different sets of characters are allowed to have their own substitution process, were considered. Partitioned models included DNA models and mixed models for combined DNA and amino acid characters, and DNA and codon characters. *A priori* partition schemes were established based on functional category (*i.e.* coding and noncoding) and gene/codon position identity (Table S1, Supporting Information). *A priori* mixed DNA-amino acid and DNA-codon partition schemes were identical to the DNA partition scheme, except that coding sites were translated to amino acids and recoded as codons, respectively. Optimal partitioning schemes were determined by hierarchical clustering of *a priori* partitions into increasingly fitter less-partitioned schemes until the model fit stops improving using ModelFinder (Kalyaanamoorthy *et al.* 2017) according to the Bayesian Information Criterion (BIC). Among-site rate variation was accommodated allowing for the new FreeRate heterogeneity model (+R), in which site rates are directly inferred from the data, in addition to the common invariable sites (+I) and gamma rates (+G) models. Tree searches were conducted starting from sets of 100 initial maximum parsimony trees using nearest neighbor interchange with default perturbation strength and a stopping rule settings. Branch-support was assessed using the ultrafast bootstrap approximation (UFboot; 1000 replicates) (Minh *et al*. 2013), and two single-branch tests including the Shimoidara–Hasegawa-like approximate likelihood ratio test (SH-aLRT; 1000 replicates) (Guindon *et al*. 2010), and the approximate Bayes parametric test (Anisimova *et al*. 2011).

## Results

### Localization and genomic sequence organization of the *Hsp70* genes in the *subobscura* subgroup.

*In situ* hybridization of a *D. subobscura* 1.9 kb long *Hsp70 cds* probe to isogenic lines for the $O_{ST}$, $O_{3+4}$, $O_{3+4+8}$ and $O_{3+4+16}$ arrangements identified a single *Hsp70* locus that invariably mapped to subsection 94A of the Kunze-Mühl & Müller (1958) standard cytological map (Figure S1, Supporting Information). Gene family reconstruction at the sequence level indicated that the *Hsp70* family consists of a single set of only two copies arranged as a head-to-head inverted repeat (*Hsp70IR*) in all the assayed lines of the *subobscura* subgroup.

MSA of the seven obtained sequences with their reverse-complemented sequences (Figure 2) showed that the two repeats are separated by a central intergenic region (CIR) of 1055±119 bp, which consists of a unique central sequence flanked by two gap-rich regions interspersed with a few short, repeat-containing aligning blocks. Outwards from both sides of the CIR, the alignment enters abruptly into the high-similarity region of the duplicated blocks, each consisting of a full-length, intronless 1929 bp long *Hsp70* gene preceded by 555±28 bp of 5' upstream sequence, and followed by 199±6 bp of 3' downstream sequence. The *Hsp70IR* is flanked by genes CG5608, at the downstream end of the *Hsp70* copy placed on the minus strand, and *Dmt* (*Dalmatian*; CG8374) at the downstream end of the *Hsp70* copy placed on the plus strand, as depicted in Figure 2. The two *Hsp70* copies were respectively named B and A, following their order of discovery.

Figure 2 shows the lengths in base pairs of the various stretches of noncoding sequence along *Hsp70IR*, including the *Hsp70* genes 3' downstream and 5' upstream sequences and the CIR, for the seven lines of this study. The aggregate region is shortest in *D. subobscura* $O_{ST}$ (2470 bp and 2485 bp, for $O_{ST1}$ and $O_{ST2}$, respectively), and longest in *D.*

*guanche* $O_{3+g}$ (2841 bp). Most of the variation in aggregate length is accounted by the CIR, which in *D. guanche* $O_{3+g}$ is 20-25% longer than in the O chromosomes from the other two species.

Figure 3 provides a schematic view of a MSA of the 5' upstream region of the 14 *Hsp70* genes, spanning from the translation start site to the beginning of the CIR. The region includes both core and proximal promoters. The proximal promoter contains four HSEs (1-4) and three completely conserved GAGA-binding sites, in agreement with most findings for *Drosophila Hsp70* genes (Tian *et al*. 2010). The spacing between HSEs shows little variation, surely because of the recentness of the species subgroup. Each HSE, except HSE2, shows the same pattern of number and orientation of pentanucleotides across all sequences, specifically HSE1: [3(TtT)]; HSE3: [4(HtH)]; and HSE4: [3(TtT)]. HSE2 exhibits potentially functional variation consisting in an indel of 10 bp between the second and the third pentanucleotide units. Comparison with the outgroup species *D. pseudoobscura* and *D. persimilis* indicates that the event is a gain of sequence in the HSE2 of *Hsp70B* in *D. guanche*. Closer inspection of the gained sequence reveals that i) its length is multiple of 5 bp; ii) when divided into two pentanucleotides, the one most upstream is similar to the canonical nTTCn unit, whereas the one most downstream bears no similarity with any of the two types of pentanucleotide unit; and iii) it is flanked by canonical motifs that are in functional phase to each other. These findings suggest that the HSE2 of *Hsp70B* in *D. guanche* evolved from an ancestral continuous four pentanucleotide-unit state to its present gap-type six pentanucleotide-unit state through acquisition of new internal sequence. All else being equal, the extra binding sites in HSE2 are expected to impart *D. guanche* with enhanced potential for heat-shock response.

Comparative analysis of gene synteny across the *Drosophila* phylogeny shows that the block formed by the tandem array of genes CG5608–*Hsp70IR–Dmt* is conserved in *D. pseudoobscura* and *D. persimilis*, two genome-draft-sequenced sibling members of the *obscura* species group that branched off from the lineage of their sister *subobscura* subgroup around 17.7 My ago (Tamura *et al*. 2004). In addition, the block is partially conserved in the *melanogaster* group, which retains the pair CG5608–*Hsp70Ba*, and in *D. willistoni*, where CG5608 and *Dmt* are flanking a putative rearrangement of the *Hsp70IR* that resulted in equal orientation of the two copies, of which the one nearest to CG5608 stands as a truncated form of the gene, and the one nearest to *Dmt* (GK10980) is annotated as a pseudogene (GK10978). BLAST queries into deeper nodes of the *Drosophila* tree did not yield additional evidence for the block. All together, these results indicate that the block CG5608–*Hsp70IR–Dmt* has remained structurally stable in the lineage leading to the modern *subobscura* species subgroups for at least ~62.2 My elapsed since the diversification of the Sophophora subgenus of *Drosophila* (Tamura *et al*. 2004).

**Molecular evolution and genealogy of the orthologous *Hsp70* genes in the *subobscura* species cluster.**

After MSA, the final alignment matrix consisted of seven taxa and 7042 characters of which 159 were parsimony-informative. Preliminary exploratory analysis using a chi2 approximation indicated that all sequences conform to the average character composition of their MSAs both from the unpartitioned dataset, and from each partition considered separately. The recombination scans did not detect significant evidence of recombination events, thereby all characters in the MSAs were assumed to share a unique branching history.

Table 1 shows the results of the modeling of the substitution process. Unpartitioned and partitioned models, with a priori partitionning and best-merging of a priori partitionning

schemes, for nucleotide (NUC), mixed nucleotide and amino acid (NUC+PROT), and mixed nucleotide and codon (NUC+CODON) data types were considered (partition identities and the specific models are provided in Table S1, Supporting Information). The best BIC model is a ModelFinder best-merging of an eight *a priori* partitions scheme of mixed noncoding nucleotide (5 partitions, 2932 characters in total) and amino acid (3 partitions, 1369 characteres) data (BIC score of 20,969.28), into a model with only two partitions to accommodate differences in the substitution process (HKY+I and JTT for the noncoding nucleotide and the amino acid partitions, respectively) between the two types of characters (BIC score of 20,466.10).

Figure 4 shows the ML tree obtained from the mixed NUC+PROT *Hsp70* data set with the best-fit two partition model of Table 1, with empirical base frequencies and edge-linked proportional branch lengths between partitions. The tree is well resolved (SH-aLRT ≥ 80; aBayes and uBFBoot ≥ 95; Anisimova *et al*. 2011), except for the trichotomy of the $O_{3+4}$ phylad of *D. subobscura*. Accordingly, *D. guanche*, known from external evidence to represent the first to split after the origination of the *subobscura* lineage, places the root between *D. subobscura* and *D. madeirensis*. Within *D. subobscura*, $O_{ST}$ and the $O_{3+4}$ phylad constitute separate monophyletic groups. The topology is robust to allowing the two partitions to have their own sets of branch lengths (edge-unlinked) in the model, and is congruent with the topologies that result from analysis of each of the two partitions separately, and with the topologies that obtain after using less-fit models of Table 1.

Figure 2 also represents the variation in nucleotide substitution rates among functional categories of sites across the *Hsp70* family region. Substitution rates were estimated with the best BIC model for the unpartitioned nucleotide MSA (TN93; Tables 1 & S1), using BaseML from the PAML vs. 4.9d package (Yang 2007) with the C option and rates scaled to the rate of substitution in the 3' flanking region of *Hsp70B*, under the topology shown in Figure 4.

From the TN93+C model, the slowest evolving characters are the first and second codon positions combined, and the fastest the CIR region. The first and second codon positions together change 12.2 and 10.7 times more slowly than third codon positions in *Hsp70A* and *Hsp70B*, respectively, as expected if the two paralogous genes were subjected to strong purifying selection at the protein level.

**Variable rates of concerted evolution between paralogous *Hsp70* genes.**

Similarity between duplicates may be accounted for by two not mutually exclusive hypotheses. The functional constraint hypothesis predicts that, with increasing time after the duplication similarity should decrease in unconstrained sites, compared to constrained sites. In contrast, the concerted evolution by gene conversion hypothesis predicts that similarity should be equal across sites, irrespective of the variation in functional constraint. The *Hsp70* duplication event investigated here predated the origin of the *Sophophora* subgenus (estimated to be ~62.2 Mya; Tamura *et al*. 2004), therefore it can be safely assumed to be at least 20 times as old as the diversification of the *subobscura* species cluster (<3 Mya; Herrig *et al*. 2013). Under a functional constraint-only scenario, the synonymous divergence (Ks; Nei & Gojobori 1986) between paralogs should be at least 20 times greater than in the ortholog comparison between *D. guanche* and either *D. madeirensis* and *D. subobscura*. On the contrary, the estimated average Ks between paralogs (0.0186±0.0049) was 5.7 and 3.7 times lower than the average Ks between orthologs for *Hsp70A* (0.1064±0.0131) and *Hsp70B* (0.0690±0.0107), respectively.

Figure 5 represents the spatial distribution of the estimated average pairwise paralogous and orthologous nucleotide divergences along the *Hsp70* repeat. Compared to the orthologous divergence, paralogous divergence is strongly U-shaped, with distinct higher values around the edges that drop abruptly to a consistent near-zero level in the intervening

central region. The central region spans from the upstream end of HSE4 to the end of the *Hsp70* gene, encompassing most of the repeat. Phylogenetic analysis from the margin and central regions clearly shows that the margin regions support clustering of orthologous sequences, whereas the central region supports clustering of paralogous sequences. This is precisely what is expected if the *Hsp70* duplicates evolved in concert by gene conversion since long before the split of the *subobscura* subgroup.

A phylogenetic network of the central region produced by split decomposition (Figure 6) shows that interparalog variation is deeply structured hierarchically by species, chromosomal arrangement within species, and isogenic line within chromosomal arrangement. This high level of substructuring is explained as a result of a shifted equilibrium towards an enhanced role of intra-chromosomal gene conversion in species segregating for chromosomal arrangements, because of the inter-chromosomal recombination suppression effect of the inversions. The presence of reticulations within the network indicates that gene conversion has not completely eroded conflicting evidence of the orthologous relationships. The tightness of the paralogous clusters is highly heterogeneous, suggesting that the rate of gene conversion has not been uniform across the different species and chromosomal arrangements. Within *D. subobscura*, $O_{ST}$ and $O_{3+4}$ are of particular *a priori* interest, because the latter chromosomal arrangement has been shown to confer enhanced thermal tolerance to its bearers compared to the former. Interestingly, $O_{ST}$ shows the tightest, and $O_{3+4}$ the loosest clustering. Table 2 shows estimated pairwise synonymous (Ks) and nonsynonymous divergences (Ka) (Nei & Gojobori 1986) among *Hsp70* orthologous and paralogous. The difference in degree of clustering between $O_{ST}$ and $O_{3+4}$, as measured by the difference in the relatively unconstrained Ks between the corresponding pairs of paralogs [0.0072±0.0032 and 0.0036±0.0025 for $O_{ST1}$(AB) and $O_{ST2}$(AB), and 0.0215±0.0058 for $O_{3+4}$(AB), respectively; see Table 2] is statistically significant (two-tailed

z-test; $P$=0.034 and $P$=0.005, for $O_{ST1}$ and $O_{ST2}$ versus $O_{3+4}$, respectively). This suggests that gene conversion has been especially active in $O_{ST}$ compared to $O_{3+4}$. At the amino acid level, however, the two chromosomal arrangements exhibit near-zero interparalog Ka values, as expected if the *Hsp70* proteins encoded by the two arrangements are highly constrained (Table 2). Overall, *D. guanche* exhibits the loosest clustering of paralogs (Ks=0.0620±0.0098; significantly greater than 0.0114±0.0041 for the average across all other strains; two-tailed z-test; $P$<0.001), suggesting that the rate of gene conversion has been strongly reduced in this species. In this case, however, the reduction in the rate of gene conversion occurs in parallel with a significant increase in interparalog Ka (0.0113±0.0029 *versus* 0.0006±0.0005; for *D. guanche versus* the average across the remainder sequences; two-tailed z-test; $P$<0.000). Maximum-likelihood-ratio tests of a local molecular clock carried out separately on i) the 5' regulatory region (521 sites), ii) the third codon positions (643), the fourfold degenerate sites (307), and iii) the translated amino acid sequences (643) of *Hsp70*, assuming respectively the TN93+dG nucleotide model and the JTT+dG amino acid model and the primary tree contained in the network of Figure 6, indicated a significant asymmetry of the evolutionary rate, with *D. guanche* copy A showing faster rate than copy B at third codon and fourfold degenerate positions ($-2\log\Lambda$=0.22, df=1, P<0.647; $-2\log\Lambda$=11.30, df=1, P=0.001; $-2\log\Lambda$=4.32, df=1, P=0.038; $-2\log\Lambda$=0.26, df=1, P=0.610 for the 5' regulatory region, the third codon positions, the fourfold degenerate sites, and the amino acid sequence, respectively; conducted using Paml vs 4.9d's BaseML and CodeML programs; Yang 2007). Associated to this increase in Ka there is a significant reduction of codon usage bias [44.885±0.006 *vs*. 43.412±0.217, for the averages for *D. guanche* Hsp70A-B *vs*. the remainder sequences; measured using the improved effective number of codons index (Nc; Sun *et al.* 2013); two-tailed z-test; P=0.014]. The increased Ka and Ks values, together with the two paralogs showing similar levels of codon usage deoptimization suggests

a relaxation of natural selection on *D. guanche*'s *Hsp70*. Still, the asymmetrry in Ks suggests that paralog B retained more of the ancestral features than paralog A.

It may be worth noting that the two copies of *Hsp70* are conspicuously more distantly spaced physically to each other in *D. guanche* than in the two other members of the *subobscura* cluster, and also that the copy that evolves more slowly in *D. guanche* (copy B) is the one associated to the newly evolved HSE2 in the proximal promoter.

## Discussion

### Genetic basis of the HS induced-like basal protein levels in $O_{3+4}$.

Herein, for the first time, we have determined the genomic sequences and conducted a comparative analysis of the HS inducible *Hsp70* family across chromosomal inversions and species of the *subobscura* subgroup. Our primary motivation has been to understand the molecular underpinnings of why *D. subobscura* flies homokaryotypic for the warm-climate $O_{3+4}$ chromosomal arrangement exhibit basal Hsp70 protein levels like those attained by their cold-climate $O_{ST}$ counterpart after a HS treatment (Calabria *et al*. 2012). Considering the genealogy of the chromosomal arrangements (Figure 1), the trait should be derived in $O_{3+4}$, and the most obvious expected causative factors would be a change in the number and/or arrangement of cis-regulatory elements and/or *Hsp70* genes. In contrast, we found a common pattern of cytological location and number of cis-regulatory elements (except for a newly evolved variant of HSE2 in *D. guanche*; see below) and gene copies, which evolve in concert through gene conversion. Gene conversion is strongly U-shaped, and precisely limited to the 5' cis-regulatory and *cds* regions, as expected for a gene family under long-term selection for efficient induction of more of the same product (Sugino & Innan 2006; Osada & Innan 2008). The pattern of concerted evolution, however, is strongly structured and idiosyncratic accross lineages as expected from the barriers to interchromosomal genetic exchange.

In all the sequenced lines, the *Hsp70* family is arranged in one cluster located in chromosome O, subsection 94A of the Kunze-Mühl & Müller (1958) standard cytological map. This location is consistent with previous findings in the *subobscura* subgroup that used *in situ* hybridization alone (Cuenca *et al*. 1998) or combined with HS induced transcriptional puffing (Moltó *et al*. 1992). These early efforts, however, did not contemplate the possibility that the cytological organization of the *Hsp70* locus could change among chromosomal arrangements. At the sequence level, the cluster consists of only two intronless and closely spaced *Hsp70* genes arranged in a head-to-head inverted repeat. A similar number and arrangement of genes is also found in the available genomes from *D. pseudoobscura* and *D. persimilis*. The lineage of these two species and that of the *subobscura* subgroup represent separate branches from the earliest split of the *obscura* group, which suggests that the genomic organization of the *Hsp70* family has remained largely unchanged during the evolution of this species group. The "two-genes" only configuration might be exceptional among *Drosophila*, for all the non-*obscura* species investigated to date show additional copies in variable numbers and orientations (Garbuz & Evgen'ev 2017). This notwithstanding, the fact that all fruitfly species investigated thus far have at least one functional *Hsp70IR* has been advanced to propose that this compact palindrome-like structure is the ancestral condition of the genus. Our comparative analysis indicates that in the *obscura* group, the *Hsp70IR* resides in a synteny block (*CG10886–Hsp70IR–Dmt*) whose origin postdated the radiation of *Drosophila*, and which is partially decayed in the *melanogaster* and *willistoni* groups. Accordingly, during the evolution of *Drosophila*, the *Hsp70IR* function would have experienced turnover of the particular paralogs on which became eventually instantiated.

If the HS induced-like basal protein levels of *Hsp70* in $O_{3+4}$ are not ascribable to a family size expansion, then they could be caused by a change in the number and/or context of cis-regulatory elements. Yet, all known key determinants of *Hsp70* promoter strength that are characteristic of *Drosophila*, including HSEs 1-4 and three GAGA sites, as well as their positions relative to the TSS (Wilkins & Lis 1997; O'Brien *et al*. 1995; Garbuz and Evgen'ev 2017), were found to be conserved.

In *D. melanogaster* the chromatin insulators *scs* (specialized chromatin sequence) and *scs*', which prevent the 87A heat shock locus (containing the *Hsp70Aa* and *Hsp70Ab* genes) from long-distance regulatory interactions, reside in the promoters of its immediate proximal and distal flanking genes (CG31211 and CG3281, respectively; Udvardy *et al*. 1985; Kuhn *et al*. 2004). Conservation of the *CG10886–Hsp70IR–Dmt* synteny block in the *obscura* group, however, makes it unlikely that the HS induced-like basal *Hsp70* protein level in $O_{3+4}$ is due to disruption of the enhancer-blocking activity of these boundary elements. Here, it might be worth noting that after long use of the polytene chromosome technique for *D. subobscura* chromosomal inversion identification in our laboratory, puffing activity at the 94A locus in $O_{3+4}$ has never been observed (Moltó *et al*. 1988), although it could be argued that the method is based on analysis of third-instar larvae only.

Considering hypothesized roles of transposable elements (TEs) as instrumental for *Hsp70* evolution in *Drosophila* (Zatsepina *et al*. 2001; Lerman *et al*. 2003; Evgen'ev *et al*. 2014), we screened the sequences of this study for repetitive sequences using the Genetic Information Research Institute (GIRI) Repbase (http://www.girinst.org/repbase/; Jurka *et al*., 2005) and RepeatMasker (http://www.repeatmasker.org/; Smit *et al*. 2015). We put particular attention on the promoter and 3' regions, as they contain constitutively nucleosome free, open chromatin domains (Karpov *et al*. 1984; Tsukiyama *et al*. 1994; Petesch & Lis 2008), with the former found to have been target of P element insertions in *D. melanogaster*

(Shilova *et al*. 2006). No evidence of TE insertions in the *Hsp70* sequences were detected, in agreement with the above-discussed stability of the *CG10886–Hsp70IR–Dmt* synteny block in the *obscura* group.

Collectively, all the above-discussed evidence indicates that the mechanistic basis of Calabria *et al*.'s (2012) observation of an atypical *Hsp70* protein expression pattern in $O_{3+4}$, either relies on more subtle genetic differences difficult to detect with our approach, or does not reside in the locus 94A. For example, they could be caused by regulation at the post-transcriptional level. Alternatively, the observation might be a false positive, and further replication is warranted.

**Chromosomal inversions promote genomic islands of concerted evolution.**

To observe as high a degree of homogeneity between intrachromosomal copies of *Hsp70* as in Figure 5 requires, first, that the rate of intrachromosomal recombination is high relative to the rate of mutation and, second, that interchromosomal recombination is rare (Liao 1999). Substructuring of concerted evolution at interspecific level is accounted for by reproductive barriers to genetic exchange. Of the three *subobscura* species, *D. guanche* is reproductively isolated from the other two (González *et al*. 1983; Krimbas & Loukas 1984). *D. madeirensis* and *D. subobscura* are partially isolated from each other, but conclusions about the degree of the isolation are mixed (Krimbas & Loukas 1984; Khadem & Krimbas 1991; Rego *et al*. 2006; Herrig *et al*. 2014). Even in the unrealistic situation that the two species would mate freely producing viable offspring, interchromosomal exchanges in the 94A locus would likely be rare, because the region is linked to the $O_3$ inversion, which is monomorphic in *D. madeirensis* and absent in *D. subobscura* (Figure 1; Larruga *et al*. 1983; Khadem *et al*. 1998).

Typically, interchromosomal recombination may occur via reciprocal crossing-over, which can be single or double, and/or non-reciprocal, so-called "copy-and-paste" gene conversion, which can occur either associated to or in absence of crossing-over (Korunes & Noor 2017). In *Drosophila*, recombination does not occur in males and, in females that are heterozygous for paracentric inversions, single crossovers are suppressed, and double crossovers are unlikely for short inversions (recombination length <20 cM; Navarro *et al*. 1997), whereas for large inversions they are more likely to affect the central part of the inversion. In contrast, noncrossover gene conversion events are expected to occur uniformly along inversions regardless their size. The inversions of this study should all be considered long [$O_3$, the shortest one, is ~23 cM long, or 12% (Loukas *et al*. 1979; Krimbas 1992) of the total O chromosome length of 190.7 cM (Pegueroles *et al*. 2010)] and old enough [0.33±0.13 *vs*. 0.35±0.05 Mya, respectively; average across the rp49 (Ramos-Onsins *et al*. 1998; Rozas *et al*. 1999); Acph-1 (Navarro-Sabaté *et al*. 1999); and Fmr1 (Pegueroles *et al*. 2013) loci] for recombination to have eroded interchromosomal differentiation. The position of the *Hsp70* locus is, however, well outside the central one-third of the inversions length in all cases, whereby the likelihood of it being affected by double crossovers should be small. On the other hand, locus CG10886, which we found located inmediately downstream of *Hsp70B* in the *subobscura* subgroup, did not show evidence of recombination in surveys of the genetic differentiation between $O_{ST}$ and $O_{3+4}$ in *D. subobscura* (Munté *et al*. 2005), and between *D. subobscura* carrying either of these two arrangements and *D. madeirensis* (Khadem *et al*. 2012). These and our results suggest that interchromosomal arrangement exchanges at the *Hsp70IR* locus in *D. subobscura*, and between this species and *D. madeirensis* have been unlikely; had they occurred, they would have been disadvantageous either directly, because they would have altered *Hsp70* function, or indirectly, because they would have interfered with co-evolution of the paralogs through tethering intrachromosomal concerted evolution.

Interaction between linkage and selection for concerted evolution may in fact be a general phenomenon contributing to the maintenance of chromosomal inversion polymorphisms in nature.

**Variation in the rate of concerted evolution between chromosomal inversions of *D. subobscura*.**

In addition to be deeply structured, the rate of concerted evolution in *D. subobscura* is also highly heterogenous depending on the chromosomal arrangement. Besides the lack of introns, several potentially synergistic factors have been proposed to contribute to enhance the rapid accumulation of mRNAs and their translation upon activation of a functional HS inducible *Hsp70* promoter, including increased compactness of the gene array, increased identity between paralogs, and codon usage bias towards efficiently translated codons (Carlini & Makowski 2015). Within *D. subobscura*, the comparison of $O_{ST}$ and $O_{3+4}$ is of the most interest on prior biological grounds. Long gathered evidence from field and laboratory studies indicate that the fitness of $O_{ST}$ is negatively impacted by heat-stress compared to $O_{3+4}$. As shown herein, this would happen in spite of $O_{ST}$ showing greater compactness and identity of the *Hsp70* paralogs than $O_{3+4}$. In fact, these differences do not seem to translate into differences in amount of *Hsp70* protein between the two arrangements after HS, at least from adults (Calabria *et al*. 2012). It rather would seem that the differential sensitivity to heat-stress of the two arrangements depends on other linked loci, and that this condition imparts stronger purifying selection to maintain an efficient heat response on $O_{ST}$ than on $O_{3+4}$. Theoretical results indicate that the effect of gene conversion in multigene families is equivalent to an increase in effective population size (Osada & Innan 2008; Fawcett & Innan 2011). Accordingly, the rate of *Hsp70* intrachromosomal gene conversion would be greater in $O_{ST}$ than in $O_{3+4}$ because it allows more efficient removal of deleterious mutations.

**Molecular evolutionary basis of vulnerability to global climate change.**

The case of *D. guanche* may shed light on the molecular mechanisms of evolutionary paths that could result in increased vulnerability to present global change. Our observation of a relaxation of natural selection on the *Hsp70* genes in this species is consistent with early experimental results on HS induced polytene chromosome puffing patterns (Moltó *et al*. 1987; 1988) showing that i) the optimum HS temperature (*i.e.*, that inducing the most intense puffing activity pattern) is higher in *D. guanche* (37°C) than in *D. subobscura* (31°C; using an $O_{ST}$ line); ii) *D. guanche* shows less total heat-shock induced puffing activity than *D. subobscura*; and iii) in contrast to *D. subobscura*, *D. guanche* does not show puff activity in locus 94A (Figure 10 in Moltó *et al*.1988), which suggests that HS induction of *Hsp70* function has become strongly diminished, if not completely eliminated in the latest species. It is worth mentioning that those experiments were based on late third instar larvae and prepupa. Unlike the adult stage, these are sessile life stages that cannot avoid heat-stress by moving to a thermally favorable site, and therefore are expected to be most reliant upon efficient upregulation of the HSR. Loss of HS inducible *Hsp70* in *D. guanche* may seem at odds with the presence of a newly evolved gap-type variant of HSE2 containing extra binding sites, hence predictably stronger, in the copy *Hsp70B* of this species. Perhaps the change segregated as a polymorphism and happened to be absent in the early experiments; or it was present, but it causes a shift in *Hsp70* cell-type/developmental regulation that was not reflected in the investigated puffing activity patterns. The apparent loss of HS inducibility in *Hsp70* is puzzling and deserves additional investigation.

*D. guanche* originated 1.8-2.8 Mya after an ancestral mainland *subobscura* propagule arrived in the Canary Islands of the Macaronesian region (Herrig *et al*. 2013). Currently, the species is endemic to the Tertiary relictual montane (600-1200 m a.s.l.) evergreen laurel forests (so-called "laurisilva") of the archipelago (Monclús 1976). The islands form as the

oceanic crust of the African Plate rotates slightly counterclocwise over a volcanic "hot-spot"; after reaching maximum area and elevation, they begin to erode and subside to below sea levels as the plate carries them to the north east (Fernández-Palacios *et al*. 2011). Today's laurisilva forests are thought to have survived the pronounced climatic change of the last 2 Myr owing to the permanent mild, sub-humid conditions provided by the influence of trade winds moisture during the geological time the islands showed a suitable altitudinal range (Nascimento *et al*. 2009). Accordingly, *D. guanche* would have transitioned from an eurythermal continental ancestor to a stenothermophile adapted to the islands long-term stable climatic conditions. Thermal niche specialization, together with reduced effective population size typical of island endemics (Woolfit & Bromham 2005) would have caused a relaxation of natural selection on *Hsp70* and, consequently, a shift toward increased accumulation of effectively neutral changes. Increasing interparalog divergence would have caused a decline in the frequency of gene conversion which, ultimately, increased the probability that the two *Hsp70* paralogs escape concerted evolution (Walsh 1987). That *D. guanche* has retained a low effective population size in the long term, is supported by reports of increased rates of synonymous divergence at other unliked regions of the species' genome (Llopart & Aguadé 1999; Pérez *et al*. 2003; Sánchez-Gracia & Rozas 2011). The idea is also consistent with an increase of about 10% of the whole genome of the *subobscura-guanche-madeirensis* (SGM) TE-related sequences specifically in this species (Miller *et al*. 2000).

Loss of HS inducibility of *Hsp70* in *D. guanche* may have increased the species' extinction risk in the face of the ongoing global warming, considering that i) human impact on the Canary Islands, initiated some 2500 yr BP (~15,000 *D. guanche* generations), has dramatically fragmented the original laurisilva forests, which currently occupy less than 12% of their human precolonization range, being largely confined to a few isolated gorges (Fernández-Palacios *et al*. 2011); and ii) its close relative *D. subobscura*, a comparatively

eurythermal species with greater population sizes and levels of standing inversion variation, is already responding to the change in the thermal environment (Rodríguez-Trelles & Rodríguez 1998; 2013; Balanyà *et al*. 2006).

**Outlook on the link between chromosomal inversions and concerted evolution**.

We have identified a previously unrealized link between chromosome inversions and concerted evolution, which, considered the pervasiveness of these phenomena, could have major implications for understanding genome evolution. Gene conversion contributes to maintain coadapted gene copies distributed across large genomic regions (Ezawa *et al*. 2006; Casola *et al*. 2010). Newly arisen inversions could disrupt existing patterns of concerted evolution through altering the relative orientation and distance between genes, which may impose constraints on their position and size. On the other side, new inversions can spread by promoting specific patterns of intrachromosomal concerted evolution through their interchromosomal recombination suppression effects.

**Acknowledgements**

This work was supported by the Spanish Ministerio de Ciencia e Innovación (CGL2013–42432P and CGL2017-89160P) and Grant 2009SGR 636 from Generalitat de Catalunya to the Grup de Genòmica, Bioinformàtica i Biologia Evolutiva. M.P.G. was supported by a PIF PhD fellowship from the Universitat Autònoma de Barcelona (Spain). We thank Antonio Fontdevila for help with the collection of *D. madeirensis*, and Montserrat Peiró Navarro for help with the in-situ hybridizations.

## References


Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C., & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology*, 60, 685-699.

Ayala, D., Fontaine, M. C., Cohuet, A., Fontenille, D., Vitalis, R., & Simard, F. (2011). Chromosomal inversions, natural selection and adaptation in the malaria vector Anopheles funestus. *Molecular Biology and Evolution*, 28, 745-758.

Balanyà, J., Oller, J. M., Huey, R. B., Gilchrist, G. W., & Serra, L. (2006). Global genetic change tracks global climate warming in *Drosophila subobscura. Science, 313,* 1773-1775.

Bettencourt, B. R., & Feder, M. E. (2002). Rapid concerted evolution via gene conversion at the *Drosophila hsp70* genes. *Journal of Molceular Evolution*, 54, 569-586.

Bradshaw, W. E., & Holzapfel, C. M. (2006). Evolutionary response to rapid climate change. *Science*, 312, 1477-1478.

Calabria, G., Dolgova, O., Rego, C., Castañeda, L. E., Rezende, E. L., Balanyà, J., Santos, M. (2012). *Hsp70* protein levels and thermotolerance in *Drosophila subobscura*: a reassessment of the thermal co-adaptation hypothesis. *Journal of Evolutionary Biology*, 25, 691–700.

Carlini, D. B., & Makowski, M. (2015). Codon bias and gene ontology in holometabolous and hemimetabolous insects. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, 324, 686-698.

Casola, C., Ganote, C. L., & Hahn, M. W. (2010). Nonallelic gene conversion in the genus *Drosophila*. *Genetics*, 185, 95-103.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17, 540-552.

Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M. A., & Petrarca, V. (2002). A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science*, 298, 1415-1418.

Cuenca, J. B., Galindo, M. I., Saura, A. O., Sorsa, V., & de Frutos, R. (1998). Ultrastructure of regions containing homologous loci in polytene chromosomes of *Drosophila melanogaster* and *Drosophila subobscura*. *Chromosoma*, 107, 113-126.

Delport, W., Poon, A. F., Frost, S. D., & Kosakovsky Pond, S. L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, 26, 2455-2457.

Deng, W., Maust, B. S., Nickle, D. C., Learn, G. H., Liu, Y., Heath, L., Kosakovsky Pond, S. L., & Mullins, J. I. (2010). DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. BioTechniques, 48, 405-408.

Dobzhansky, T. (1947). Genetics of natural populations. XIV. A response of certain gene arrangements in the third chromosome of *Drosophila pseudoobscura* to natural selection. *Genetics*, 32, 142–160.

Dobzhansky, T. G. (1970). *Genetics of the evolutionary process*. New York: Columbia University Press.

Evgen'ev, M. B., Garbuz, D. G., & Zatsepina, O. G. (2014). Heat Shock Proteins and Whole Body Adaptation to extreme Environments (218 pp.) Springer, Berlin-New-York-London.

Ezawa, K., Oota, S., & Saitou, N. (2006). Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Molecular Biology and Evolution*, 23, 927-940.

Fawcett, J. A., & Innan, H. (2011). Neutral and non-neutral evolution of duplicated genes with gene conversion. Genes, 2, 191-209.

Fernández-Palacios, J. M., de Nascimento, L., Otto, R., Delgado, J. D., García-del-Rey, E., Arévalo, J. R., & Whittaker, R. J. (2011). A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography*, 38, 226-246.

Garbuz, D. G., & Evgen'ev, M. B. (2017). The evolution of heat shock genes and expression patterns of heat shock proteins in the species from temperature contrasting habitats. *Russian Journal of Genetics*, 53, 21-38.

García Guerreiro, M. P., & Fontdeila, A. (2007). The evolutionary history of *Drosophila buzzatii*. XXXVI. Molecular structural analysis of Osvaldo retrotransposon insertions in colonizing populations unveils drift effects in founder events. *Genetics*, 171, 301-310.

González, A. M., Cabrera, V. M., Larruga, J. M., & Gullón, A. (1983). Molecular variation in insular endemic *Drosophila* species of the Macaronesian archipelagos. *Evolution*, 37, 1128-1140.

Grant, P. R., Grant, B. R., Huey, R. B., Johnson, M. T. J., Knoll, A. H., & Schmitt, J. (2017). Evolution caused by extreme events. *Philosophical Transactions of the Royal Society B*, 372, 20160146.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59, 307-321.

Hashikawa, N., Mizukami, Y., Imazu, H., & Sakurai, H. (2006). Mutated yeast heat shock transcription factor activates transcription independently of hyperphosphorylation. *Journal of Biological Chemistry*, 281, 3936-3942.

Herrig, D. K., Modrick, A. J., Brud, E., & Llopart, A. (2013). Introgression in the *Drosophila subobscura—D. madeirensis* sister species: evidence of gene flow in nuclear genes despite mitochondrial differentiation. *Evolutio*n, 68, 705–719.

Hoekstra, L. A., & Montooth, K. L. (2013). Inducing extra copies of the *Hsp70* gene in *Drosophila melanogaster* increases energetic demand. *BMC Evolutionary Biology*, 13, 68.

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation?. *Annual Review of Ecology, Evolution, and Systematics*, 39, 21-42.

Hoffmann, A. A., Sgrò, C. M., & Weeks, A. R. (2004). Chromosomal inversion polymorphisms and adaptation. *Trends in Ecology and Evolution*, 19, 482-488.

Hughes, L. (2000). Biological consequences of global warming: is the signal already apparent? *Trends in Ecology and Evolution*, 15, 56-61.

Huson, D. H., & Bryant, I. D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254-67.

IPCC. (2001). Climate Change 2001: Impacts, Adaptation, and Vulnerability. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.

Jedlicka, P., Mortin, M. A., & Wu, C. (1997). Multiple functions of *Drosophila* heat shock transcription factor in vivo. *EMBO Journal* 16, 2452-2462.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, 110, 462-467.

Kapun, M., Fabian, D. K., Goudet, J., Flatt T. (2016). Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution*, 33, 1317-1336.

Karpov, V. L., Preobrazhenskaya, O. V., & Mirzabekov, AD. (1984). Chromatin structure of *hsp 70* genes, activated by heat shock: selective removal of histones from the coding region and their absence from the 5' region. *Cell*, 36, 423-431.

Khadem, M., & Krimbas, C. B. (1991). Studies of the species barrier between *Drosophila subobscura* and *D. madeirensis* I. The genetics of male hybrid sterility. *Heredity*, *67*, 157–165.

Khadem, M., Munté, A., Camacho, R., Aguadé, M., & Segarra, C. (2012). Multilocus analysis of nucleotide variation in *Drosophila madeirensis*, an endemic species of the Laurisilva forest in Madeira. *Journal of Evolutionary Biology*, 25, 726-739.

Khadem, M., Rozas, J., Segarra, C., Brehm, A., & Aguadé, M. (1998). Tracing the colonization of Madeira and the Canary Islands by *Drosophila subobscura* through the study of the *rp49* gene region. *Journal of Evolutionary Biology*, 11, 439-452.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589.

Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, 33, 511-518.

Kent, W. J. (2002). BLAT – The BLAST-like alignment tool. *Genome Research*, 12, 656-664.

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, 8, e1000501.

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, 173, 419-434.

Korunes, K. L., & Noor, M. A. F. (2017). Gene conversion and linkage: effects on genome evolution and speciation. *Molecular Ecology*, 26, 351-364.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). *Molecular Biology and Evolution*, 23, 1891–1901.

Krimbas, C. B. (1992). The inversion polymorphism of *Drosophila subobscura*. In C. B. Krimbas and J. R. Powell (Ed.), *Drosophila Inversion Polymorphism* (pp. 128-220). CRC Press, Boca Raton, FL.

Krimbas, C. B., & Loukas, M. (1984). Evolution of the *obscura* group drosophila species. I. Salivary chromosomes and quantitative characters in *D. subobscura* and two closely related species. *Heredity*, 53, 469–482.

Kuhn, E.J., Hart, C. M., & Geyer, P. K. (2004). Studies of the role of the *Drosophila scs* and *scs*' insulators in defining boundaries of a chromosome puff. *Molecular Biology and Evolution*, 4, 1470-1480.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870-1874.

Kunze-Mühl, E., & Müller, E. W. (1958). Untersuchungen über die chromosomale Struktur und die natürlichen Strukturtypen von *Drosophila subobscura* Coll. *Chromosoma*, 9, 559–570.

Labrador, M., Naveira, H., & Fontdevila, A. (1990). Genetic mapping of the *Adh* locus in the repleta group of *Drosophila* by in situ hybridization . *Journal of heredity*, 81, 83-86.

Larruga, J. M., Cabrera, V. M., González, A. M., & Gullón, A. (1983). Molecular and chromosomal polymorphism in continental and insular populations from the southwestern range of *Drosophila subobscura. Genetica, 60,* 191–205.

Lerman, D. N., Michalak, P., Helin, A. B., Bettencourt, B. R., & Feder, M. E. (2003). Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. *Molecular Biology and Evolution*, 20, 135-144.

Levitan, M., & Etges, W. J. (2005). Climate change and recent genetic flux in populations of *Drosophila robusta*. *BMC Evolutionary Biology*, 5, Art, No. 4.

Liao, D. Q. (1999). Concerted evolution: Molecular mechanism and biological implications. *American Journal of Human Genetics*, 64, 24-30.

Loukas, M., Krimbas, C. B., Mavragani-Tsipidou, P., & Kastritsis, D. (1979). Genetics of *Drosophila subobscura* populations: VIII. Allozyme loci and their chromosome maps. *Journal of Heredity*, 70, 17–26.

Llopart, A., & Aguadé, M. (1999). Synonymous rates at the *RpII215* gene of *Drosophila*: Variation among species and across the coding region. *Genetics*, 152, 269-280.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* 1, vev003.

Martin, D., & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, 16, 562-563.

Maynard Smith J. (1992). Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, 34, 126-129.

Menozzi, P., & Krimbas, C. B. (1992). The inversion polymorphism of *D. subobscura* revisited: Synthetic maps of gene arrangement frequencies and their interpretation. *Journal of Evolutionary Biology*, 5, 625-641.

Miller, W. J., Nagel, A., Bachmann, J., & Bachmann, L. (2000). Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. *Molecular Biology and Evolution*, 17, 1597-1609.

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30, 1188-1195.

Monclús, M. (1976). Distribución y ecología de drosofílidos en España. II. Especies de *Drosophila* de las Islas Canarias, con la descripcion de una nueva especie. *Boletín de la Real Sociedad Española de Historia Natural. Sección Biológica*, 74, 197-213.

Moltó, M. D., de Frutos, R., & Martínez-Sebastián, M. J. (1988). Gene activity of polytene chromosomes in *Drosophila* species of the *obscura* group. *Chromosoma*, 96, 382-390.

Moltó, M. D., Pascual, L., & de Frutos, R. (1987). Puff activity after heat shock in two species of the *Drosophila obscura* group. *Experientia*, 4, 1225-1227.

Moltó, M. D., Pascual, L., Martínez-Sebastián, M. J., & de Frutos, R. (1992). Genetic analysis of heat shock response in three *Drosophila* species of the *obscura* group. *Genome*, 35, 870-880.

Munté, A., Rozas, J., Aguadé, M., & Segarra, C. (2005). Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. *Genetics*, 169, 1573-1581.

de Nascimento, L., Willis, K. J., Fernández-Palacios, J. M., Criado, C., & Whittaker, R. J. (2009). The long-term ecology of the lost forest of La Laguna, Tenerife (Canary Islands). *Journal of Biogeography*, 36, 499-514.

Navarro, A., Betrán, E., Barbadilla, A., & Ruiz, A. (1997). Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics*, 146, 695-709.

Navarro-Sabaté, A., Aguadé, M., & Segarra, C. (1999). The relationship between allozyme and chromosomal polymorphism inferred from nucleotide variation at the *Acph-1* gene region of *Drosophila subobscura*. *Genetics*, 153, 2 871-889.

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, 3, 418–426.

Nikolaidis, N., & Nei, M. (2004). Concerted and nonconcerted evolution of the *Hsp70* gene superfamily in two sibling species of nematodes. *Molecular Biology and Evolution*, 21, 498-505.

O'Brien, T., Wilkins, R. C., Giardina, C., & Lis, J. T. (1995). Distribution of GAGA protein on Drosophila genes in vivo. *Genes & Development*, 9, 1098-1110.

Osada, N., & Innan, H. (2008). Duplication and gene conversion in the *Drosophila melanogaster* genome. *PLOS Genetics* , 4, e1000305.

Padidam, M., Sawyer, S., & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology*, 265, 218-225.

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37-42.

Parsell, D. A., & Lindquist, S. (1993). The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annual Review of Genetics*, 27, 437-496.

Pegueroles, C., Aquadro, C. F., Mestres, F., & Pascual, M. (2013). Gene flow and gene flux shape evolutionary patterns of variation in *Drosophila subobscura*. *Heredity*, 110, 520–529.

Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D., & Pupko, T. (2010). GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Research*, 38, W23-W28.

Pérez, J. A., Munté, A., Rozas, J., Segarra, C., & Aguadé, M. (2003). Nucleotide polymorphism in the *RpII215* gene region of the insular species *Drosophila guanche*: reduced efficacy of weak selection on synonymous variation. *Molecular Biology and Evolution*, 20, 1867–1875.

Perisic, O., Xiao, H., & Lis, T. T. (1989). Stable binding of *Drosophila* heat shock factor to heat-to-head and tail-to-tail repeats of a conserved Sbp recognition unit. *Cell*, 59, 797-806.

Petesch, S. J., Lis, J. T. (2008). Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at *Hsp70* loci. *Cell*, 134, 16-18.

Piñol, J., Francino, O., Fontdevila, A., & Cabré, O. (1988). Rapid isolation of *Drosophila* high molecular weight DNA to obtain genomic libraries *Nucleic Acids Research*, 16, 2736-2736.

Posada, D., & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 13757-13762.

Ramos-Onsins, S., Segarra, C., Rozas, J., & *Aguadé*, M. (1998). Molecular and chromosomal phylogeny in the obscura group of *Drosophila* inferred from sequences of the rp49 gene region. *Molecular Phylogenetics and Evolution*, 9, 33-41.

Rego, C., Balanyà, J., Fragata, I., Matos, M., Rezende, E. L., & Santos, M. (2010). Clinal patterns of chromosomal inversion polymorphisms in Drosophila subobscura are partly associated with thermal preferences and heat stress resistance. *Evolution*, 64, 385–397.

Rego, C; Matos, M; Santos, M. (2006). Symmetry breaking in interspecific *Drosophila* hybrids is not due to developmental noise. *Evolution*, 60, 746-761.

Rezende, E. L., Balanyà, J., Rodríguez-Trelles, F., Rego, C., Fragata, I., Matos, M., Serra, L. & Santos, M. (2010). Climate change and chromosomal inversions in *Drosophila subobscura*. *Climate Research*, 43, 103-114.

Rozas, J., Segarra, C., *Ribó*, G., & *Aguadé*, M. (1999). Molecular population genetics of the *rp49* gene region in different chromosomal inversions of *Drosophila subobscura*. *Genetics*, 151, 189-202.

Rodríguez-Trelles, F., Alvarez, G., & Zapata, C. 1996.Time-series analysis of seasonal changes of the O inversion polymorphism of *Drosophila subobscura*. *Genetics*, 142: 179-187.

Rodríguez-Trelles, F., & Rodríguez, M. A. (1998). Rapid micro-evolution and loss of chromosomal diversity in *Drosophila* in response to climate warming. *Evolutionary Ecology*, 12, 829-838.

Rodríguez-Trelles, F., & Rodríguez, M. A. 2007 Comment on 'Global genetic change tracks global climate warming in *Drosophila subobscura*', *Science* 315

Rodríguez-Trelles, F., & Rodríguez, M. A. 2010. Measuring evolutionary responses to global warming: cautionary lessons from *Drosophila*. *Insect Conservation and Diversity*, 3: 44-50.

Rodríguez-Trelles, F., Tarrío, R., & Santos, M. 2013. Genome-wide evolutionary response to a heat wave in *Drosophila*. *Biology Letters*, 9:e20130228.

Sánchez-Gracia, A., & Rozas, J. (2011). Molecular population genetics of the OBP83 genomic region in *Drosophila subobscura* and *D. guanche*: contrasting the effects of natural selection and gene arrangement expansion in the patterns of nucleotide variation. *Heredity*, 106, 191-201.

Sawyer, S. (1989). Statistical test for detecting gene conversion. *Molecular Biology and Evolution*, 6, 526-538.

Schmidt, E. R. (2002). A simplified and efficient protocol for nonradioactive in situ hybridization to polytene chromosomnes with a DIG-labeled DNA probe. In Non-radioactive In situ hybridization Aplication Manual. Third Edition. Edited by: Grünewald-Janho S, Keesey J, Leous M,van Miltenburg R, Schroeder C. Mannheim , Roche; pp. 108-111.

Sela, I., Ashkenazy, H., Katoh, K., & Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, 43, W7-W14.

Shilova, V. Y., Garbuz, D. G., Myasyankina, E. N., Chen, B., Evgen'ev, M. B., Feder M. E., & Zatsepina, O. G. (2006). Remarkable site specificity of local transposition into the *Hsp70* promoter of *Drosophila melanogaster*. *Genetics*, 173, 809-820.

Smit, A. F. A., Hubley, R., & Green, P. 2015. RepeatMasker Open-4.0.

Sperlich, D., Feuerbach-Mravlag, H., Lange, P., Michaelidis, A., & Pentzos-Daponte, A. (1977). Genetic load and viability distribution in central and marginal populations of *Drosophila subobscura*. *Genetics*, 86, 4 835-848.

Sugino, R. P., & Innan, H. (2006). Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. *Trends in Genetics*, 22, 642–644.

Sun, S., Evans, B. J., & Golding, G. B. (2011). "Patchy-tachy" leads to false positives for recombination. *Molecular Biology and Evolution*, 28, 2549-2559.

Tamura, K., Subramanian, S., & Kumar, S. (2004). Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Molecular Biology and Evolution*, 21, 36-44.

Tarrío, R., Rodríguez-Trelles F., & Ayala, F. J. (2001). Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the drosophilidae. *Molecular Biology and Evolution*, 8, 1464-1473.

Tian, S., Haney, R. A., & Feder M. E. (2010). Phylogeny disambiguates the evolution of heat-shock *cis*-regulatory elements in *Drosophila.Plos ONE* , 5, e1066.

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44, W232-W235.

Tsukiyama, T., Becker, P. B., & Wu, C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature*, 367, 525 – 532.

Udvardy, A., Maine, E., & Schedl, P. (1985). The 87A7 chromomere. Iden-tification of novel chromatin structures flanking the heat-shock locus that may define the boundaries of higher order domains. *Journal of Molecular Biology*, 185, 341– 358.

Umina, P. A., Weeks, A. R., Kearney, M. R., McKechnie, S. W., & Hoffmann, A. A. (2005) A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science*, 308, 691–693.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T. Geurts, R, & Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, 35, W71-W74.

Walsh, J. B. (1987). Sequence-dependent gene conversion: Can duplicated genes diverge fast enough to escape conversion?. *Genetics*, 117, 543-557.

Wilkins, R. C., & Lis, J. T. (1997). Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. *Nucleic Acids Research*, 25, 3963-3968.

Woolfit, M., & Bromham, L. (2005). Population size and molecular evolution on islands. *Proceedings of the Royal Society B-Biological Sciences*, 272, 2277-2282.

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, 24, 1586–1591.

Zatsepina, O. G., Velikodvorskaia, V. V., Molodtsov, V. B., Garbuz, D., Lerman, D. N., Bettencourt, B. R., Evgen'ev, M. B. (2001). A *Drosophila Melanogaster* strain from sub-equatorial Africa has exceptional thermotolerance but decreased *Hsp70* expression. *Journal of Experimental Biology*, 204, 1869-1881.

## Data Accessibility

Nucleotide sequences are deposited in GenBank under accessions: MG780233-MG780239

## Author Contributions

M.P.G.G., R.T. and F.R-T. designed the research. M.P.G., M.P.G.G., R.T. and F.R.-T., performed the research. M.P.G, R.T. and F.R.-T., analyzed the data. M.P.G., M.S., F.J.A., R.T. and F.R.-T. wrote the paper.

**Tables**

**Table 1**. Optimal model selection for the evolution of the *Hsp70IR* nucleotide region (NUC) in the *Drosophila subobscura* species cluster. Shown are the unpartitioned, *a priori*, and ModelFinder best-merging partitioning schemes with their corresponding numbers of partitions (*P*), and the log-likelihood (lnL) and Bayesian Information Criterion (BIC) scores, and the number of parameters (*k*) of each model (the specific models and characters of the partitions are provided in Table S1).

| Data set | Partitioning scheme | P | lnL | BIC | *k* |
|---|---|---|---|---|---|
| NUC | Unpartitioned | 1 | -13,198.93 | 26,548.48 | 17 |
| | A priori | 14 | -12,449.37 | 26,714.97 | 205 |
| | Best merging | 4 | -12,557.49 | 25,655.41 | 61 |
| Mixed NUC + PROT | A priori | 8 | -10,032.84 | 20,969.28 | 108 |
| | Best merging | 2 | -10,120.10 | 20,466.10 | 27 |
| Mixed NUC + CODON | A priori | 8 | -12,429.85 | 25,964.10 | 132 |
| | Best merging | 2 | -12,314.87 | 25,382.74 | 90 |

**Table 2**. Nei & Gojobori (1986) pairwise synonymous (Ks; above diagonal) and nonsynonymous (Ka; below diagonal) distances between coding sequences of *Hsp70* ortholog and paralog genes in species and inversions of the *subobscura* cluster ($O_{ST}$, $O_{3+4}$, $O_{3+4+8}$ and $O_{3+4+16}$ gene arrangements are from *D. subobscura*; Dm: *D. madeirensis*; Dg: *D. guanche*).

| | | Copy A | | | | | | | Copy B | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $O_{ST1}$ | $O_{ST2}$ | $O_{3+4}$ | $O_{3+4+8}$ | $O_{3+4+16}$ | Dm | Dg | $O_{ST1}$ | $O_{ST2}$ | $O_{3+4}$ | $O_{3+4+8}$ | $O_{3+4+16}$ | Dm | Dg |
| Copy A | $O_{ST1}$ | – | 12.5 | 50.1 | 50.1 | 39.3 | 25.1 | 104.9 | 7.2 | 8.9 | 39.3 | 50.1 | 41.1 | 26.9 | 73.5 |
| | $O_{ST2}$ | 0.0 | – | 41.1 | 41.1 | 30.4 | 19.7 | 101.3 | 12.5 | 3.6 | 30.4 | 41.1 | 35.7 | 21.5 | 68.1 |
| | $O_{3+4}$ | 1.5 | 1.5 | – | 21.5 | 14.3 | 46.5 | 110.3 | 46.5 | 44.7 | 21.5 | 14.3 | 12.5 | 48.3 | 69.9 |
| | $O_{3+4+8}$ | 2.9 | 2.9 | 1.5 | – | 17.9 | 50.1 | 115.7 | 46.5 | 44.7 | 25.0 | 14.3 | 30.4 | 51.9 | 77.1 |
| | $O_{3+4+16}$ | 1.5 | 1.5 | 0.0 | 1.5 | – | 39.4 | 108.5 | 35.8 | 34.0 | 10.7 | 14.3 | 12.5 | 41.2 | 64.5 |
| | Dm | 2.2 | 2.2 | 0.7 | 2.2 | 0.7 | – | 97.8 | 21.5 | 19.7 | 39.4 | 50.1 | 37.6 | 9.0 | 70.0 |
| | Dg | 11.3 | 11.3 | 9.9 | 11.3 | 9.9 | 10.6 | – | 103.1 | 101.4 | 110.3 | 113.9 | 108.5 | 96.1 | 62.0 |
| Copy B | $O_{ST1}$ | 0.0 | 0.0 | 1.5 | 2.9 | 1.5 | 2.2 | 11.3 | – | 8.9 | 35.8 | 46.5 | 37.5 | 23.3 | 69.9 |
| | $O_{ST2}$ | 0.7 | 0.7 | 2.2 | 3.7 | 2.2 | 2.9 | 12.1 | 0.7 | – | 34.0 | 44.7 | 35.8 | 21.5 | 68.1 |
| | $O_{3+4}$ | 1.5 | 1.5 | 0.0 | 1.5 | 0.0 | 0.7 | 9.9 | 1.5 | 2.2 | – | 21.5 | 16.1 | 41.2 | 66.3 |
| | $O_{3+4+8}$ | 2.9 | 2.9 | 1.5 | 1.5 | 1.5 | 2.2 | 10.6 | 2.9 | 3.7 | 1.5 | – | 26.8 | 51.0 | 69.9 |
| | $O_{3+4+16}$ | 1.5 | 1.5 | 0.0 | 1.5 | 0.0 | 0.7 | 9.9 | 1.5 | 2.2 | 0.0 | 1.5 | – | 39.4 | 68.1 |
| | Dm | 2.9 | 2.9 | 1.5 | 2.9 | 1.5 | 2.2 | 11.3 | 2.9 | 3.7 | 1.5 | 3.3 | 1.5 | – | 71.8 |
| | Dg | 9.5 | 9.5 | 8.0 | 9.5 | 8.0 | 8.8 | 11.3 | 9.5 | 10.2 | 8.0 | 8.0 | 8.0 | 9.5 | – |

**Figure Legends**

**Figure 1**. Phylogeny of O chromosomal inversions in the *Drosophila subobscura* subgroup. The cladogram on the left depicts the evolutionary history of chromosomal rearrangement on the standard cytological map of Kunze-Mühl & Müller (1958) shown on the right. The hypothetical ancestral gene order is $O_3$, on which inversion *g* (orange) arose in *D. guanche*, originating the $O_{3+g}$ gene arrangement, and inversions ST (blue) and 4 (red) arose independently in *D. subobscura*, originating the gene arrangements $O_{ST}$ and $O_{3+4}$. Inversion 3 went extinct in *D. guanche* and *D. subobscura*. In this latter species, inversions 8 and 16 arose independently on $O_{3+4}$, originating the complex gene arrangements $O_{3+4+8}$ (green) and $O_{3+4+16}$ (pink), which together with $O_{3+4}$ represent the $O_{3+4}$ phylad. *D. madeirensis* is currently monomorphic for $O_3$. Arrows indicate the location of the *Hsp70* locus (subsection 94A) on the chromosomal rearrangement background.

**Figure 2**. Genomic organization and rates of nucleotide substitution in the *Hsp70* locus. *Upper panel*: the boxes represent the different functional elements, including coding (*Hsp70A* and *Hsp70B*, and their upstream *CG5608* and downstream *Dmt* flanking genes) and noncoding (3' and 5' flanking regions and the CIR) elements drawn to scale. Arrow boxes indicate the sense of transcription. Horizontal brackets denote the edges of recognizable homology between duplicated blocks, and black boxes inside the span of the brackets denote the region of the repeats affected by concerted evolution. *Middle panel*: the numbers are base pair lengths across the species and inversions of the corresponding functional elements on the upper panel. Notice that, 3' lengths are referred to the segments of the flanking regions included in the brackets on the upper panel. *Lower panel*: the heights of the vertical bars are estimates of the rates of nucleotide substitution in the functional elements on the upper panel. Rates were estimated assuming the tree topology in Figure 1 and the TN93+C model with the BaseML program of the PAML v.4.9d package (Yang 2007), and were scaled to the rate of substitution in the 5' flanking region of *Hsp70B*. "1+2" indicate first and second codon positions, and "3" third codon position of the *Hsp70* genes.

**Figure 3**. Schematic representation of a MSA of the 5' cis-regulatory region of the *Hsp70* ortholog and paralog genes from the *subobscura* subgroup. Boxes on the top of the MSA denote different cis-regulatory elements, including four HSEs (1-4; pink), GAGA sites (G+: GAGAGAG, G-: CTCTCTC; yellow), and the core promoter (green) with underlined TATA-box and TSS. Colored background columns highlight the central three nucleotides of each HSE pentanucleotide motif (black), nucleotide variants that differ between paralogs (red), and nucleotide variants that are shared between paralogs allegedly as a result of gene conversion (dark green).

**Figure 4**. Unrooted ML tree of the phylogenetic relationships among species and chromosomal inversions of the *subobscura* cluster based on the best BIC model from Table 1 (see also Table S1). The model allows for two partitions to accommodate differences in the substitution process between noncoding nucleotide (HKY+I) and amino acid (JTT) characters. Branch lengths are proportional to the scale given in character substitutions per site, where character is either nucleotide or amino acid. Numbers at nodes represent, respectively from left to right, SH-aLRT, aBayes, and UFBoot branch support values based on 1000 replicates.

**Figure 5**. Spatial distribution patterns of the average pairwise proportion of nucleotide differences between duplicated blocks from the same chromosome (continuous line) and between orthologs (*D. guanche* vs the rest; dashed line) obtained using a non-overlapping window analysis of size 100 bp. Boxes on the top denote functional elements, including from left to right HSE 4 to 1, and the Hsp70 coding region. The arrow box indicates the sense of transcription.

**Figure 6**. Phylogenetic network of *Hsp70* orthologous and paralogous sequences from the species and chromosomal arrangements of the *subobscura* subgroup. The split network was constructed using the NeighborNet method as implemented in SplitsTree v.4.14.5 (Huson & Bryant 2006), on TN93+I (% of invariable sites 85.1) model, which is the best-fit model from the unpartitioned data (Tables S2-S3, Supporting Information), distances obtained using the DIVEIN web server (https://indra.mullins.microbiol.washington.edu/DIVEIN/) (Deng *et al*. 2010). Sets of parallel edges represent conflicting topological signals.

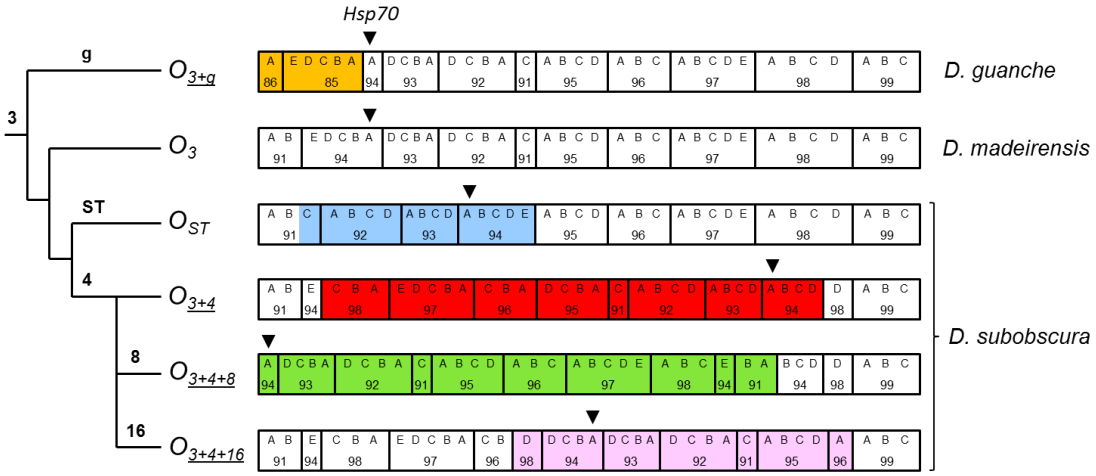


Figure 1

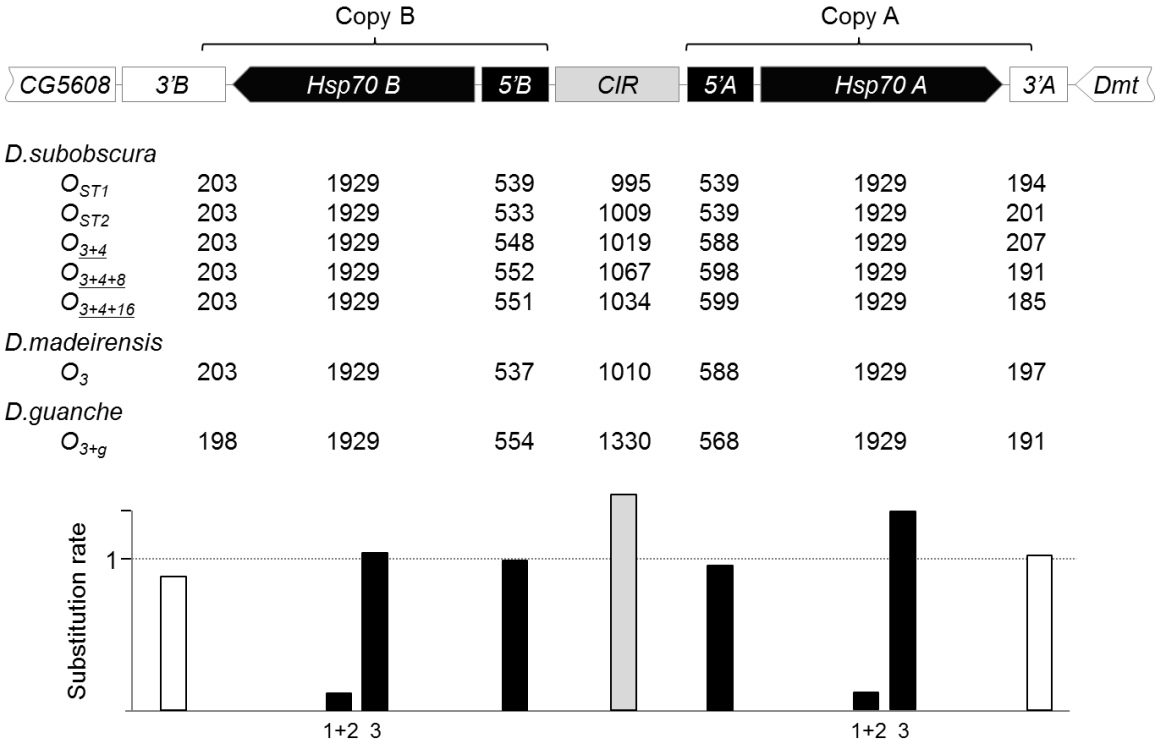


Figure 2

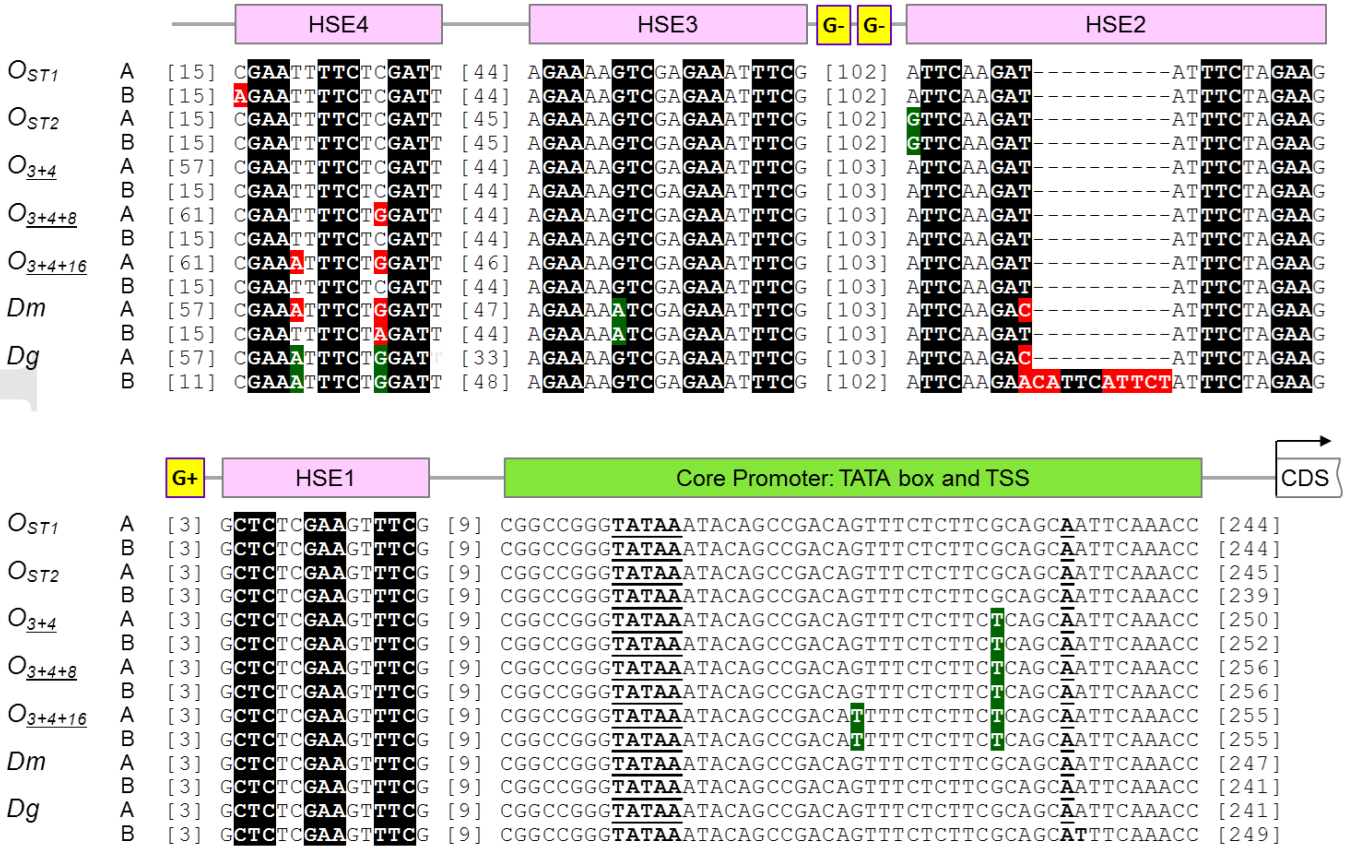


Figure 3

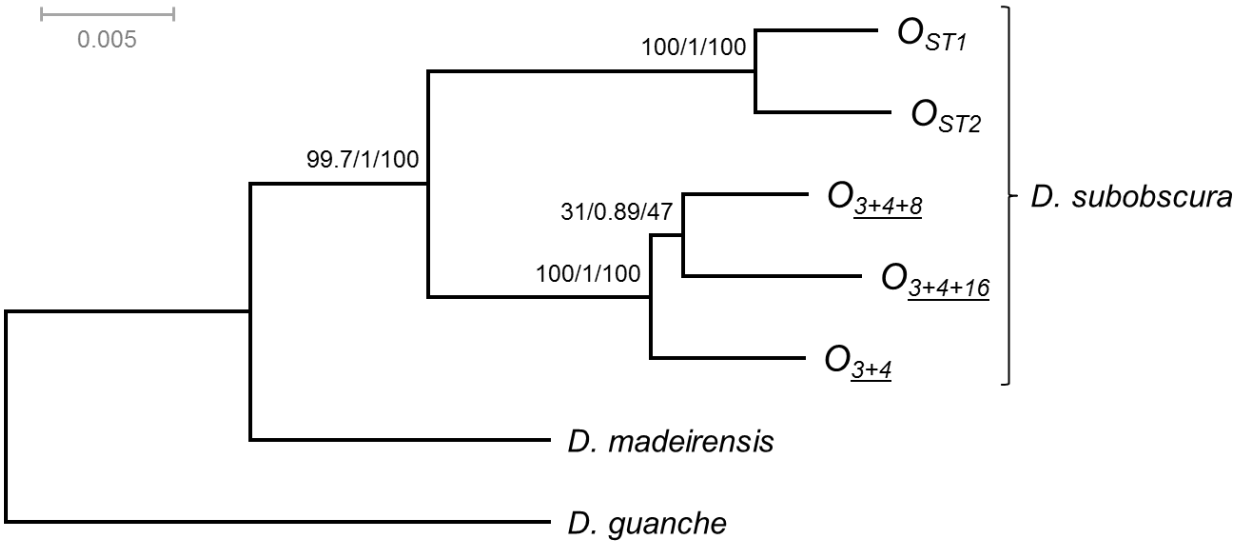


Figure 4

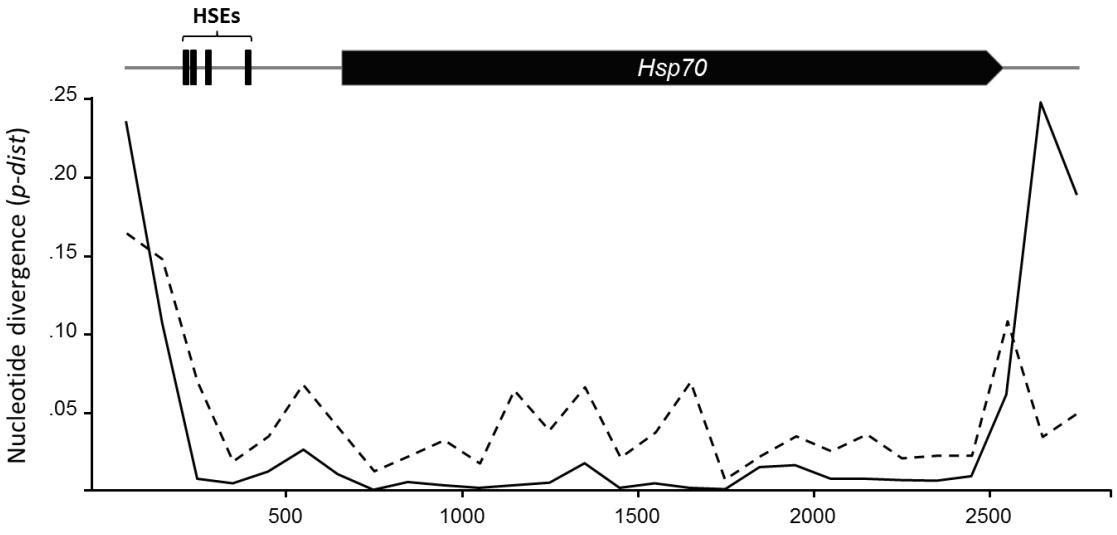


Figure 5

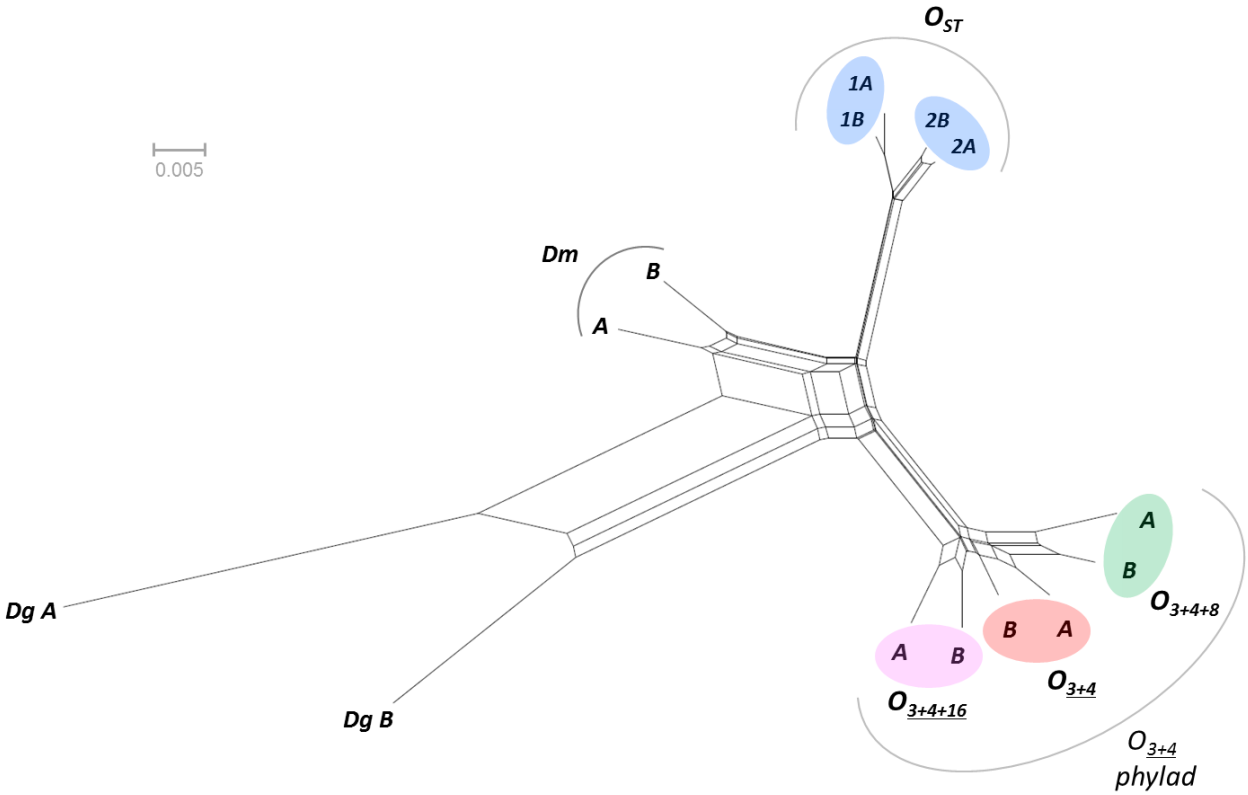


Figure 6