Evolution of the hsp70 gene family at the nucleotide, genome organization and gene expression levels in Drosophila subobscura

Marta Puig Giribets

Evolution of the *hsp70* gene family at the nucleotide, genome organization and gene expression levels in *Drosophila subobscura*.

Marta Puig Giribets

Departament de Genètica i Microbiologia

Universitat Autònoma de Barcelona

Thesis submitted for the degree of Doctor in Genetics (PhD)

Directors: Dr. Maria Pilar García Guerreiro
Dr. Francisco José Rodríguez-Trelles Astruga
Dr. Rosa María Tarrío Fernández

Mentor: Dr. Rosa María Tarrío Fernández

June 2018

Doctor and Associate Professor **Mª Pilar García Guerreiro**, Doctor and Associate Professor **Francisco José Rodríguez-Trelles Astuga** and Doctor and Adjunct Lecturer **Rosa Mª Tarrío Fernández** of the Department of Genetics and Microbiology, the Faculty of Biosciences of the Autonomous University of Barcelona,

Certify that: **Marta Puig Giribets** has conducted under their direction the research work in the Department of Genetics and Microbiology, the Faculty of Biosciences of the Autonomous University of Barcelona, which has led to the development of the Doctoral Thesis entitled: Evolution of the *hsp70* gene family at the nucleotide, genome organization and gene expression levels in *Drosophila subobscura*.

Dr. Mª Pilar García Guerreiro	**Dr. Francisco Rodríguez-Trelles**	**Dr. Rosa Tarrío**
Associate Professor	Associate Professor	Adjunct Lecturer
Faculty of Biosciences	Faculty of Biosciences	Faculty of Biosciences
Autonomous University of Barcelona	Autonomous University of Barcelona	Autonomous University of Barcelona

Per al *cugino*, que segurament no entendrà gran cosa però sempre hi és...

‘Aut viam inveniam aut faciam’
(*Trobaré el camí o el faré*)

Proverbi llatí

Agraïments

Diuen que el dia de la tesi és un dels més importants de les nostres vides. La meva racionalitat científica fa que no pugui creure en dies més importants que d'altres, però tot i així, intentaré fer l'esforç per unes hores.

Comencem pel principi de la història. Poc abans de finalitzar la carrera de Biologia, que havia començat feia quatre anys, vaig anar a veure el Dr. Francisco Rodríguez-Trelles, professor de l'assignatura 'Evolució', una de les meves preferides. L'evolució era una branca de la biologia que m'apassionava de feia temps i que volia investigar més a fons, i vaig pensar que anant al seu despatx tindria alguna oportunitat d'expandir els meus coneixements en aquest camp. Ell em va dirigir a la Dra. Mª Pilar García-Guerreiro, qui potser tindria una plaça per oferir-me. Recordo una de les primeres converses que vaig tenir amb ella, en què em va preguntar: "i tu, què voldries investigar?" – i jo vaig contestar sense gairebé ni pensar-m'ho "les seqüències!". En aquell moment no sabia que el meu desig es faria realitat molt poc temps després.

La Pilar em va explicar que per poder *investigar* calia que fes primer el màster en Genètica Avançada i que després, en finalitzar-lo, comencés el doctorat. Em va semblar un pla interessant; de fet, no volia començar a treballar de seguida sense haver acumulat més coneixements, i el doctorat era una de les poques vies que m'oferia aquesta opció. Tampoc no sabia si era la millor opció, però donades les circumstàncies d'aquell moment, vaig decidir acceptar-la i més endavant ja tindria temps de decidir si la investigació era per mi o no. Desitjava de tot cor que fos que sí, perquè sempre m'havia fascinat la idea de poder dedicar-me a la recerca, però al mateix temps n'era conscient de la manca de beques i els alts requisits (en gairebé tots els camps) que se'ns exigien als estudiants per poder-nos dedicar a la investigació, un sector no massa ben valorat al nostre país ara per ara.

El primer any de màster em va ensenyar a treballar pel meu compte, un hàbit que vaig descobrir que no tenia del tot desenvolupat malgrat els meus 25 anys i els sis que m'havia passat a la universitat (els dos primers, cursant la carrera de Psicologia que no vaig arribar a acabar mai). També vaig aprendre a conviure amb gent molt diferent de mi. Al principi això em va resultar difícil perquè em trobava que sovint no tenia les eines que necessitava per prendre la decisió correcta en cada moment. Aquesta habilitat, com tantes altres, me l'ha proporcionada l'experiència acumulada al llarg d'aquests últims cinc anys.

Els quatre anys de tesi han passat de manera ràpida. L'objectiu principal que he tingut present en tot moment ha estat el d'acabar el més aviat possible abans que finalitzés el meu contracte PIF, que té una durada finita de tres anys. Em preocupava el tema de no poder complir els terminis –com m'imagino que els passa a bona part dels doctorands– i sentia que malgastava el temps quan no em passava prou hores al laboratori; potser és per això que ara sento que he deixat oblidada una part important de la meva vida durant tots aquests anys. Potser ja està bé que fos així, o potser no. Ara tant és, perquè ja no puc tornar enrere. Recordo que els últims cinc anys han estat una lluita constant perquè els experiments sortissin, una lluita que no he volgut abandonar en cap moment malgrat

les nombroses batalles perdudes. Una lluita en què m'alçava cada cop que queia i tornava a començar com si res.

Arribats en aquest punt, crec que hauria d'agrair el suport proporcionat per totes les persones del meu voltant, i que d'alguna manera o d'una altra han contribuit a la realització d'aquesta tesi. Voldria començar per la meva directora de tesi, la Dra. Pilar, qui sense pràcticament conèixer-me, em va donar l'oportunitat de treballar al seu laboratori per tal d'endinsar-me en el món de la recerca i que ha fet possible la realització d'aquesta tesi. Li voldria agrair la confiança que va dipositar en mi des del primer moment i tot el suport –moral i pràctic– que m'ha donat durant tot aquest temps. Al mateix temps, també tinc una sèrie d'agraiments pels altres dos directors de tesi, el Dr. Francisco Rodríguez-Trelles i la Dra. Rosa Tarrío. Al Dr. Francisco li voldria agrair la motivació que m'ha inspirat en tot moment per continuar investigant, els incomptables coneixements que m'ha proporcionat, el seu tracte amable i empàtic i la seva enorme dedicació en la meva tesi. Així mateix, a la Dra. Rosa Tarrío voldria agrair-li el seu gran ajut en els experiments, la seva simpatia que dibuixava amb el més sincer dels seus somriures, els abundants coneixements científics que m'ha aportat i la seva contribució com a tutora d'aquesta tesi. També voldria agraïr a en Sebastián Ramos-Onsins la seva col·laboració en aquest treball, sense la qual no hagués pogut completar-lo. I no voldria acabar aquest paràgraf sense esmentar el Dr. Mauro Santos, el Dr. Julio Rozas i el Dr. Antoni Fontdevila que han contribuit també a expandir els meus coneixements científics.

A banda dels directors de tesi, també voldria agrair el suport dels meus companys de tesi, doctorands i estudiants de màster o de grau. A la Valèria li voldria agrair de tot cor que m'hagi ajudat indubtablement a reforçar la meva confiança i autoestima i el fet que ha estat una companya fora de sèrie en tots els aspectes. A en Víctor voldria comentar-li que també ha estat un excel·lent company, i que agraeixo profundament l'immens respecte i consideració que ha mostrat per les meves qualitats com a investigadora i com a persona. Després també voldria agrair a en Sergi els bons moments que hem passat junts fora del laboratori i a l'Alba el seu suport incondicional i la seva simpatia. No voldria oblidar pas la Hayley i el seu caràcter amable ni la seva generositat. Tampoc no oblidaré la Carmen, la companya que vaig conèixer durant el primer any de màster, que em va semblar una persona agradable, dedicada i intel·ligent. I per descomptat, vull esmentar també totes les persones, estudiants de grau, màster i doctors, amb qui he tingut l'oportunitat de coincidir més o menys breument durant aquests cinc anys: Joan, Eila, Inma, Raquel, Jone, Laia, Doris, Luz, Sergio i tants d'altres que probablement em descuido.

A banda dels companys de laboratori, també volia agrair el suport dels tècnics i altres treballadors del departament de Genètica. A la Lourdes voldria agrair-li tot l'ajut que m'ha proporcionat en els aspectes més quotidians del funcionament d'un laboratori, que no m'eren del tot coneguts quan vaig començar i que, gràcies a ella, me n'he pogut sortir. Voldria mostrar el meu agraiment a la Montse Peiró, una tècnic excel·lent amb qui malauradament només vaig coincidir els primers mesos, així com també a la Raquel Ferraz i a la Montse Sales pel seu ajut indispensable en el manteniment de les soques del laboratori i la seva paciència i dedicació per la seva feina. Voldria esmentar la Sònia per veure'm quan em sentia invisible. I finalment, voldria donar les gràcies a les secretàries del departament, Meritxell, Mª Josep, Maite i Elena, per assessorar-me en tot moment i per la seva eficàcia i eficiència en les seves tasques.

I would also like to thank all the people that guided me during my three month stay in Denmark and accepted me in their research group. They have not only made the experiments succesful, but also showed me how to work in a completely different envirnoment. Likewise, they made the experience positive and unique, a experience that I would not mind to repeat in the future: Jesper, Volker, Doth i Marie: *Mange tak!*

Deixant de banda el laboratori, no puc oblidar de donar les gràcies a tots els que d'alguna manera m'han fet costat durant tots aquests anys. Així mateix, voldria donar les gràcies a les 'Fronterizas': Rota, Hri i Roger (i a en Widi, que al cel sia) per haver convertit l'estada en un pis compartit de l'Hospitalet en una experiència única i inoblidable. També voldria agrair el suport de tots aquells amb qui he compartit bona part dels meus dies però que ja no hi són: tia Antònia, oncle Josep, avi Esteve, àvia Maria... i a la meva família (pares, germans, nebots, cosins, cunyats, i a la Bona, naturalment!) per tot el suport i ajuts proporcionats durant aquest temps. I a tu també encara que no ho arribis a llegir mai, Maria.

Per últim voldria donar les gràcies al *cugino*, per haver-me proporcionat bona part de les millors experiències d'aquests últims cinc anys i per haver-me animat a no desistir mai i per acceptar-me tal com sóc. A ell i a la seva família, que m'han tractat com a un més, els hi agraeixo de tot cor tots els bons moments.

Contents

ABSTRACT

Numerous studies have confirmed the adaptive value of the rich chromosomal inversion polymorphism in the drosophilid *D. subobscura*. However, until recently very little was known about the molecular basis behind its maintenance in natural populations. In this sense, the characterization of the thermally inducible *hsp70* gene family in this species could shed light on a hotly debated topic in *Drosophila* population genetics as is the adaptive inversion polymorphism. In *D. subobscura*, the *hsp70* gene family has been ascribed to a locus that is found in the distal segment (SI) of the O chromosome, a region where numerous chromosomal arrangements have been described. Some of them, such as the warm $O_{\underline{3+4}}$ and the cold O_{ST}, show a north-south clinal geographical distribution and opposed mid- and long-term seasonal variation patterns. In search of candidate loci, a previous heat shock experiment quantified Hsp70 protein levels in homokaryotypic strains for the O_{ST}, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4}}$ arrangements. Unexpectedly, individuals of the warm-adapted $O_{\underline{3+4}}$ arrangement showed increased levels in absence of thermal stress that did not boost after the heat shock. Unfortunately, by the time this experiment was performed there was very little data available on the molecular organization of the *Hsp70IR* locus in this and in other closely related species of the *obscura* group to explain these unpublished results.

The previously mentioned results led to the present thesis work, whose objectives are to locate the *Hsp70IR* locus in the karyotype and to know the genomic organization, molecular characteristics and gene expression patterns in several representative chromosomal arrangements that comprise the genomic region where the *hsp70* gene family is located: $O_{\underline{3+4+16}+\underline{2}}$, $O_{\underline{3+4+8}}$, $O_{\underline{3+4}}$ and O_{ST}. Using the sequence of a clone from an O_{ST} line genomic library and contigs from the unassembled genome of *D. subobscura*, we designed a probe from the *hsp70* coding region that enabled us to determine the location of the locus by *in situ* (ISH) hybridization. Concomitantly, we completed the sequencing of a 9-10 kb region in the *Hsp70IR* locus in 12 lines isogenic for the aforementioned arrangements and in *D. madeirensis* and *D. guanche* to shed light on the evolution of this locus in the last 1.8 - 2.8 million years (myr). Lastly, we quantified basal *hsp70* mRNA and protein levels in adult males and females wearing the cold-

adapted O_{ST} and the warm-adapted $O_{\underline{3+4}}$ arrangements to see if they differed between them.

ISH results showed a single hybridization site in the 94A band in the distal segment (SI) of the O chromosome coincident in the 4 studied karyotypes: $O_{\underline{3+4+16+2}}$, $O_{\underline{3+4+8}}$, $O_{\underline{3+4}}$ and O_{ST}. The sequences corresponding to the 12 isogenic lines and to *D. madeirensis* and *D. guanche* indicated that in these three species of the *subobscura* cluster, the *Hsp70IR* locus consists of two 2.5 - 3.0 kb long paralogous copies in divergent orientation separated by a 0.5 - 1.4 kb nonduplicated central spacer region. The two copies show a high degree of conservation between the different gene arrangements and species analyzed, while the central spacer region is highly polymorphic. This organization is similar to that found in *D. pseudoobscura* and *D. persimilis*, two species that diverged from the *subobscura* cluster 10-15 Myr ago.

Among the most relevant aspects of polymorphism analyses, we highlight the high degree of conservation in the coding regions (CDSs) and the *cis*-regulatory elements (CREs) in the proximal promoter of all the analyzed *hsp70* genes, which might indicate that these are functional in all studied lines, and that their regulation be similar. On the other hand, analyses of polymorphism in several discernible regions of the locus indicates that it fits a neutral model in the four previously mentioned arrangements. Although the variability patterns across the ~10 kb sequenced region are similar between the four arrangements, these are highly variable among the different regions, which is explained by the high selective constrainment in the coding regions of the two genes and their proximal promoters. Curiously, at the sequence level, the paralogous 5'-UTR and CDS regions tend to be significantly more similar within the same arrangement and, in some cases, within the same line. Taking into account that the ancestral duplication occurred before the radiation of the *obscura* group, 10-15 Myr ago, the most parsimonious explanation is that the two copies evolve in concert through ectopic recombination mechanisms in the four studied arrangements of *D. subobscura* and in *D. madeirensis*. In contrast, the divergence between the same regions of the two paralogous genes is much higher in *D. guanche*, which could indicate that the gene conversion mechanisms are not active in this species.

Lastly, we carried out the quantification of basal *hsp70* mRNA and protein levels in adult males and females of six lines isogenic for the cold-adapted O_{ST} and six for the

warm-adapted O_{3+4} arrangements. Basal mRNA quantification results indicate that the two arrangements exhibit similar levels, yet significant differences are observed between males and females of the warm O_{3+4} arrangement, where the mRNA levels of males are significantly lower than those of females. Regarding the quantification of basal Hsp70 protein levels, these suggest that there are no differences between sexes nor between the two arrangements, but instead we observe a significant interaction between sex and arrangement, where O_{ST} males tend to have higher levels than females while the opposite pattern is observed in O_{3+4}. Overall, the results for both, mRNA and protein data, indicate that *hsp70* expression might be influenced by sex. Finally, it should be stressed that our results do not match those of the previous study where higher basal Hsp70 protein levels had been noted for O_{3+4} males and females relative to their O_{ST} and O_{3+4+8} counterparts. Although our results are preliminary, the quantification of basal mRNA and protein levels together with the *in situ* hybridization experiments and the sequencing data indicate that the *Hsp70IR* locus is highly conserved in O_{ST} and O_{3+4}, and that other candidate loci must be searched to explain their adaptive character.

RESUM

Nombrosos estudis han constatat el valor adaptatiu del ric polimorfisme d'inversions cromosòmiques al drosofíl·lid *D. subobscura*. No obstant això, fins ara es coneixien molt poc les bases moleculars que hi ha darrere del seu manteniment a les poblacions naturals. En aquest sentit, la caracterització de la família gènica termoinduïble *hsp70* en aquesta espècie podria aportar-nos informació d'un tema prou debatut dins la genètica de poblacions de *Drosophila* com és el polimorfisme adaptatiu de les inversions. A *D. subobscura*, el locus on es troba aquesta família gènica es situa al segment distal (SI) del cromosoma O, una regió on han estat descrits nombrosos ordenaments cromosòmics. Alguns d'ells, com el càlid O_{3+4} i el fred O_{ST}, mostren una distribució geogràfica clinal nord-sud i patrons de variació estacional oposats a mitjà i llarg termini. En cercar loci candidats, un experiment previ de xoc tèrmic va quantificar els nivells de la proteïna Hsp70 en soques homocariotípiques dels ordenaments O_{ST}, O_{3+4+8} i O_{3+4}. Inesperadament, els individus de l'ordenament càlid O_{3+4} mostraven nivells incrementats d'aquesta proteïna, en absència d'estrès tèrmic, que no augmentaven després del xoc tèrmic. Malauradament, en el moment en què es va dur a terme

l'experiment hi havia moltes incògnites sobre l'organització molecular del locus *Hsp70IR* en aquesta i en d'altres espècies properes del grup *obscura,* per a poder donar una explicació a aquests resultats inèdits.

Els resultats prèviament esmentats, van donar peu al present treball de tesi, els objectius del qual són localitzar el locus *Hsp70IR* al cariotip i conèixer-ne l'organització genòmica, característiques moleculars i expressió gènica en diversos ordenaments cromosòmics d'interès que inclouen la regió genòmica on es troba la família gènica *hsp70*: $O_{3+4+16+2}$, O_{3+4+8}, O_{3+4} i $O_{ST.}$ Gràcies a la seqüència d'un clon de la genoteca d'una línia O_{ST} i a còntigs del genoma de *D. subobscura*, hem pogut dissenyar una sonda a partir de la regió codificant de *hsp70* que ens ha permès determinar la localització del locus on es troba aquesta família gènica (*Hsp70IR*) mitjançant hibridació *in situ* (ISH). Paral·lelament, hem pogut completar la seqüenciació d'una regió de 9-10 kb al locus *Hsp70IR* en 12 línies isogèniques per als ordenaments esmentats i a les espècies properes *D. madeirensis* i *D. guanche* per aclarir l'evolució d'aquest locus en els darrers 1,8 - 2,8 milions d'anys (Ma). Finalment, hem quantificat els nivells basals de mRNA i proteïna en mascles i femelles adults portadors de l'ordenament fred O_{ST} i el càlid O_{3+4} per veure si diferien significativament entre ells.

Els resultats de la ISH van mostrar un únic punt d'hibridació a la banda 94A del segment distal (SI) del cromosoma O que coincidia els 4 cariotips estudiats: $O_{3+4+16+2}$, O_{3+4+8}, O_{3+4} i O_{ST}. Les seqüències corresponents a les 12 línies isogèniques i a *D. madeirensis* i *D. guanche* indiquen que en aquestes tres espècies del clúster *subobscura*, el locus *Hsp70IR* consta de dues còpies paràloges de 2,5 – 3,0 kb en orientació divergent i separades per una regió espaiadora central no duplicada de 0,5 – 1,4 kb. Les dues còpies mostren un elevat grau de conservació entre els diferents ordenaments i espècies analitzats, mentre que la regió espaiadora central és altament polimòrfica. Aquesta organització seria similar a la que trobem a *D. pseudoobscura* i *D. persimilis*, dues espècies que van divergir del clúster *subobscura* fa 10-15 Ma.

Entre els aspectes més rellevants de l'anàlisi del polimorfisme, destaquem l'elevada conservació de les regions codificadores (CDSs) i els diferents elements reguladors en *cis* (CREs) al promotor proximal de tots els gens *hsp70* analitzats, que indicarien que aquests són funcionals a totes les línies estudiades, i que la seva regulació podria ser similar. D'altra banda, l'estudi del polimorfisme en diferents regions discernibles del

locus indica que aquest s'ajusta a les expectatives del model neutre en els quatre ordenaments prèviament esmentats. Tot i que els patrons de variabilitat al llarg de la regió seqüenciada de ~10 kb són similars entre els quatre ordenaments, aquests són altament variables entre les diferents regions, cosa que s'explica per l'elevat constrenyiment selectiu al qual es troben sotmeses les regions codificadores dels dos gens i el seu promotor proximal. Curiosament, a nivell de seqüència, les regions paràlogues del promotor proximal i el CDS tendeixen a ser significativament més similars dins del mateix ordenament i, en alguns casos, dins la mateixa línia. Tenint en compte que la duplicació ancestral es va produïr abans de la radiació del grup *obscura*, fa 10-15 Ma, l'explicació més parsimoniosa és que les dues còpies evolucionen concertadament mitjançant mecanismes de recombinació ectòpica en els quatre ordenaments estudiats de *D. subobscura* i a *D. madeirensis*. En canvi, la divergència entre aquestes mateixes regions dels dos gens paràlegs és molt més elevada a *D. guanche*, cosa que podria indicar que els mecanismes de conversió gènica no són actius en aquesta espècie.

Per últim, hem dut a terme la quantificació dels nivells basals de mRNA i proteïna en mascles i femelles adults de sis línies isogèniques per a l'ordenament fred O_{ST} i sis per a l'ordenament càlid $O_{\underline{3+4}}$. La quantificació dels nivells de mRNA indica que els nivells són similars entre els dos ordenaments però en canvi aquests difereixen entre mascles i femelles de l'ordenament càlid $O_{\underline{3+4}}$, on els nivells de mRNA dels mascles són significativament inferiors als de les femelles. Així mateix, la quantificació dels nivells de la proteïna Hsp70 suggereix que no hi ha diferències entre sexes ni entre els dos ordenaments, però en canvi observem una interacció significativa entre sexe i ordenament, on els nivells en mascles O_{ST} tendeixen a ser més elevats que en femelles i el patró invers s'observa a l'ordenament $O_{\underline{3+4}}$. Aquests resultats, tant per a mRNA com per a proteïna, indiquen que l'expressió de *hsp70* podria estar influïda pel sexe. Finalment, cal remarcar que les nostres dades no coincideixen amb les de l'estudi precedent on s'havien trobat nivells basals de la proteïna Hsp70 més elevats en mascles i femelles $O_{\underline{3+4}}$ que en O_{ST} i $O_{\underline{3+4+8}}$. Encara que els resultats del nostre estudi siguin preliminars, la quantificació dels nivells basals de mRNA i proteïna juntament amb els experiments d'hibridació *in situ* i les dades de seqüenciació indiquen que el locus *Hsp70IR* es troba altament conservat en O_{ST} i $O_{\underline{3+4}}$, i que cal cercar altres loci candidats per poder explicar-ne el caràcter adaptatiu.

1. INTRODUCTION

The introduction section has been divided into three parts. In the first part, I will describe what are chromosomal inversions, how do they form, the significance of inversion polymorphim in evolution and the current hypotheses to explain its maintenance in natural populations. In the second section, I will briefly clarify the classification, biology and ecology of *Drosophila subobscura*, a species that exhibits a rich inversion polymorphism across its five acrocentric chromosomes. I will provide evidence on the adaptive value of some *D. subobscura* inversions whose frequencies co-vary with diverse environmental variables, including temperature and latitude while mentioning studies of loci involved in thermal adaptation. In the last section, I will describe the highly conserved and thermally inducible *hsp70* gene family in the genus *Drosophila*, a candidate locus to explain the adaptive value of certain *D. subobscura* inversions in the O chromosome. Finally, I will try to relate the extant differences in *hsp70* copy number variation, *cis*-regulatory polymorphism and genomic organization in *Drosophila* with the contrasting patterns in Hsp70 protein levels, thermotolerance and thermal preference exhibited by *D. subobscura* strains homokaryotypic for certain adaptive chromosomal arrangements.

1.1. The importance of chromosomal inversion polymorphism in natural populations

'*Evidence is a very sticky thing... it may seem to point very straight to one thing, but if you shift your own point a little, you may find it pointing in an equally uncompromising manner to something entirely different.*'

Sherlock Holmes – The Boscombe Valley Mystery

1.1.1. What are chromosomal inversions?

Inversions are described as reversed end to end segments in the chromosome relative to a standard, non-inverted segment. They occur when a chromosome undergoes breakage and rearrangement within itself. Inversions can vary a lot in their size; the longest ones expand several megabases and involve thousands of loci, while others are small enough to be only detectable by direct comparison of DNA sequences. Inversions are classified as pericentric or paracentric. Pericentric inversions are those that comprise the centromere, while paracentric ones involve a segment in one chromosomal arm. Most of the described inversions in *Drosophila* are paracentric (Powell 1997). One consequence of inversions is the inhibited recombination in the heterokaryotype because the inverted segment is unable to make a stable synaptic interaction with the standard one. When recombination does occur, it results in a high proportion of unbalanced and inviable gametes as depicted in Figure 1. Therefore, recombination rates between two inverted segments, although possible, are estimated to be several orders of magnitude lower in heterokaryotypes relative to homokaryotypes. Consequently, the loci included within two inverted segments evolve mostly separately from each other, mirroring a speciation process. For example, even though the human and chimpanzee genomes are about 98% identical, more than 1.500 microinversions have been fixed between these two species contributing to their reproductive isolation (Feuk et al. 2005). Still, the reduced recombination rates can represent an important source of genetic variation within inversions over long timescales (Kirkpatrick 2010). Moreover, the importance of inversions has widely been recognized in the evolution of sex chromosomes and local

adaptation. Likewise, the evolution of mammal Y chromosomes seems to have been largely driven by inversions that inhibit recombination with its X counterpart. It is thought that this recombination inhibition led eventually to a loss of genetic material and the accumulation of male-specific traits in Y (Graves 2006).

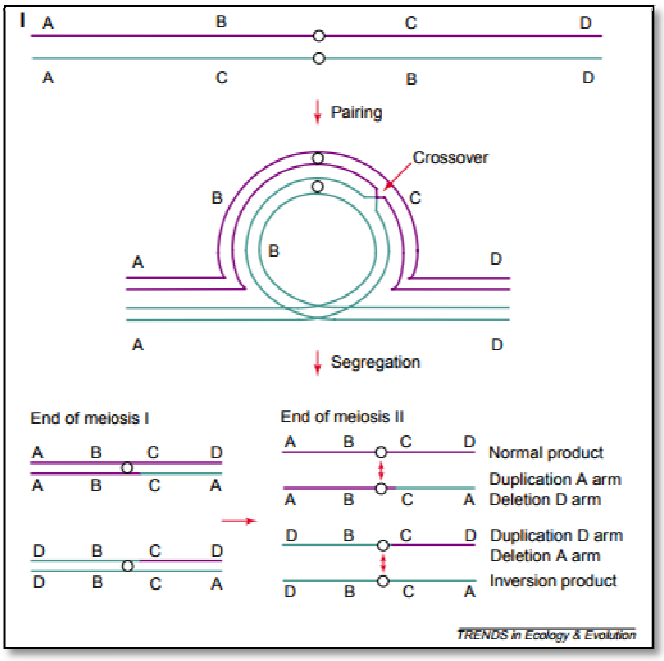


Figure 1: Recombination inhibition in heterokaryotypes: gametes resulting from a single crossover between an original segment and a segment with a pericentric inversion results in about 50% of unbalanced gametes and therefore inviable (adapted from Rieseberg (2001)).

1.1.2. Mechanisms of chromosomal inversions

Inversions are formed when two random and simultaneous breakdowns occur in a chromosome and the resulting segment is inserted, by mistake, in an inverted orientation during the reparation process. Although chromosomes can break spontaneously, certain chemical and physical mutagenic agents can increase the frequency of these events. Two main mechanisms were proposed to explain inversions formation: ectopic recombination (ER) and staggered double-strand breaks (SB). ER involves recombination between two inverted homologous sequences as transposable elements (TEs), tRNA genes or segmental duplications (Kellis et al. 2003; Cáceres et al. 2007; Delprat et al. 2009), in the same chromatid. In the staggered-break mechanism (SB), the inversion is produced when two staggered breaks are formed at both sides of a DNA segment and the repair mechanism does not rejoin a 5' end to its own 3' but to the

3' of the other breakpoint. The reparation process is completed after the remaining gaps are filled up, resulting in two inverted duplicate segments at both ends of the inversion. The presence of these inverted duplicate segments are a differential feature of SB mechanism (Casals and Navarro 2007) and can potentially confer positive fitness effects to the bearers as described in *Drosophila* (Puerma et al. 2016). In *Drosophila*, it is thought that one or the other mechanism predominates in different lineages. For example, SB is the preferred mechanism in the *Sophophora* subgenus, while most inversions in the *repleta* group seem to have originated via ER (Puerma et al. 2016). Inversions are generally assumed to be monophyletic, originating from a single chromosome in one individual (Powell 1997), making them good candidates to study chromosome evolution and phylogenetic relationships between closely related taxa. Nevertheless, some evidence suggests that certain breakpoints could have broken up multiple times (Puerma et al. 2014). This observation indicates that polyphyletic origins for the same inversion, albeit quite unlikely, remain a possibility. Interestingly, it seems that certain regions in the chromosome, called “hot” points of breakage, are more prone to break up than others. These “hot” points appear to be enriched with repetitive sequences and TEs in some species, which indicates that the latter might play a role in their formation (Lyttle and Haymer 1992; Mathiopoulos et al. 1999; Casals and Navarro 2007).

1.1.3. The role of TEs in inversion generation

TEs are relatively short sequences that can move in the genome and are found in almost all organisms. The different TE families are conventionally classified into class-I or retrotransposons and class-II or DNA transposons. The principal difference between them is their mode of replication; retrotransposons produce a RNA intermediate, while DNA transposons simply cut and paste their DNA segment during transposition (reviewed in Wicker et al. 2007). Initially considered “junk” DNA, TEs are now recognized as an important source of mutations that might contribute to species evolution (García Guerreiro 2012). TE insertions in a vicinity of a gene can prevent or modulate its expression, which might alter some cell functions. To minimize the impact of harmful mutations, the genomes of higher eukaryotes are equipped with mechanisms that suppress TE mobilization (Slotkin and Martiessen 2007). Interestingly, this mobilization can be triggered by environmental changes and genomic stress resulting from interspecific hybridization (García Guerreiro 2012; Romero-Soriano and García

Guerreiro 2016). Since most TEs seem to be transmitted vertically (from parents to their offspring) they can be useful to study phylogenetic relationships between closely related phyla (Clark and Kidwell 1997). Interestingly, some phylogenetic studies strongly indicate that TEs can also be transmitted horizontally (Clark and Kidwell 1997; Fortune et al. 2008). Some authors have proposed that bursts of TE mobilization may drive genome restructuring that in rare cases might become favorable for the host (Slotkin and Martienssen 2007; Belyayev 2014). Nevertheless, provided that TEs are generally deleterious for the host, the fate of most TE families is to degenerate to a point where they are not able to transpose anymore. The evidence of TE involvement in inversion generation and other gross rearrangements comes from phyla as diverse as bacteria (Daveran-Mingot et al. 1998), yeasts (Roeder 1983) and mammals (Schwartz et al. 1998). However, examples of it are rather scarce in *Drosophila*. One of the best described examples is the DNA transposon *Galileo*, which is thought to have generated several inversions in *D. buzzatii* via ectopic recombination between inverted copies (Cáceres et al. 2001; Delprat et al. 2009). Interestingly, this element seems to be relatively abundant in species with high inversion polymorphism (Gonçalves et al. 2014). In the case of *D. subobscura*, Felger and Sperlich (1989) found that, although most of TE probes hybridized in the sexual A chromosome that shows the highest degree of reorganization between closely related species, very few hybridization signals coincided with inversion breakpoints. These results do not back up the hypothesis of a direct engagement of TE sequences in inversion generation in these species Moreover, in *D. subobscura* and the closely related, *D. madeirensis* and *D. guanche*, *P* element sequences were localized in one cluster in the O chromosome. Since no transposase-encoding sequence could be recognized, *P* element sequences are likely inactive in these three species. Interestingly, the sequences necessary for its repressor activity are conserved, suggesting that these could have acquired a new function (Krimbas 1993). In contrast, insertions of the *P* element DNA transposon tended to localize in inversion breakpoints in *D. willistoni*, suggesting a relationship with inversion generation in this species (Regner et al. 1996).

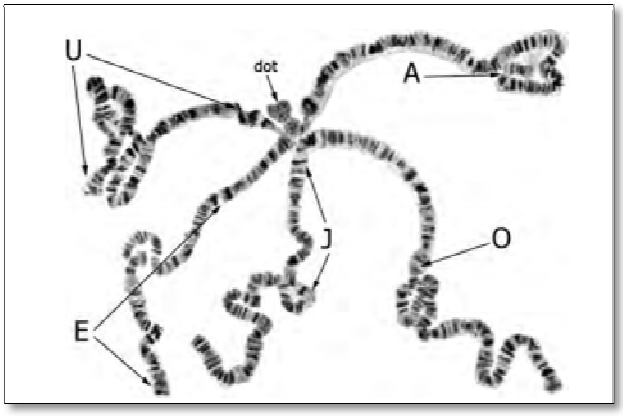


Figure 2: Polytene chromosomes that show the standard karyotype of *D. subobscura* with its 5 acrocentric chromosome pairs and a dot chromosome pair. Arrows indicate chromatin peculiarities used to identify each chromosomal pair (adapted from Serra (2013)).

1.1.4. Visualization

Inversions are easier to study and characterize in the polytene chromosomes that are common to many dipteran insects (Figure 2). These giant chromosomes occur in certain specialized cells of salivary glands and other tissues with high metabolic rates. They are observed in the late third instar larval stage because they form when homologous chromosomes pair together in the synpasis and undergo multiple rounds of DNA replication without cell division, a process called endomitosis. In addition to increase cell volume, polytene chromosomes are thought to participate in metabolism, as multiple copies of genes allow a higher level of gene expression. In their visualization, we can distinguish the chromomeres that are dark bands where chromatin is most compact. They are depleted of genes relative to the lighter bands or interchromomeres, where chromatin is more relaxed (Figure 3A above). The pattern of chromomeres has been used to identify inversions and track chromosomal homologies between closely related species. They are also useful to localize different kinds of loci, including genes, microsatellites and inversion breakpoints. Bubble-like structures called chromosomal puffs are sometimes observed and consist in regions where chromatin is locally decondensed to allow intensive gene transcription (Figure 3A below). In most *Drosophila*, puffs may also show up in response to a change in the environment just as after heat stress (Moltó et al. 1992). Interestingly, heat shock puffing patterns exhibit similarities within the same inversion in different strains, but also some differences depending on the strain and its geographical origins (Krimbas 1993). In the

heterokaryotypes, several chromosomal loops can be seen, which result from the incomplete synapsis between the two inverted segments (Figure 3B). Since the position, shape and size of a loop is unique to the bearer's karyotype, they can help to identify the inversions carried by it.

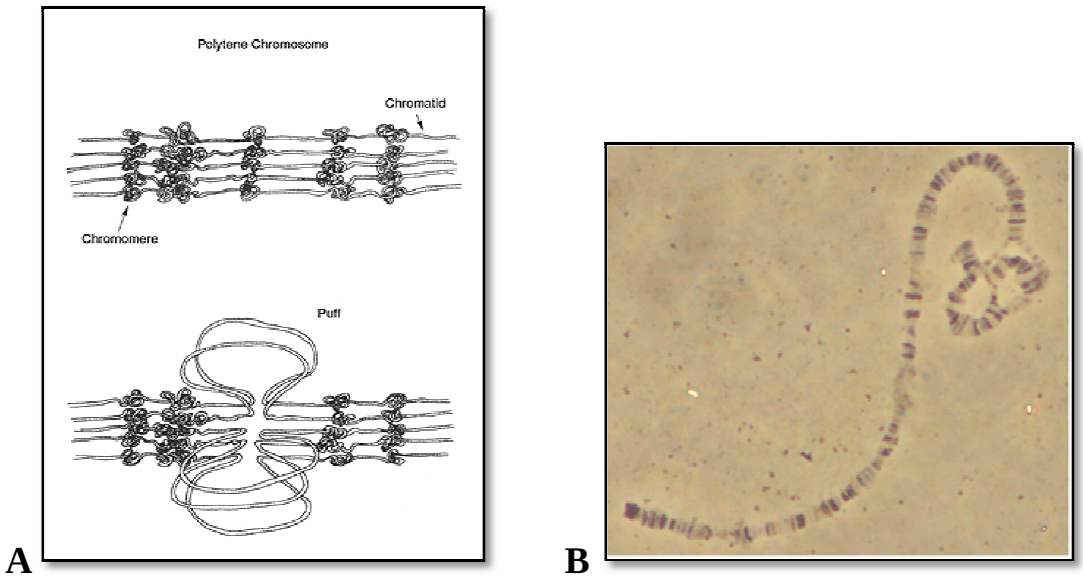


Figure 3: A: Schema representing chromomeres (above) and a puff (below) in a chromosomal region (adapted from: http://www.cmb.ki.se/research/daneholt/br.html). **B:** An O_{ST}/O_{3+4} homologous pair showing its distinctive loop (adapted from Pegueroles et al. 2013).

1.1.5. Maintenance of inversion polymorphism in natural populations

Chromosomal inversions are often described as adaptive cassettes that can become advantageous in a population by capturing favorable allele combinations that might contribute to evolutionary processes such as local adaptation and speciation by several mechanisms (Hoffmann and Rieseberg 2008). Any advantageous allele combination or haplotype captured by an inverted segment will remain stable for longer due to suppression of recombination (Dobzhansky 1970; Kirkpatrick and Barton 2006). Although inhibited recombination rates are general between alternate inversions, these can differ substantially based on the characteristics of the inversion. It is generally accepted that gene flow between two inverted segments can produce viable gametes by two molecular mechanisms: double crossover and gene conversion (GC). Due to structural constraints, double crossovers can only form in the more central regions of relatively long inversions, while GC is thought to be the predominant mechanism near inversion breakpoints. Likewise, the occurrence of double crossovers is reduced in

small inversions and complex arrangements. Another worth noting difference between these two mechanisms is that while recombinant tracts resulting from GC typically involve only a few ten to hundreds of base pairs, those originating from double crossovers are generally longer, sometimes comprising more than 3 kb (Rozas et al. 1999; Chen et al. 2007). Due to structural constraints, gene flow rates among inverted segments are expected to be negligible near the breakpoints. Consequently, favorable epistatic interactions among blocks of genes that have been captured by an inversion might be preserved for longer in these regions (Rozas et al. 1999).

The adaptive value of inversions has been corroborated in taxa as diverse as plants and humans (Hoffmann and Rieseberg 2008). In *Drosophila*, a latitudinal component has been described for certain inversions that in turn show an association with phenotypic traits including body and wing size, wing-loading, flight activity and plant toxin resistance (Stalker 1980; Iriarte et al. 2003; Kennington et al. 2007; Pegueroles et al. 2016). The adaptive value of inversions has also been shown in the genus *Anopheles*, where some inversions have been linked to insecticide resistance, behavioral characteristics and enzymatic resistance (Ayala et al. 2014). Highly divergent regions between inverted segments are a common target to identify candidate genes responsible of adaptive inversion polymorphism (White et al. 2007; Cheng et al. 2012). Interestingly, certain gene pathways associated with similar selective environmental pressures are shared between species of different genera such as *A. gambie* and *D. melanogaster*, suggesting that different insect phyla may often reuse the same set of genes to adapt to analogous environmental niches (Kirkpatrick and Kern 2012).

Even though these examples evidence the adaptive value of inversions, there is no consensus on how an inversion polymorphism is preserved in a population once it originates. Researchers have come up with at least six hypotheses to explain the maintenance of inversion polymorphism in natural populations. Interestingly, some models may be more prevalent in certain phyla than others due to biological and demographic factors. These models are briefly explained as follows, emphasizing the studies in our model species *D. subobscura* and other closely related *Drosophila* (see Hoffmann and Rieseberg 2008 for a more detailed review).

I. The **coadaptation** hypothesis originally proposed by Dobzhansky (1950) and later upgraded by Haldane (1957), states that inversions preserve sets of allele combinations

that are adaptive in a local environment. These favorable allele combinations interact epistatically between themselves in the same inversion and also with alleles in the alternate inversion of a homologous chromosome, which results in a higher fitness in the heterokaryotype or heterosis (Figure 4A). The introduction of alleles from different populations will annulate these favorable interactions, lowering the fitness of the individual carrying them.

For instance, the cage experiments of Dobzhansky and Epling (1948) demonstrated that in some *D. pseudoobscura* populations, certain arrangements are coadapted with alternative arrangements of the same population, and that recombination between the same inversion from different populations results in the elimination of favorable epistatic interactions in the heterokaryotypes. In contrast, no evidence in favor of the coadaptation model has yet been found in *D. subobscura* (Krimbas 1993). For instance, heterokaryotypic F1 individuals did not show heterosis and there is no evidence on breakdown of coadapted gene complexes in F2 (McFarquhar and Roberston 1963; Pentzos-Daponte and Sperlich 1965). This last finding is disputed by Prevosti (1967) that found a possible breakdown of coadapted complexes in F2 relative to wing length, although he also failed to find heterosis in F1 hybrids. Interestingly, Pegueroles et al. (2010a) noted that *D. subobscura* heterokaryotypes were more frequent than expected regardless of the origins of the populations used (Greece, Sweden and Spain). These results are not compatible with the coadaptation hypothesis, where only heterokaryotypes of the same population are expected to have increased fitness. Also, according to this hypothesis, the genetic content of a particular inversion is predicted to differ between populations. However, little or no genetic differentiation has been noted for the same arrangement in separate European *D. subobscura* populations (Pinsker and Böhm 1989; Rozas et al. 1995; Pegueroles et al. 2013).

II. The concept of gene complexes or **supergenes** detailed in Wasserman (1968), suggests that different sets of coadapted combinations of alleles or 'supergenes' occur in the same inversion. In contrast with the coadaptation hypothesis, briefly detailed above, epistatic interactions will only take place between alleles localized in the same inverted segment (Figure 4B). According to this hypothesis, structural homogeneity does not equal to genetic homogeneity, and the same inversion is expected to differ genetically between populations. One consequence of genetic heterogeneity in the same inversion is that recombination can easily disrupt favorable epistatic interactions. Thus,

homokaryotypes are expected to exhibit a lowered fitness relative to heterokaryotypes because they display higher recombination rates.

The supergene selection in *D. subobscura* is supported by some linkage disequilibrium (LD) studies of seasonal variations of allozymes that reported evidence of epistatic interactions between loci within the O_{ST} inversion (Fontdevila et al. 1983; Santos 2009). However, Santos (2009) indicates that this hypothesis is not compatible with O_{ST} latitudinal clines and the high migration rates of this species. To reconcile these observations, Santos (2009) proposes that migration could be contributing more to the observed LD among loci in this inversion than epistatic interactions between them. His explanation for the patterns observed in Fontdevila et al. (1983) is that there is a certain degree of population structure in the O_{ST} arrangement, and that periodical fluctuations in gene exchange between populations would contribute more to the observed cyclical changes in LD in O_{ST} than epistatic interactions.

III. The **local adaptation** scenario formulated by Kirkpatrick and Barton (2006) proposes that inversions are favored when they capture at least two alleles that are individually adaptive in their local environment. A worth noting difference, regarding the coadaptation hypothesis, is that epistatic interactions between alleles may or may not take place in the local adaptation model. Once they have captured an advantageous haplotype, inversions are maintained in the population by selection-migration equilibrium. No epistatic interactions are required for this to happen, although this scenario does not negate them and these could even assist chromosomal inversion evolution (Figure 4C). This hypothesis is thus less restrictive than the coadaptation model.

If there are no epistatic interactions between loci within the same inversion, the two inverted segments will eventually become homogenized by double crossover and GC except at inversion breakpoints (Navarro et al. 2000). Due to physical constrainments, gene flow rates among inverted regions are expected to be lowest around breakpoints and highest in the central area of the inverted segments. Consequently, the genetic content around inversion breakpoints is expected to be rather similar in the same inversion from different populations and noticeably different between arrangements in the same population.

The allele frequencies of several loci located within the O_{3+4} inverted region in *D. subobscura* indicate little between-population differentiation for the O_{ST} and O_{3+4} inversions, which stands in contrast with their clear latitudinal clines (Rozas and Aguadé 1990; Rozas et al. 1995; Pegueroles et al. 2013). At the same time, strong genetic differentiation exists between alternate arrangements in the same population of this species (Rozas et al. 1999; Munté et al. 2005; Pegueroles et al. 2013). These outcomes are compatible with the local adaptation hypothesis, but not with the coadaptation hypothesis. On the whole, the local adaptation hypothesis is to date the one that better explains the maintenance of inversion polymorphism in *D. subobscura* populations.

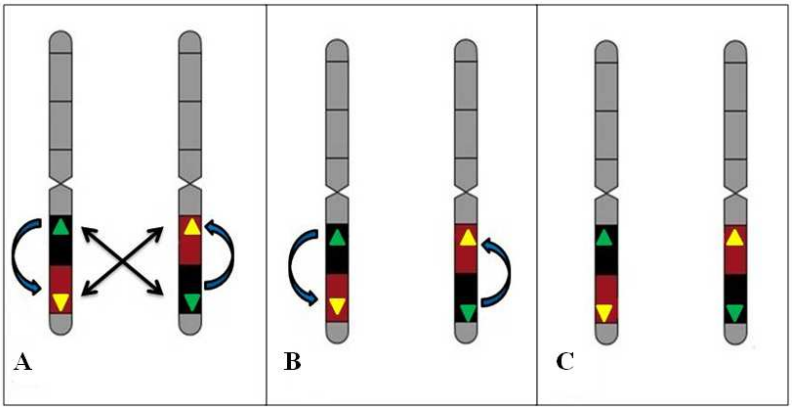


Figure 4: Epistatic interactions under three models of inversion maintenance in populations. A: the coadaptation hypothesis, where epistatic interactions (represented as arrows) take place among different loci (represented as green and yellow triangles) within and between different chromosomal arrangements. **B:** Supergene selection, where epistatic interactions are expected to occur only among loci within the same inversion. **C:** the local adaptation hypothesis, where epistatic interactions between loci might or might not occur.

IV. The hypothesis of **positive selection** states that a new inversion can become adaptive if its formation is accompanied by mutations that alter the expression patterns of its neighboring loci. Once an inversion is produced, the new gene expression patterns will become the target of natural selection. For instance, there is solid evidence for inversion polymorphism being adaptive in many *Drosophila*, including *D. subobscura*. Evidence suggests that most inversions in the latter and in most *Sophophora* species seem to have originated by the SB mechanism that produces duplicated fragments that might confer positive effects on their bearers (Puerma et al. 2016).

V. Adaptive inversions: under this scenario, selection will act on an entire inversion that has captured multiple adaptive alleles. Inversions can become **underdominant** or **overdominant** relative to the original chromosome.

> **-Underdominance** occurs when the heterokaryotypes show decreased fitness relative to the homokaryotypes. This is mainly caused by the high proportion of unbalanced gametes formed by single crossovers during meiosis. Interestingly, single crossovers seem to occur at a higher frequency in plant species (Rieseberg 2001). Demographic changes and high autofecundity rates often result in a decrease of the efective population effective size, which may lead to the fixation or loss of the inversion.
>
> **-Overdominance** is observed when the fitness of the heterokaryotype is higher than that of the homokaryotypes. Nevertheless, the reason for this is different from those postulated by the first three models. Under the overdominance model, deleterious recessive alleles that appear in one inversion, will come together in the homokaryotype much more often than in the heterokaryotype. Consequently, the former would exhibit a lowered fitness relative to the latter.

VI. Finally, some inversions could be considered selectively **neutral**. When this is the case, their frequency in the population will depend on the populational specific parameters such as its effective size, migration rates and genetic drift.

1.2. *Drosophila subobscura* as a model species to study adaptive inversion polymorphism

1.2.1. Introduction to the species

Drosophila subobscura (Collin 1936) is one of the most common *Drosophila* species in Europe and its adjacent areas in the Palearctic region. Since it is relatively abundant in the field and easy to cultivate in the laboratory, it has been extensively studied by many research groups in Europe from different specialties, including population genetics, ecology and insect physiology. Since its discovery in the Americas, a few decades ago,

D. subobscura has attracted the interest of many scientists from this continent as well. This species exhibits a rich chromosomal inversion polymorphism that has received the attention of many population geneticists and evolutionists in the last decades, with some inversions showing an association with temperature and latitude. To try to decipher the molecular mechanisms behind the adaptive character of *D. subobscura* inversion polymorphism, in this work we have studied the genomic location, the nucleotide polymorphism and the gene expression patterns of the thermally inducible *hsp70* gene family in four representative arrangements of the O chromosome, including the warm climate-associated $O_{\underline{3+4}}$ and the cold climate-associated O_{ST}.

1.2.2. Classification

D. subobscura belongs to the *obscura* species group within the *Drosophila* genus. *Obscura* is closely related to the *melanogaster*, *willistoni* and *saltans* groups, all members of the 25-40 million years ago (MYO) *Sophophora* subgenus (Russo et al. 1995; Obbard et al. 2012). The sequencing of several genes suggests that *obscura* can be divided into the Nearctic and Palearctic groups that split 10-15 MYA (Russo et al. 1995; Barrio and Ayala 1997). The Palearctic cluster includes the monophyletic *obscura* and *subobscura* subgroups and a few species of unresolved phylogeny. The monophyletic Nearctic cluster is in turn represented by lineages that colonized the New World, including the *pseudoobscura* and *affinis* subgroups (Barrio and Ayala 1997). Two of the most representative *obscura* species, with publicly available genomes are the closely related *D. pseudoobscura* and *D. persimilis*. The *subobscura* subgroup is represented by three species, including *D. subobscura*, *D. madeirensis* and *D. guanche*. DNA and inversion polymorphisms have consistently shown that *D. subobscura* and *D. madeirensis* are sister taxa that diverged only 0.6-1.0 MYA, while the divergence between *D. guanche* from *D. subobscura*/*D. madeirensis* occurred 1.8-2.8 MYA (Ramos-Onsins et al. 1998). *D. subobscura* coexists sympatrically with *D. madeirensis* in the Madeira island and with *D. guanche* in the Canary islands. An analysis of the *rp49* gene region suggests that *D. subobscura* populations in Madeira and Tenerife likely represent two independent colonization events from the continent (Khadem et al. 1998).

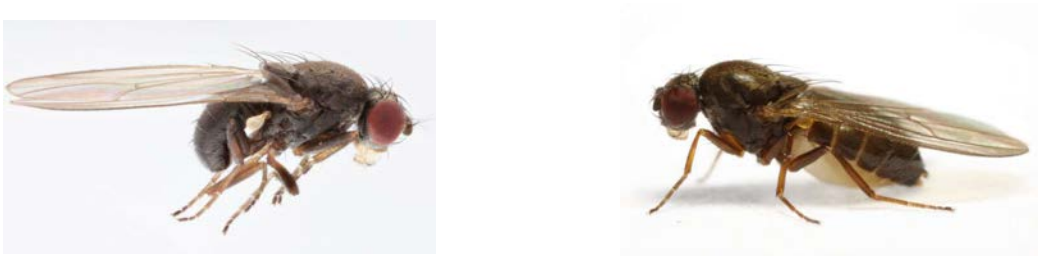


Figure 5: Unscaled images of adult *D. subobscura* individuals: male (left) and female (right).

1.2.3. Biology

Much like other drosophilids and some insect phyla, *D. subobscura* life cycle undergoes complete metamorphosis or holometabolism. In holometabolic species, the life cycle is divided into four stages, namely egg, larva, pupae and adult. The complete life cycle of *D. subobscura* lasts about 23 days at 18ºC in optimal larvae density (Orengo and Prevosti 1994), but it is highly sensitive to environmental factors. Higher and lower temperatures result in shorter and longer life cycles, respectively. Adult lifespan is also affected by temperature, being significantly longer (≥50 days) at 10ºC-12ºC than at 20ºC-22ºC (±20 days) (Krimbas 1993). Some adults raised at benign conditions can exceptionally survive for more than 100 days. The optimum developmental temperature for this species has been established at 18ºC, although adults can tolerate a wide range of temperatures and exhibit fly activity from 3ºC to 37ºC (Krimbas 1993). Temperatures above 25ºC are generally considered stressful for this species, causing damage to the ovaries and reduced fertility (Krimbas 1993). In addition to temperature, humidity is another potential stressor. For example, *D. subobscura* flies are more active at sunset and sunrise during summer, while at noon they prefer to rest at cooler microenvironments to avoid low humidity conditions (Dyson-Hudson 1956).

Mating behavior has also been studied in *D. subobscura*. In standard conditions, females are expected to reach sexual maturity about 24 hours after eclosion. Similar to the other two species of the *subobscura* cluster, *D. madeirensis* and *D. guanche*, but in contrast with most *Drosophila*, *D. subobscura* flies mate only in presence of light. Although it is often assumed that *D. subobscura* females mate only once in their lifetime (Smith 1956), a small but significant percentage of them have shown to mate at least twice (Loukas et al. 1981). Interestingly, the sperm of *D. subobscura* males shows more than one morphology, a phenomenon called polymegaly that is typical of species with polyandrous females.

1.2.4. Ecology and geographical distribution

D. subobscura is viewed as a generalist species based on its diet. Larvae feed themselves typically of decaying fruits, vegetation, fungi and fermenting sap, being fruits the preferred medium. Females oviposit in fruits, preferably bananas. Indeed, fermented bananas are commonly used as baits to collect adult flies in the wild. Although its ecology is poorly known, *D. subobscura* has been collected in forests of different tree species, including *Quercus*, *Ulmus*, *Acer*, *Castanea*, *Abies* and *Pinus* (Calabria 2012). It can also be found in urban areas, associated with human activity (Kenig et al. 2010; Calabria 2012).

This species is native to the Palearctic ecozone that consists of temperate and boreal ecoregions. Its native habitat includes most of Europe, North Africa and some parts of Asia Minor (Figure 6). Towards the Atlantic, it has been detected in some islands, including Madeira, Azores and the Canary islands. It has also been collected in many Mediterranean islands, including the Balearic, Sardinia, Sicily, Crete, Corsica, Cyprus and other smaller ones.

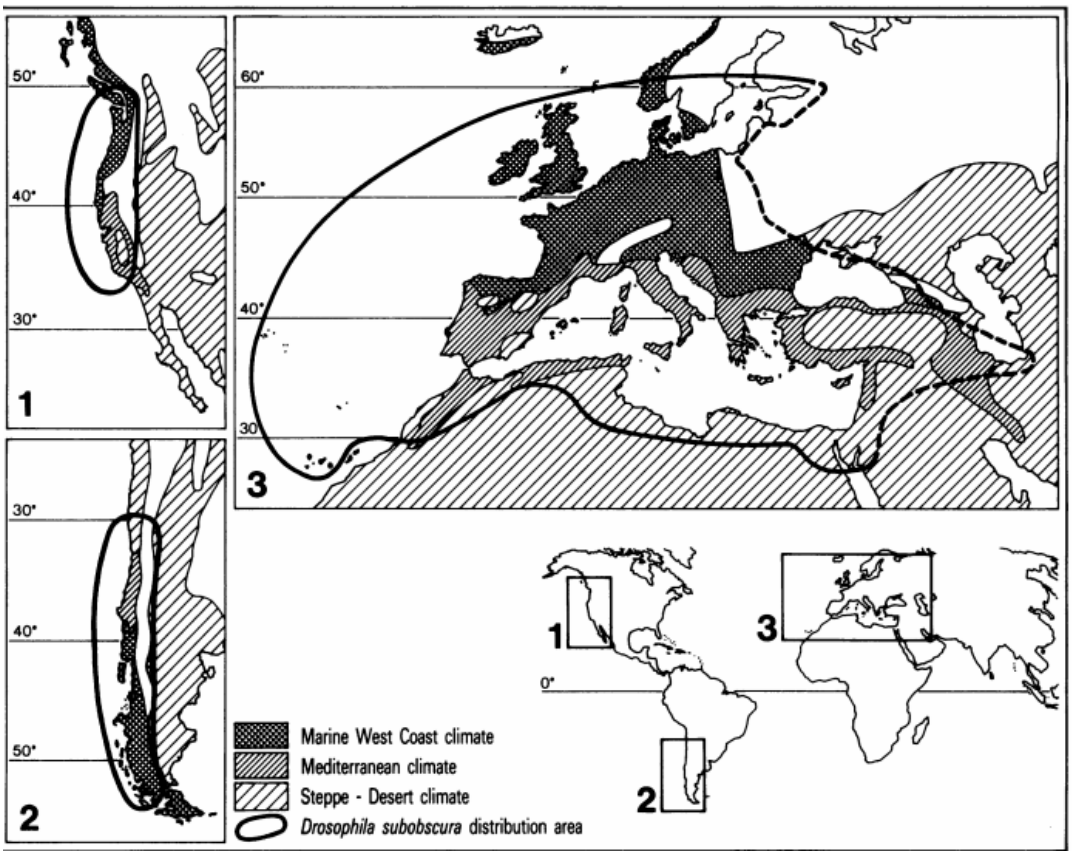


Figure 6: Climatic characteristics of *D. subobscura* habitat range (adapted from Prevosti et al. (1988)).

Human activity has been responsible for its rapid spread to various locations, including South America, North America and possibly Egypt. This successful colonization indicates that, similar to its relative *D. melanogaster*, *D. subobscura* has a great adaptation capacity. According to some population models, the colonizer American populations are descended from 10 to 150 individuals that were brought to Chile in late 1970s and successfully spread to the surrounding areas. The origin of this founding population is disputed, being the western Mediterranean region a plausible source based on inversion frequencies and microsatellite loci (Prevosti et al. 1985; Ayala et al. 1989; Pascual et al. 2007). In 1982, *D. subobscura* was detected for the first time in Townsed, Washington (USA). Studies on inversion polymorphism and lethal allelism indicate that, in this case, the founding population had its origins in South America (Ayala et al. 1989, Mestres et al. 1992). Interestingly, it was noted that already in 1988 the American populations showed latitudinal clines in inversion frequencies that mirrored those in the Old World (Prevosti et al. 1988; Ayala et al. 1989; Prevosti et al. 1990). Not without reason, the colonization of the Americas was described as "*a grand experiment in nature*" by these authors. Newer studies have confirmed that inversion clines did indeed form in the American populations (Balanyà et al. 2003, 2006), but a more recent survey from this continent indicates that they are vanishing (Castañeda et al. 2013). Behavioural thermoregulation, where individuals use physiological adjustments and behavior to maintain their body temperature, has been proposed as a mechanism to explain the fading of these adaptive inversion clines (Castañeda et al. 2013).

D. subobscura coexists with other drosophilids across its habitat. In Northern and Central Europe it coincides with its close relative *D. obscura*, while the latter becomes rarer towards the South, where *D. subobscura* is the most common *Drosophila* species. In the Americas, it coexists with the closely related *D. pseudoobscura* in the West Coast of North America and with various *Drosophila* in South America that choose the same fruits to oviposit (Alamiri 2000).

Population size is affected by environmental conditions, such as temperature, humidity and food availability. Studies on population dynamics in the Palearctic region indicate that *D. subobscura* population size fluctuates along the year. Peak population sizes are reached in summer in Northern Europe and in spring and autumn in Southern Europe (Calabria 2012). One consequence of such fluctuations is that population effective sizes are difficult to calculate. Some estimates have been made, in most cases, when

populations were in their peak. The figures seem to be higher in populations sampled in the Balkan mountains and decrease gradually as we move away from there. As such, the European populations from the Balkans would be ecologically central in comparison with populations from more peripheral areas, including those from Northern Europe, the Middle East and Africa. Ecologically, central regions include multiple and numerous habitats where the species can thrive, while the more marginal regions exhibit fewer and rarer habitats. In accordance with this, population effective sizes and chromosomal diversity are expected to be higher in more central areas. The calculated frequencies of O chromosomes carrying one or multiple lethal alleles in various populations, indicate that the lethal charge is significantly lower in the more marginal populations of *D. subobscura*. This is explained because in populations with low effective sizes there is a faster elimination of lethal recessive alleles due to higher rates of inbreeding. Nonetheless, the low chromosomal diversity expected for marginal *D. subobscura* populations is not meet in certain urban populations from peripheral areas. Some authors have hypothesized that urbanization results in the creation of many new ecological niches, which might be accompanied by an increase in chromosomal diversity (Kenig et al. 2010).

The rapid spread to various areas in the New World suggest that *D. subobscura* has high dispersal capabilities (Prevosti et al. 1988). This is in line with low values for the fixation index (F_{ST}) reported between continental European populations using microsatellite data (Pascual et al. 2001) and the lack of population subdivision, for the same inversion, using restriction site polymorphisms (Rozas and Aguadé 1990; Rozas et al. 1995). On the contrary, the Canarian populations exhibit notable genetic differentiation from continental and Madeira ones (Khadem et al. 1998). As expected, strong genetic differentiation has been detected between alternate inversions in the O chromosomes from several European populations, which results from the inhibited recombination in heterokaryotypes. The high genetic differentiation between alternate inversions contrasts with an apparent lack of structure for the same arrangement between different populations, which indicates high gene flow levels (either passive or active) among continental European populations (Pinsker and Böhm 1989; Rozas and Aguadé 1990; Rozas et al. 1995; Calabria 2012; Pegueroles et al. 2013).

1.2.5. *D. subobscura* inversion polymorphism: overview

The karyotype of *D. subobscura* consists in 6 homologous chromosome pairs. These are divided into five acrocentric chromosomes including the X chromosome (A) plus four autosomes: O, J, U and E. A small dot-like chromosome constitutes the sixth pair. The male Y chromosome is acrocentric and no euchromatic regions in it have been confirmed (Menozzi and Krimbas 1992; Krimbas 1993).

Compared with other *Drosophila*, this species exhibits a rich inversion polymorphism. More than 60 inversions and 85 arrangements have been discovered in natural populations, which are distributed across its 5 acrocentric pairs (Menozzi and Krimbas 1992). Some of these arrangements are restricted to certain areas of the species distribution. The nomenclature used here is described in Krimbas (1993) and detailed as follows: each inversion is labeled first by the letter of its chromosome followed by a subscript that reflects the time of their discovery. For example, A_1 was the first inversion described in the A chromosome. Many inversions are simple (O_7, O_{12}) while others overlap with each other forming complex arrangements ($O_{\underline{3+4+23}}$, $O_{\underline{3+4+8}}$). Combinations of nonoverlapping arrangements and inversions occuring in different chromosomal segments have been observed, where the same inversion can combine with different arrangements and vice versa ($O_{\underline{3+4}+\underline{1}}$). Overlapping inversions like $O_{\underline{3+4+8}}$ are underlined, while nonoverlapping ones are represented by a dashed line ($O_{\underline{3+4+16}+\underline{2}}$). An exception to this nomenclature are the "standard" inversions, which are designed with a ST subscript. Five "standard" inversions distributed across its five acrocentric pairs (E_{ST}, J_{ST}, O_{ST}, U_{ST}, A_{ST}), exist in *D. subobscura*. Provided that this nomenclature was established by convention, it is curious that in most cases the ST inversions appear to be more abundant at higher latitudes in both the Palearctic region and the Americas (Orengo and Prevosti 1996; Zivanovic and Mestres 2011). Also, it is worth noting that some inversion combinations are observed in natural populations more often than expected by chance. For instance, the O_7 inversion usually combines with $O_{\underline{3+4}}$ to form the $O_{\underline{3+4}}$ arrangement that occurs at a high frequency in many Mediterranean populations. This biased pattern is explained in some cases by strong suppression of recombination in heterokaryotypic females (*Drosophila* males usually do not recombine) and the formation of lethal gametes during meiosis (Krimbas 1993). Although certain inversion combinations in the same chromosome appear to be more

abundant than others, no LD is observed between inversions in different chromosomes that seem to combine at random (Zivanovic et al. 2014, 2016).

1.2.6. *D. subobscura* inversion polymorphism: the case of the O chromosome

In this thesis we will focus only in inversions and arrangements that occur in the O chromosome of *D. subobscura*. This chromosome corresponds to Müller E element and is homologous to the 3R arm in *D. melanogaster* and chromosome 2 in *D. pseudoobscura*/*D.persimilis*/*D. miranda* (Segarra et al. 1996). Of the six pairs is the longest, having a length of 190 ≈31 Mb (Pegueroles et al. 2010b). It is conventionally divided into segments I (SI) and II (SII), where the first is the most distal to the centromere. About 24 natural inversions and 46 arrangements, which include combinations of overlapping and nonoverlapping inversions, have been described in the Palearctic populations for this chromosome (Krimbas 1993). Curiously, no inversions comprising simultaneously segments I and II seem to exist.

1.2.7. Adaptive inversion polymorphism in *D. subobscura*

D. subobscura inversion polymorphism was initially classified as "rigid" by Dobzhansky (1962), as it did not seem to respond to environmental changes. This stood in contrast with the more "flexible" nature of the genome of its relative *D. pseudoobscura*, where seasonal clines for some of its inversions had been described. A few years later, Sperlich and Feuerbach (1966) described *D. subobscura* inversion polymorphism as semirigid/semiflexible, based on the observation that, although most inversions are quite stable, some of them do respond to environmental fluctuations. Nowadays, there is ample evidence on short, mid and long-term changes in *D. subobscura* inversion polymorphism that correlate with certain environmental variables. Consequently, it has become a valuable piece of information to monitorize climate changes over time.

1.2.7.1. Geographical changes

The inversion polymorphism in *Drosophila* is subject to respond to a variety of environmental variables. One of the most studied and well-known examples include the

altitudinal and/or latitudinal components exhibited by certain inversions of various *Drosophila* species (Powell 1997). What forces are behind these geographical patterns is a hot discussion topic among evolutionary geneticists, and diverse clarifications have been proposed. These include interpretations of selection based on environmental heterogeneity, behavioral temperature regulation, genetic drift and historical considerations.

1.2.7.1.1. Latitude

Latitudinal trends have been observed in some cosmopolitan inversions of *D. melanogaster* and other *Drosophila* species For example, in *D. melanogaster* the frequency of the In(3R)Payne inversion exhibits a north-south cline along Australia's East Coast. Interestingly, this cline has shifted its position in the last 20 years, a trend probably influenced by global climate change (Anderson et al. 2005). Other species that exhibit latitudinal clines for some of their inversions are *D. buzzatii* (Hasson et al. 1995), *D. robusta* (Levitan 1992) and *D. persimilis* (Powell 1992).

Studies in several *D. subobscura* populations from the Palearctic region indicate that certain inversions display a marked latitudinal component, which reinforces their adaptive nature (Menozzi and Krimbas 1992; Krimbas 1993). In addition, the frequencies of some clinal inversions in the colonizer American populations have evolved closer to their Old World pattern in less than two decades after the initial colonization event, a rapid process that cannot be explained by genetic drift alone (Prevosti et al. 1988; Balanyà et al. 2003; Rego et al. 2010).

1.2.7.1.2. Altitude

An altitudinal component, for certain inversions, has been recorded in some *Drosophila*, including *D. mediopunctata* (Ananina et al. 2004), *D. buzzatii* (Hasson et al. 1995), *D. robusta* (Etges 1984) *D. pseudoobscura* (Dobzhansky 1948), *D. annanassae* (Reddy and Krishnamurthy 1974) and *D. persimilis* (Dobzhansky 1948; Mohn and Spiess 1963) among others. The evidence for an altitudinal component in some *D. subobscura* inversions is rather limited. For instance, Martínez-Sebastián et al. (1984) failed to find an association with altitude after studying the inversion frequencies of three populations located at different altitudes in Sierra de Gúdar (Aragón, Spain) considering their different ecological and climatic conditions. Similarly, Prevosti (1971, 1972) did not

detect any changes associated to altitude after revisiting the inversion frequencies of populations from the Canary islands and Madeira. In contrast, Burla et al. (1986) found differences in the J chromosome inversion polymorphism between samples captured at different latitudes in the Swiss Alps, with the cold-associated J_{ST} arrangement increasing its frequency at higher altitudes.

1.2.7.2. Temporal changes

Besides the conspicuous latitudinal and altitudinal clines, strong frequency shifts associated with seasonality and climate change have been reported for certain inversion polymorphisms. In the next lines I will provide some evidences on these aspects.

1.2.7.2.1. Seasonal (short-term) changes

Short-term changes in inversion polymorphism are usually defined by a significant variation in the frequency of a certain inversion in a population in a period of less than a year. Seasonal trends in inversion frequencies have been observed in various *Drosophila*, including *D. pseudoobscura* (Dobzhansky 1948), *D. persimilis* (Dobzhansky 1948; Dobzhansky and Ayala 1973), *D. robusta* (Levitan 1992) and *D. funebris* (Sperlich and Pfriem 1986).

In *D. subobscura*, several studies have pointed out short-term changes in frequencies of some arrangements that were associated with season changes. In Burla and Götz (1965) a seasonal component was detected for some inversions in flies sampled three times along the years 1963 and 1964 in Zurich. Similarly, Fontdevila et al. (1983) identified two arrangements in the O chromosome that exhibited contrasting seasonal shifts in flies sampled periodically along 5 years in Mount Pedroso (Galicia, N.W. Spain). More precisely, the warm climate-associated $O_{\underline{3+4}}$ arrangement increased its frequency during the late summer period, while the cold climate-associated O_{ST} showed the opposite trend, decreasing in summer and increasing its frequency towards autumn. A revision over a 15 year time span, in flies collected periodically in Mount Pedroso, confirmed these seasonal trends (Rodríguez-Trelles et al. 1996).

Rodríguez-Trelles et al. (2013) developed a “genome-wide warm dose” index after carefully examining the seasonal shifts in inversion frequencies in two populations in the North of the Iberian peninsula. This index was calculated based upon the

frequencies of the inversions that increased during summer 2011 and 2012 in the Mount Pedroso population and in a second one located ~600 km eastwards (Berbikiz, Basque Country, N. Spain). Interestingly, the April month was exceptionally warm in 2011 in western Europe and, suprisingly, flies collected over that month at both places showed an increased "warm dose" that reminded the typical values of the late summer period. These outcomes reinforced the conjecture that temperature may drive adaptive evolutionary shifts and that the genome is able to accomodate itself to rapid climatic changes. In contrast with these findings, some early research lines have failed to record cyclic seasonal changes in *D. subobscura* inversion polymorphism (Pentzos-Daponte 1964; Prevosti 1964; De Frutos and Prevosti 1984). However, the evidence against seasonality presented in these ancient works should be taken with caution, as it comes from a small number of populations and only short seasonal periods were covered (reviewed in Rodríguez-Trelles et al. 1996).

1.2.7.2.2. Long-term changes

Long-term changes in inversion polymorphism can be defined as a significant variation in the frequency of a certain inversion in a population occurring over a long period of time (usually, one or several decades). Some examples include the PP and TL arrangements in *D. pseudoobscura*, two inversions whose frequencies have significantly increased during the last decades in the west coast of North America (including British Columbia and California). Curiously, their increase does not show any evident association with environmental changes (Anderson et al. 1991). This suggests that innovations in the genetic content of these inversions have been *per se* the target of selection.

A gradual loss of chromosomal diversity accompanied by an expansion of low latitude genotypes poleward has been noted in long-term studies in several populations. On this line, the warm climate-associated $O_{\underline{3+4}}$ arrangement in *D. subobscura* seems to have increased its frequency over time, while the cold climate-associated O_{ST} has become increasingly rare. This has resulted in a more "southern" chromosomal composition in the most recent collections (Rodríguez-Trelles et al. 1996; Rodríguez-Trelles and Rodríguez 1998; Solé et al. 2002; Orengo et al. 2016). This trend is often interpreted as evidence of rapid adaptation to global warming (Orengo and Prevosti 1996; Balanyà et al. 2006; Rezende et al. 2010). In agreement with this hypothesis, the newest collections

of *D. subobscura* in Europe indicate that this species is expanding northwards, likely because these regions are becoming warmer (Krimbas 1993).

Although most studies suggest that temperature is the main factor responsible inversion frequency shifts in *D. subobscura*, this is likely far from being so simple (Calabria 2012). In fact, certain temperature-associated arrangements show some differences among them. For example, the warm climate-associated O_{3+4} has a strong seasonal component that is not so apparent in the warm climate-associated O_{3+4} (Rodríguez-Trelles et al. 1996; Rodríguez-Trelles et al. 2013). Likewise, O_{3+4} is relatively more freqüent than O_{3+4} in certain Southern European populations, and shows a more pronounced latitudinal component in the continental Europe (Calabria 2012) and the newly colonized areas (Balanyà et al. 2003). In line with these findings, *D. subobscura* populations from Chile maintained in the laboratory for several generations at three thermal regimes (13, 18 and 22ºC) showed that, although some of the inversions mirrored the patterns found in nature, temperature alone was unlikely to explain all the variation (Santos et al. 2004, 2005). However, a pitfall of these laboratory studies is that populations are not in a constant temperature regime in the wild, but rather exposed to significant temperature fluctuations. Since these fluctuations are likely more extreme in certain areas of its habitat, it makes sense that some inversions are better at buffering their effects than others. An indirect relationship between temperature and inversion polymorphism is supported by certain warm climate-associated inversions that are in LD with allelic variants that confer resistance to a toxin produced by plants endemic to the warm Mediterranean regions (Pegueroles et al. 2016).

Apart from temperature, humidity, rainfall and solar radiation are other environmental variables that can contribute to cyclic changes in inversion polymorphism (Rodríguez-Trelles et al. 1996). In addition to climatic factors, heavy metals from highly polluted areas can exert selective pressure on inversion polymorphism in this species as well (Kenig et al. 2015).

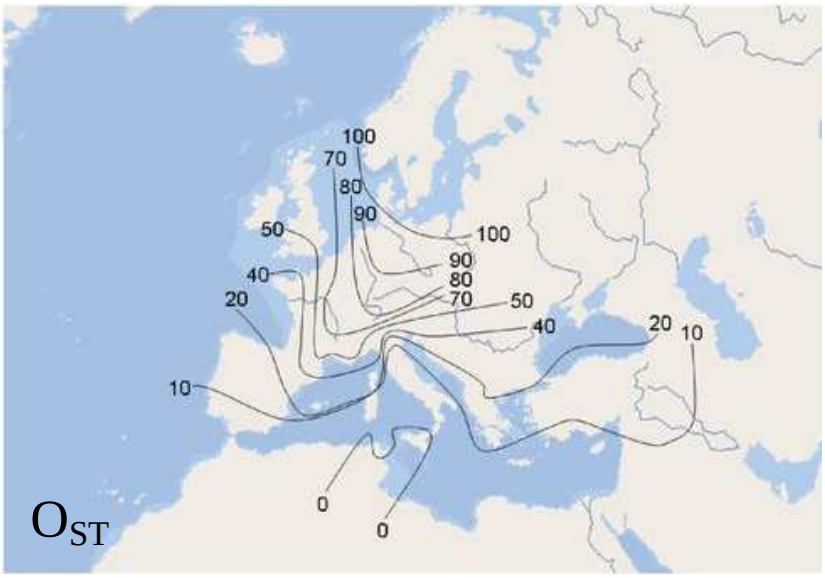


Figure 7: Frequency of the O_{ST} inversion in the Palearctic region (adapted from Dolgova et al. (2010)).

1.2.8. Candidate loci to explain adaptive inversions

Although there is ample evidence on the adaptive value of some *D. subobscura* inversions, very few is known about the molecular mechanisms responsible of it. The unassembled *D. subobscura* genome makes it a difficult task to search for candidate loci in different inversions. Furthermore, the closest species to *D. subobscura* with sequenced genomes are *D. pseudoobscura* and *D. persimilis*, which diverged too long ago to clarify any intra-specific events. Still, several studies have tried to identify loci linked with thermal adaptation and found their location in the *D. subobscura* genome.

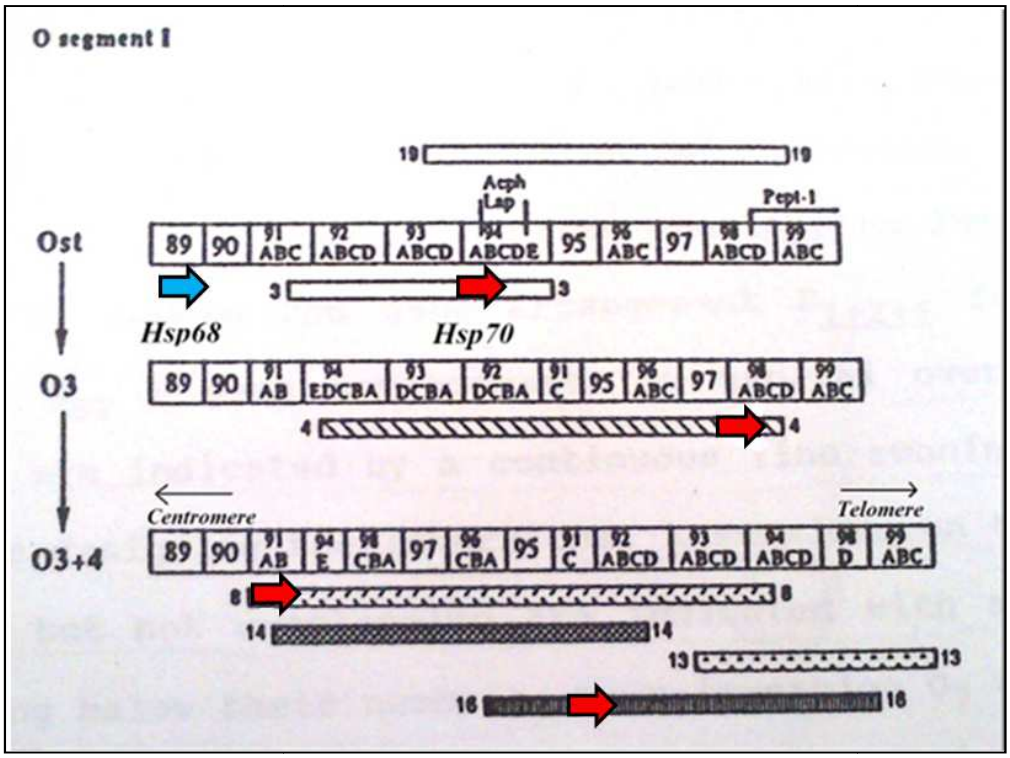


Figure 8: **Location of candidate loci to explain the adaptive inversion polymorphism of *D. subobscura*.** Approximate locations of *hsp68* and *hsp70* in the 89A (blue) and 94A (red) arrows in the SI segment of the O chromosome of an O_{ST} strain (Moltó et al. 1992). The 94A band was marked in a red arrow in three additional arrangements studied in this work ($O_{\underline{3+4}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+16}}$, adapted from Krimbas (1993)).

For instance, the heat shock resistance experiments carried out by Quintana and Prevosti (1991) revealed the presence of possible heat shock factors in the A chromosome and in the $O_{\underline{3+4}}$ inverted region in the SI segment of the O chromosome. In another heat shock experiment, DNA probes were constructed from several *D. melanogaster* genes encoding proteins of the heat shock family (HSPs) to find their homologous counterparts in *D. subobscura, D. guanche* and *D. madeirensis* strains. The probes hybridized on eight chromosomal puffs, indicating that loci in these regions augmented their transcription rates as a response to temperature increase. More precisely, two heat shock genes, *hsp70* and *hsp68*, localized respectively in the 94A and 89A bands in the SI segment of the O chromosome in these three closely related species (Moltó et al. 1992), which is in agreement with the results of Quintana and Prevosti (1991). Interestingly, the 94A band lies within the $O_{\underline{3+4}}$ inverted region in *D. subobscura*, where several adaptive inversions had been previously described, including $O_{\underline{3+4}}$ and O_{ST} (Figure 8). Another heat shock study measured Hsp70 protein levels before and after heat stress in flies homokaryotypic for different representative O-chromosome

arrangements that comprised the 94A band. Suprisingly, the warm climate-associated $O_{\underline{3+4}}/O_{\underline{3+4}}$ flies showed increased expression levels, under basal conditions, that did not boost with heat stress (Calabria et al. 2012). This result could be explained if the $O_{\underline{3+4}}$ strains had an extra number of *hsp70* copies that resulted in higher basal protein amounts. In fact, higher levels of *hsp70* mRNA and protein after heat stress are often observed in thermotolerant *Drosophila* strains with an extra number of *hsp70* genes (Garbuz et al. 2002; Garbuz et al. 2003; Evgen'ev et al. 2004). However, a higher number of copies does not always correlate with increased thermotolerance and with Hsp70 protein production because other *Drosophila* lineages have achieved high inducible Hsp70 levels, thermal resistance and a broad geographical distribution without increasing their *hsp70* copy number (Bettencourt and Feder 2001). Aside from copy number variation, *hsp70* mRNA and protein production can be modulated by transposable element (TE) insertions in the proximal promoter regions of *hsp70* genes (Bettencourt et al. 2002; Lerman et al. 2003; Lerman and Feder 2004). In Laayouni et al. (2007), an analysis of the transcriptome of *D. subobscura* stocks adapted to three nonstressful temperature regimes (13, 18 and 22ºC) revelated that a larger than expected number of loci, likely involved in thermotolerance, mapped inside inverted chromosomal segments. Interestingly, some of these loci, including genes of the heat shock protein family (HSPs), *Fst* and *Treh* had previously been linked with thermal resistance in *Drosophila*.

1.2.9. Inversions and arrangements considered in this work

To explore the molecular mechanisms behind extant *D. subobscura* adaptive inversion polymorphism, we have chosen four representative combined arrangements of the O chromosome. These include O_{ST}, $O_{\underline{3+4+16}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4}}$. Four of these arrangements occur in the segment I ($O_{\underline{3+4+8}}$, $O_{\underline{3+4}}$, O_{ST}, and $O_{\underline{3+4+16}}$) and two in the segment II (O_2, O_7) (see Figure 9 for a schematic representation). In this work we will focus on the arrangements in the SI segment, which includes chromosomal bands where candidate loci for adaptive inversion polymorphism have been localized as mentioned in the section above (Quintana and Prevosti 1991; Moltó et al. 1992) (see also Figure 8).

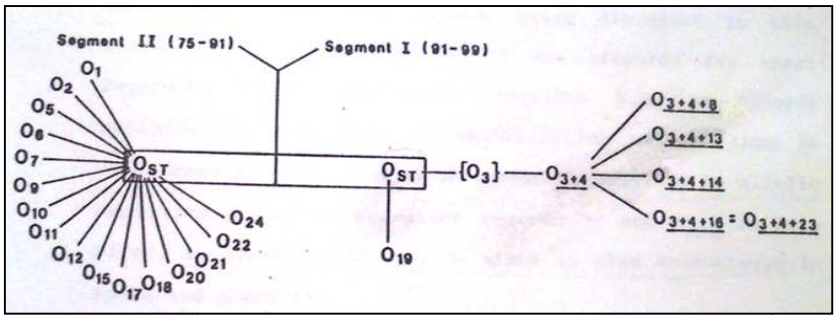


Figure 9: Schema representing the 24 described inversions and arrangements of the O chromosome in the Palearctic *D. subobscura* populations distributed across the I and II segments (adapted from Krimbas 1993). Notice that O_{3+4+16} and O_{3+4+23} represent the same arrangement.

According to various studies, the O_{ST} and O_{3+4} arrangements show marked and opposed latitudinal and seasonal components (Krimbas and Loukas 1980; Prevosti et al. 1988; Rodríguez-Trelles et al. 1996; Balanyà et al. 2003; Rodríguez-Trelles et al. 2013). In the Palearctic populations, the cold climate-associated O_{ST} is more frequent towards the North Pole, where it is close to fixation (90%). In contrast, the warm climate-associated O_{3+4} arrangement is less abundant in northern Europe and reaches its highest frequencies in the warmer Mediterranean environments, where O_{ST} is much less frequent or absent (Calabria 2012). The association between O_{3+4+8} and environmental factors is far from clear. Some studies have found a negative correlation between O_{3+4+8} and latitude that indicates that it belongs to the warm climate-associated category, reaching its highest frequencies (>80%) in certain North African populations (Krimbas 1993). In contrast, clinal frequencies of O_{3+4+8} in the recently colonized areas of South America do not parallel those in Europe and its association with allozyme alleles is more similar to those in O_{ST} (Prevosti et al. 1985; Balanyà et al. 2003). More recent collections indicate that O_{3+4+8} has increased its frequency north of the Pyrenees in the last years, which indicates that it is somehow adaptive (Balanyà et al. 2004; Rezende et al. 2010; Calabria 2012). The fourth and last studied arrangement, $O_{3+4+16+2}$, exhibits low frequencies (<5%) in most studied populations and it has not been associated with temperature fluctuations.

The phylogenetic relationship between the four arrangements in the segment I of the O chromosome (O_{3+4+16}, O_{ST}, O_{3+4}, O_{3+4+8}) is most consistent with an ancestral O_3 inversion giving rise to O_4 and O_{ST} independently in the *D. subobscura* lineage (Krimbas and Loukas 1984). The O_3 inversion is now extinct in contemporary *D. subobscura* populations, but it is fixed in the closely related species *D. madeirensis*. In

D. guanche, O_3 gave rise to the $O_{\underline{3+g}}$ inversion that is now fixed in this species. The nucleotide variation in the *rp49* gene region suggests that the $O_{\underline{3+4}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+16}}$ (sometimes labeled as $O_{\underline{3+4+23}}$, see Krimbas (1993)) are nested in a clade that excludes O_{ST} (Rozas et al. 1999) (Figure 10).

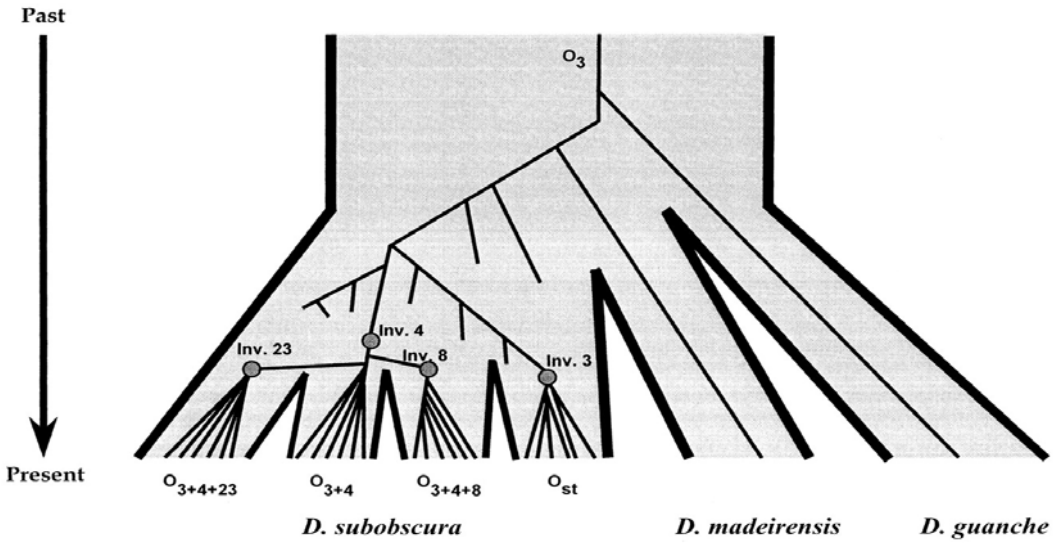


Figure 10: Schema showing the diversification of the four considered arrangements in *D. subobscura* and its two closest relatives, *D. madeirensis* and *D. guanche* (adapted from Rozas et al. (1999)).

$O_{\underline{3+4}}$ and O_{ST} sequences usually form well-separated clusters and show high F_{ST} values as indicated by the extant polymorphism of several loci located within the inverted segment (Navarro-Sabaté et al. 1999; Munté et al. 2005; Rozas et al. 1999; Pegueroles et al. 2013). In contrast, a rather limited degree of population genetic structure is observed between the closely related $O_{\underline{3+4}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+23}}$ arrangements (Rozas et al. 1999). In spite of high genetic differentiation, several GC tracts have been detected between $O_{\underline{3+4}}$ and O_{ST} *rp49* sequences, indicating that suppression of recombination is not complete in the heterokaryotype (Rozas and Aguadé 1994; Rozas et al. 1999).

1.3. Contribution of the *hsp70* gene family to thermal adaptation and adaptive inversion polymorphism

'The reason why plants and animals (including men) are so sensitive to temperature is that they are chemical machines… Many of them do burn even at ordinary temperatures, but so slowly that we don't notice any great change even within a lifetime.'

JBS Haldane, *Keeping cool and other essays* (1940, pp. 60-70)

The thermally inducible *hsp70* gene family is a candidate locus that might shed light on the molecular mechanisms responsible of the adaptive character of some *D. subobscura* inversions and the indisputable protagonist of this thesis work. Consequently, in the next lines I will focus on describing its most salient features in *D. subobscura* and other closely related *Drosophila*, including its classification, cellular roles, participation in the heat shock response and genomic organization.

1.3.1. The heat shock proteins

The proteins of the heat shock family (HSPs) are produced by cells in response to exposure to stressful conditions. HSPs are one of the most conserved family of proteins, being present in virtually all living organisms, from bacteria to humans. Although they were first described in relation to heat shock, it is now known that the expression of HSPs can be triggered by a broad scope of physical and chemical stressor agents, including exposure to cold, UV light, inflammation, exposure to toxins, hipoxia and tissue remodeling. A well-described function within the cell is the chaperone function. In this process, some HSPs participate by stabilizing nascent proteins to ensure correct folding or by helping to refold proteins damaged by cell stress. Other roles include intraorganellar protein shuttling and the targeting of old or terminally misfolded proteins to proteases for degradation. HSPs are also believed to play a role in the presentation of pieces of peptides on the cell surface to help the immune system recognize afflicted cells. All members of this highly conserved group of proteins share the common domain structure yet differ in their expression patterns, subcellular localization and cell functions.

HSPs are named according to their molecular weight, for example, Hsp60, Hsp70 and Hsp90 refer to families of HSPs on the order of 60, 70, and 90 kilodaltons in size, respectively. The small 8-kilodalton protein ubiquitin, which marks proteins for degradation, also has features of a heat shock protein. Most HSPs with chaperone activity belong to six conserved groups: Hsp40, Hsp60, Hsp70, Hsp90 and Hsp100 plus the small HSPs (sHSPs) that have molecular masses between 12 and 43 kilodaltons. Although these groups are conserved among virtually all taxa, certain homologues may differ in their molecular weight between species. For example, mammal Hsp90 proteins are orthologs of Hsp83 in *D. melanogaster*.

1.3.2. The heat shock response

A temperature increase of a few degrees may trigger the heat shock response (HSR), which is activated to protect cells against damage (Lindquist 1986; Morimoto 1993). The HSR consists in the transcriptional upregulation of genes encoding HSPs to prevent or minimize cell damage. It is generally controlled by a single transcription factor that in eukaryotes corresponds to the heat shock factor (HSF) and the sigma factor in *E. coli*. Under basal conditions, HSF exists as a monomer, but when stress induces protein damage, HSF is activated to trimerize. The trimer migrates to the nucleus, where it binds to heat shock elements (HSEs) in the promoter of heat shock genes to activate or substantially increase their transcript levels. Since constitutively high levels of the Hsp70 protein are deleterious for the cell in absence of stress (Feder et al. 1992), the mRNA is rapidly degraded after cellular stress by a self-limiting expression mechanism (Balakrishnan and De Maio 2006).

The HSR has been characterized in some *Drosophila*. In *D. melanogaster*, Hsp70 proteins are the most induced during thermal stress and are strongly associated with thermotolerance (Krebs 1999; Gong and Golic 2006), and several other HSPs have been shown to increase their expression levels on a par with the former, after thermal stress. Nonetheless, different HSP families may disagree considerably in the magnitude and tempo of their stress response. For example, some HSPs like Hsp60 appear to be thermally insensitive, while others like Hsp40 and Hsp83 show a significant increase in their transcript levels after heat exposure that confers substantial thermoprotection (Neal et al. 2006; Bettencourt et al. 2008). Nonetheless, mutant *D. melanogaster* strains with

no functional *hsp70* copies (*hsp70*-) show that the upregulation of some HSPs do not seem to improve their thermotolerance after severe heat shock (Bettencourt et al. 2008).

The HSR is influenced by the environmental context and is therefore prone to vary between organisms adapted to different niches. For example, in some fish species that survive at 0ºC, the HSR is activated at 5ºC, while in thermophilic bacteria that thrive at 50ºC the HSR starts at 60ºC. Variations in the HSR are also observed in strains of closely related species inhabiting contrasting environments. For example, the maximum induction of Hsp70 after heat shock is detected at 36-37ºC in wild-type *D. melanogaster* strains, while substantially increased levels of this protein are already observed after exposure at 32ºC in the more temperate *D. subobscura*. Interestingly, HSPs also respond to cold stress in the so-called cold stress response (CSR). Although this response has received less attention than HSR, there are some worth mentioning outcomes. A set of heat-shock genes are activated after exposing *D. melanogaster* adults during several hours at 0ºC and most of these genes are induced by heat stress as well. However, in contrast with the HSR, the highest transcript levels of these *hsp* genes are reported in the recovery phase, after returning the adults to their optimum temperature (Colinet et al. 2010).

1.3.3. The *hsp70* gene family: overview

The 70 kilodalton HSPs (Hsp70s or DnaK) belong to a highly conserved gene family with homologs from bacteria to mammals. For example, the amino acid sequences of human Hsp70 family members are roughly 50% identical to their homolog in *E. coli* (DnaK) and 75% to their Hsp70 homologs in *Drosophila.* Surprisingly, these identities are similar at the nucleotide (DNA) level, which indicates that synonymous sites are also subject to selection, for exmaple, by providing additional information in the nucleotide sequence or in mRNA secondary structure (Hunt and Morimoto 1985).

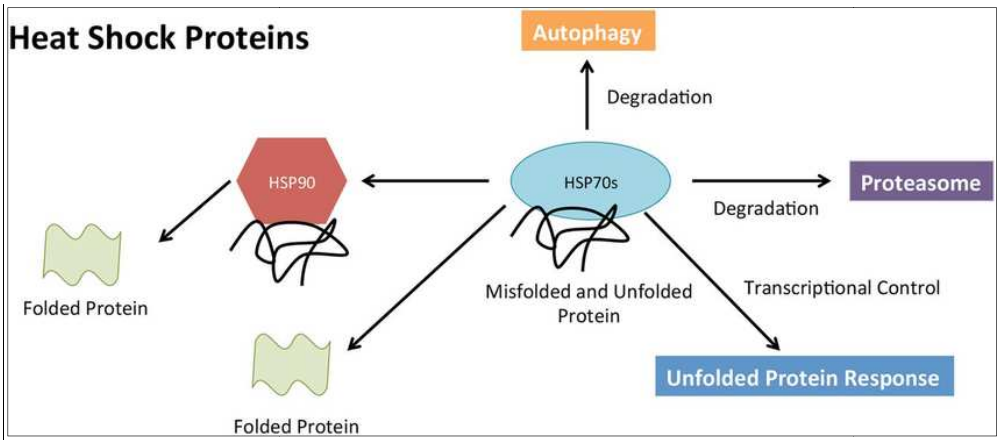


Figure 11: Schema showing the involvement of Hsp70 and other HSPs in protein folding and degradation (adapted from Nikesitch and Ling (2015)).

Hsp70s have housekeeping functions in the cell, where they are essential components of folding and signal transduction pathways. They also exhibit quality control functions in which they repair misfolded proteins and proofread their structure. The broad spectrum of cellular functions of Hsp70 proteins has been achieved through amplification and diversification mechanisms. Co-chaperone proteins are selectively recruited by Hsp70 chaperones to fulfill specific cellular functions, such as the stimulation of chaperone ATPase activity. Interestingly, Hsp70 chaperones have been shown to cooperate with other chaperone systems like Hsp90, which has broadened their activity spectrum (Figure 11).

The detection of Hsp70 in the tissues of diverse taxa has been used as a biomarker for environmental stress (Nadeau et al. 2001; Ireland et al. 2004; De Jong et al. 2006). Perhaps not surprisingly, Hsp70 proteins seem to play an important role in senescence. Indeed, Hsp70s are upregulated in tissue-specific patterns during aging, where they are thought to counteract the toxicity of misfolded proteins. Since Hsp70 protein levels have shown to correlate with life span, they can also be used as biomarkers of aging (Tower 2011).

1.3.4. The *hsp70* gene family in *Drosophila*

In the genus *Drosophila*, the *hsp70* gene family includes proteins that share a high degree of homology among themselves. These are divided into the HSCs or heat shock cognates (*hsc70s*), *hsp70* and *hsp68*. The *hsp70* and *hsp68* genes are generally described as inducible (meaning that basal levels are *per se* low and normally only increase under certain stressful conditions), whereas Hsc70s are constitutively expressed

in the cytosol. The *hsp70* genomic organization has been extensively studied in *Drosophila*. For instance, in the 25-40 MYO *Sophophora* subgenus that encompasses among others the *obscura* and *melanogaster* groups (Russo et al. 1995; Obbard et al. 2012), *hsp70* homologs are organized in groups of two or more copies that are distributed across one or two gene clusters located in Müller's E element (Figure 12). For example, some *Drosophila* groups, including *obscura* and *willistoni*, appear to have two *hsp70* genes at one cluster in a peculiar inverted-pair arrangement that resembles the ancestral condition, while some *melanogaster* subgroups have four or more copies copies distributed in two clusters (Moltó et al. 1992; Bettencourt and Feder 2001). Interestingly, the *hsp70* organization of *obscura* is analogous to that in the mosquito *A. albimanus* and is thought to have been present in the last common ancestor of all dipterans. This indicates that the original inverted pair duplicated at least once in *Drosophila*, resulting in a different number of copies between groups. Different mechanisms have been postulated to explain the *hsp70* copy number variation in *Drosophila*. For instance, a retrotransposition-mediated duplication probably took place in the common ancestor of all *melanogaster* subgroups 10-15 MYA, yielding two additional copies in these. In contrast, the fifth copy in *D. melanogaster*, the extra copies in *D. virilis* and *D. lummei* and a third copy in the *montium* subgroup seem to have originated via tandem duplication (Konstantopoulou et al. 1998; Bettencourt and Feder 2001; Evgen'ev et al. 2004). The five to six *hsp70* copies in *D. melanogaster* occur in two separate clusters inside the 3R chromosome, namely 87A7 and 87C1 (Brown and Ish-Horowicz 1981). The extra copies in this species do not merely serve to rapidly produce large *hsp70* transcript amounts, but have been shown to be under specific developmental-stage regulation (Lakhotia and Prasanth 2002). As exemplified above, *hsp70* copy number variation occurs between closely related species and strains of the same species as well; on this line, a maximum of seven copies are found in the more thermotolerant *D. virilis* and five in the less tolerant *D. lummei* (Evgen'ev et al. 2004).

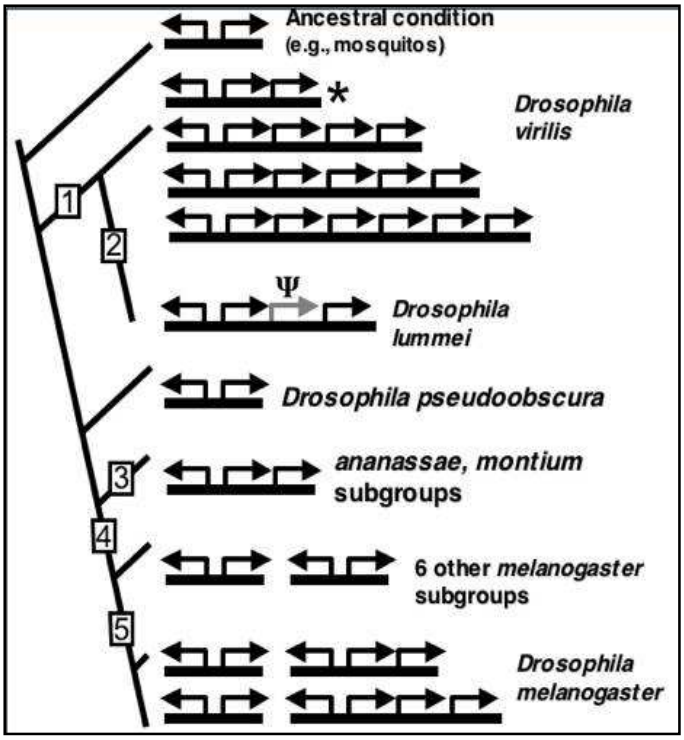


Figure 12: *Hsp70* copy number variation in *Drosophila* (adapted from: http://pondside.uchicago.edu/~feder/researchD.html).

1.3.5. The *hsp70* gene family in *D. subobscura*

Until recently, very little was known about the genomic organization and copy number variation of the *hsp70* gene family in *D. subobscura*. A heat shock experiment using a set of probes that targeted several *hsp* genes determined that *hsp70* localized to the 94A band in a O_{ST} strain (Moltó et al. 1992). Interestingly, the 94A band is located within the O_{3+4} inverted region, which shows high genetic differentiation between the cold climate-associated O_{ST} and the warm climate-associated O_{3+4} arrangements (Rozas et al. 1999; Munté et al. 2005). A second hybridization signal was detected in the 89A band in the same chromosomal region that corresponded to *hsp68*, also member of the *hsp70* gene family. In *D. subobscura*, *hsp68* shows moderate identity (~75%) to *hsp70* at the sequence level and might be associated with thermal resistance as well (Laayouni et al. 2007). Interestingly, it has been postulated that *hsp68* originated from a duplication of the *hsp70* locus that took place early in *Drosophila* evolution, with subsequent loss of one copy in most lineages. The primitive duplicated pair has been preserved only in some *Drosophila* species like *D. virilis* (Velikodvorskaia et al. 2005).

1.3.6. Gene conversion and the evolution of multigene families

1.3.6.1. Overview and mechanisms

The multiple *hsp70* copies in *Drosophila* are usually found in clusters that include two or more members (Bettencourt and Feder 2001, 2002). The different copies are thought to have originated from successive duplication events and should therefore be described as paralogs. Surprisingly, some paralogs in the same cluster are more similar within the same species than to their orthologuous counterparts in more distant lineages as evidenced in *D. melanogaster* (Bettencourt and Feder 2002). Since in many cases the duplication is thought to have occurred long before the divergence of these lineages, this pattern indicates that gene conversion (GC) takes place at these loci. Considering that GC seems to have played an important role in the evolution of multigene families like *hsp70*, I will spare some lines to describe the molecular mechanisms enabling recurrent GC events and their evolutionary significance.

GC is described as a process by which one DNA sequence replaces a homologous sequence yielding two identical sequences. This process is termed *allelic* when one allele replaces another allele of the same gene, or *ectopic* when it takes place between two different loci. Interlocus GC may be divided into *cis*, if GC involves two paralogs that are found in the same chromatid, or *trans* if these are located in sister chromatids or homologous chromosomes. The mechanisms by which GC occurs are not well understood, but it is thought to be initiated when a double strand break (DSB) is produced. DSBs can be repaired by a series of homologous recombination (HR) interrelated pathways. Two pathways that can produce GC events are the double-strand break repair (DSBR) and the synthesis-dependent strand annealing (SDSA) (see Figure 13). These two mechanisms show some similarites and differences. Briefly, the two processes start when a 3' end invasion of a single strand into a homologous double strand occurs, forming a (D)-loop structure and the newly paired strand is used as template to initiate the reparative synthesis (**a**). The two pathways diverge in the next step: in the case of DSBR, DNA synthesis takes place at the two broken ends, with the consequence that information is copied from both strands of the complex (**b**), whereas the genetic information is only copied to one strand in the SDSA pathway (**c**). Subsequently, in SDSA, the extended strand hybridizes with a single strand resulting from the resection of the other end of the break, while in DSBR a double Holliday

junction (dHJ) is formed. A notable difference is that SDSA exclusively results in non-crossover GC tracts, while in the case of DSBR the dHJ resolution can generate both crossover and non-crossover GC products. There is no precise way to distinguish between recombinant tracts originating from one or the other mechanism. Nonetheless, the tracts produced by double crossovers are expected to be somewhat longer (>1 kb) than GC tracts that rarely exceed 1 kb (Rozas et al. 1999; Chen et al. 2007).

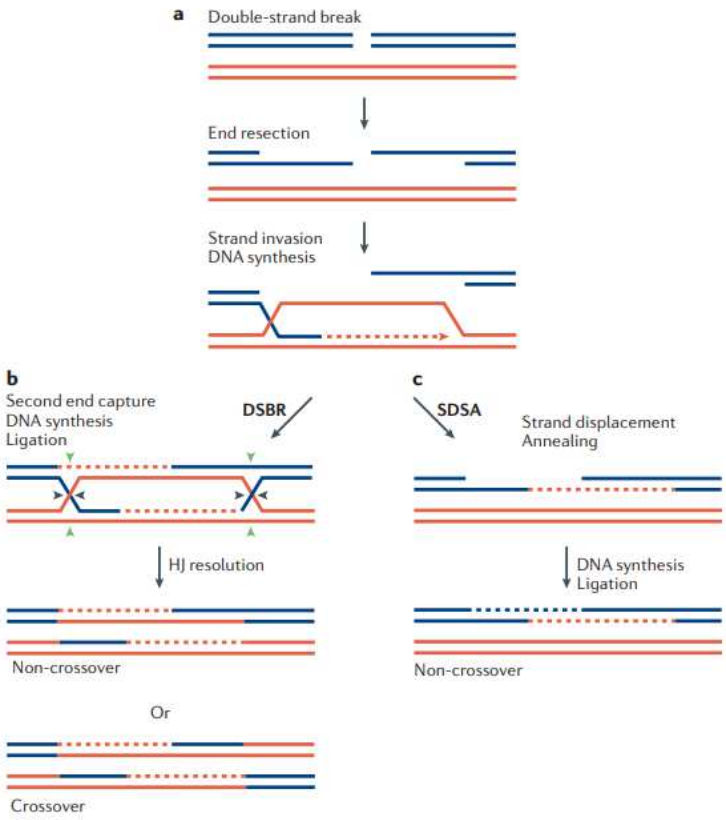


Figure 13: DSBR and SDSA, two proposed mechanisms to explain GC events (extracted from Sung and Klein 2006).

Efficient GC seems to require high (>80%) identity at sequence level and it tends to increase as the physical distance between the interacting sequences diminishes (Ezawa et al. 2006). In addition, although GC can occur between identical regions as small as 10 bp (Mézard et al. 1992), later studies have shown that GC rates are significantly higher for paralogs that have >250 bp identical regions, which introduced the concept of minimal efficient processing segment or MEPS (Jinks-Roberston et al. 1993; reviewed in Mansai et al. 2011). It is thought that certain motifs, including alternating purine and pyrimidine nucleotides, polypurine and polypyrimidine palindrome-like sequences, minisatellites and Z-DNA among others might influence the frequency of GC events.

However, many GC examples do not include any of these motifs (reviewed in Chen et al. 2007).

It has been noted that interlocus GC events may show biased directionality. In this line, some authors have identified a donor (or "master") and an acceptor (or "slave") locus. Sometimes, the bias between the donor and the acceptor might be as high as 20-fold. It is thought that directionality of GC correlates with the relative levels of expression of the two genes involved, with the more expressed gene converting more often the gene that is less expressed. In the human genome, it has been observed that pseudogenes that do not follow this rule are often associated with certain diseases. Consequently, the slave-to-master direction is rarely observed in the human genome because it would probably generate defective alleles that are doomed to extinction. Still, deleterious mutations will sometimes spread to functional alleles. In this line, a number of human diseases including some cancer types seem to be caused by GC between a pseudogene as donor and an active gene as receptor through the introduction of frameshifting, aberrant splicing, nonsense mutations and deleterious missense mutations among others. At the same time, it has been proposed that GC could reverse inherited deleterious alleles by a process termed "natural gene therapy". In this process, the mutant phenotype is reverted back to the normal phenotype by the wild-type allele or by interallelic GC in those diseases caused by different mutations in the two alleles. Such mechanisms could be used in the future to treat the clinical manifestation of these diseases (Chen et al. 2007).

1.3.6.2. Evolutionary significance

Interlocus GC is thought to play an important role in the concerted evolution of multigene families and repetitive DNA. The most direct consequence of this mechanism is that paralogous loci become more similar at the sequence level to each other than they are to their orthologous counterparts in other closely related species. On this line, a pitfall of using genes that are suspicious to undergo recurrent GC events in phylogenetic analyses is that GC periodically erases the divergence between paralogous duplicates, which results in gene trees that are inconsistent with species trees. This is due to the fact that the observed divergence between duplicates will always be less than the expected under a molecular clock model.

GC has been detected between duplicates that lie close in the genome of diverse species (Russnak and Candido 1985; Balding et al. 1992; Rozen et al. 2003) and its biological and evolutionary implications are many and diverse. For instance, interallelic GC seems to frequently occur in gene families that are highly polymorphic, including the AB0 system and the HLA class II gene region in humans. Indeed, a large number of new alleles in the HLA system seem to have been produced by GC between the parental alleles, indicating that GC increases genetic diversity via the formation of new haplotypes (Chen et al. 2007; Fawcett and Innan 2011). Another important evolutionary consequence of GC is the retention of sequence identity between gene duplicates when there is selection to produce higher amounts of the same protein (Fawcett and Innan 2011). Interestingly, it is thought that GC may slow down the Y-chromosome degradation process in mammals (Trombetta and Cruciani 2017). Since primate Y chromosomes are unable to recombine for the most part, GC between duplicates may work as an analogous mechanism to correct genetic mistakes and maintain the integrity of the few functional genes in this chromosome (Rozen et al. 2003).

An evolutionary explanation for recurrent GC events between *hsp70* duplicates in some *Drosophila* postulates that duplicates allow greater evolutionary exploration of mutational space while buffering against deleterious changes, and that the interaction of selection and conversion ensures that only neutral or beneficial mutations spread (Bettencourt and Feder 2001, 2002). In other words, GC in multigene families such as *hsp70* in *Drosophila* would simulate an increase in the effective population size that enables more efficient selection by promoting a faster spread of benefical mutations and a more efficient elimination of deleterious ones (Mano and Innan 2008; Fawcett and Innan 2011).

1.3.6.3. Parameters influencing the rate of GC events

The frequency of GC events is determined by several factors. It has been theorized that substantial identity at the nucleotide level between duplicates, is necessary to enable recurrent GC events. According to the mathematical model developed in Rozen et al. (2003) for palindrome arms on primate Y chromosomes, the nucleotide divergence (*d*) between two *de novo* arisen duplicates will increase until reaching an equilibrium value after a certain number of generations. This equilibrium value will be reached sooner or later depending on the values of two parameters that counteract each other: the mutation

rate (μ) and the GC rate (c) that periodically erases the accumulated divergence between the two duplicates. Other models suggest that once the nucleotide divergence between the two duplicates surpasses a threshold, it will start increasing roughly linearly. At this stage, concerted evolution is said to be terminated (Fawcett and Innan 2011). Interestingly, simulations involving HERV15 proviral paralogous sequences in the Y chromosomes of three primate species (Hurles et al. 2004) indicate that GC can exaggerate sequence divergence between orthologs when the two paralogs have already accumulated some variation before speciation. These effects are not observed when the two paralogs are identical before speciation.

1.3.6.4. GC and the evolution of *hsp70* in *Drosophila*

Examples of GC between duplicates in *Drosophila* include the five to six *hsp70* copies in *D. melanogaster* that are distributed across two ~500 kb apart clusters in the chromosome arm 3R (Bettencourt and Feder 2001). The 87A7 cluster contains two palindromic and divergent copies separated by a 1.7 kb long intergenic region that includes a transposable element, while the 87C1 locus includes the proximal *hsp70Ba* copy that is transcribed toward the centromere and is separated by ~38 kb of repetitive DNA from the three other copies, *hsp70Bbb*, *hsp70Bb*, and *hsp70Bc* that are transcribed in the opposite direction (Gong and Golic 2004). The copies within each cluster are almost identical between each other, while they show moderate divergence with those in the other cluster (Figure 14). Other *Drosophila* species whose *hsp70* duplicates undergo recurrent GC include *D. orena* and *D. simulans* (Bettencourt and Feder 2001). Outside *Drosophila*, convergence has also been detected in the *hsp70* inverted pair of the mosquito *A. albimanus* (Benedict et al. 1993). In contrast, divergence seems to predominate over convergence in certain species like *D. mauritiana*, where five of the eight analyzed *hsp70* alleles had likely degenerated into pseudogenes (Bettencourt and Feder 2001). In that study, no association between thermotolerance and *hsp70* copy number or identity between paralogs in the same species was noted after examining various lineages of the *melanogaster* subgroup. However, although *hsp70* gene loss or duplication is not the only mechanism to attain increased thermotolerance and *hsp70* expression levels, it appears to be an important one that has occurred multiple times in evolution.

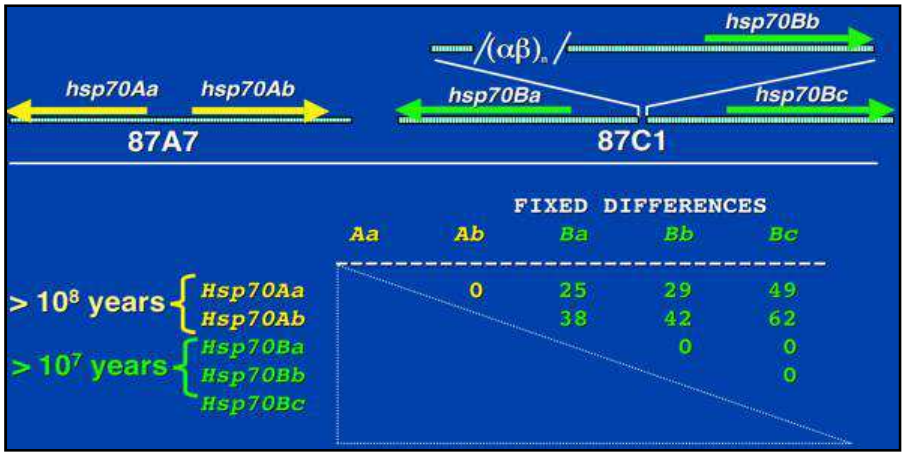


Figure 14: GC and the evolution of the *hsp70* gene family in *Drosophila*. The schema shows the number of fixed differences at the nucleotide level between *hsp70* copies distributed across the 87A7 and 87C1 clusters in *D. melanogaster*. Notice that copies within the same cluster are indistinguishable from each other (adapted from: http://pondside.uchicago.edu/~feder/researchD.html).

1.3.6.5. Is GC driving the evolution of the *hsp70* genes in the *obscura* group, too?

To explore the organizaton of *hsp70* genes in the *obscura* group of *Drosophila*, we searched for *hsp70* homologs in the genomes of the closely related *D. pseudoobscura* and *D. persimilis* that are available from the FlyBase servers (www.flybase.org). This enabled us to assemble a ~10 kb genomic region comprising two *hsp70* copies and their 5' and 3' flanking regions. These two sibling species belong to the Nearctic subgroup within *obscura*, and are therefore much more closely related between themselves than to the *subobscura* species cluster, from which they diverged 10-15 MYO (Russo et al. 1995). Consistent with previous work (Moltó et al. 1992) the *hsp70* genomic organization in *D. pseudoobscura* and *D. persimilis* appears to be similar to the ancestral two-copy cluster that also exists in several other *Drosophila* groups (Bettencourt and Feder 2001) and consists in two ~2.7-kb long divergent copies separated by ~2.5 kb of intergenic, nonduplicated DNA. Sequence alignments of the of the two *hsp70* copies in *D. pseudoobscura* and *D. persimilis* indicate that their coding regions (CDS) and proximal promoters are somewhat more similar within the same species. Although our results are preliminary, this observation is compatible with periodic GC events taking place in both *D. pseudoobscura* and *D. persimilis* assuming that the duplication that led to the inverted pair predates their divergence.

1.3.6.6. Other mechanisms to explain the evolution of multigene families

It is worth mentioning that some *hsp70* copies seem to have escaped GC and are distinguishable from each other. Sometimes, one or various copies have become pseudogenes, as in *D. lummei* and *D. mauritiana* (Bettencourt and Feder 2001; Evgen'ev et al. 2004). GC is often difficult to tell apart from the birth-and-death evolution model of multigene families. One difference between the two mechanisms is that GC predicts that the duplicated sequences will be more similar within the same species, while the birth-and-death model postulates that highly identical gene duplicates will coexist side-by-side with more divergent genes and pseudogenes originating from ancient duplication events (Figure 15). This is so because under the birth-and-death model, groups of new functional genes are constantly evolving, whereas concerted evolution does not allow the functional differentiation of genes, because all family members evolve as a unit (Nei and Rooney 2005). The evolution of the *hsp70* gene family in *Drosophila* has been cited by some authors as an example involving both processes, GC and the birth-and-death process. An interplay between these two processes has apparently driven the evolution of the *hsp70* genes in two sibling nematode species as well (Nikolaidis and Nei 2004).

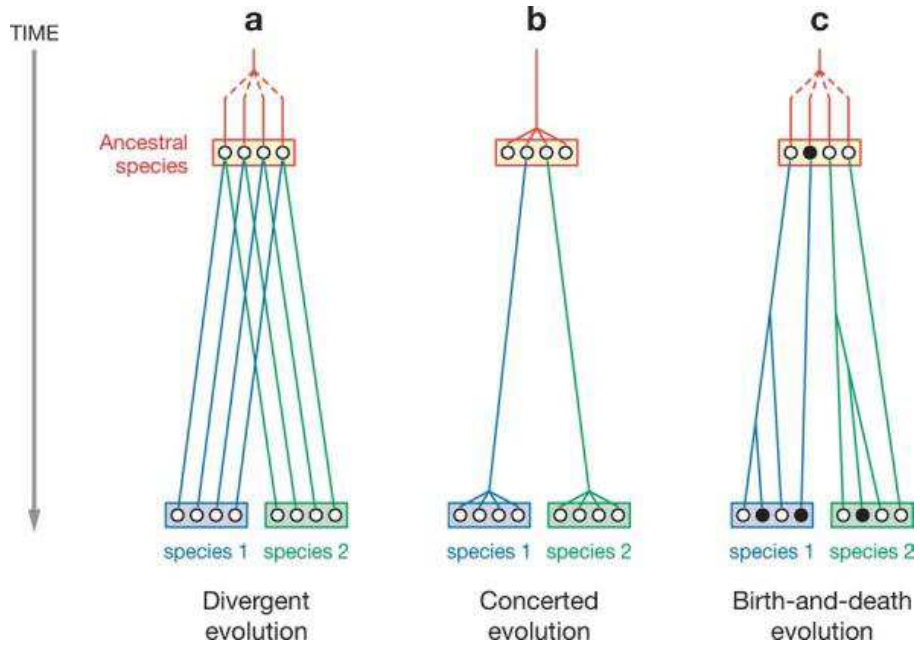


Figure 15: Three models of multigene family evolution and their predicted genes trees (extracted from Nei and Rooney 2005).

1.3.7. *Hsp70* cluster organization in *Drosophila* and its relationship with thermotolerance

Interestingly, the copy organization within a cluster seems to have undergone substantial rearrangements between certain closely related lineages. A peculiar compact disposition and a high degree of identity between paralogous *hsp70* copies is sometimes observed in more thermotolerant species, while interspersed and more divergent copies are found in less thermotolerant ones. Some authors have proposed that close proximity and recurrent GC may be two requirements to mount a fast and highly coordinated response favorable in challenging conditions (Evgen'ev et al. 2004; Garbuz et al. 2011). Still, high compactness could result from increased rates of DNA loss rather than natural selection (Petrov and Hartl 1998; Garbuz et al. 2011). In summary, the evidence presented here suggests that the genomic organization of *hsp70* copies is prone to vary among strains of the same species in spite of its high conservation, although it is not clear to what extent do these rapid changes relate to environmental pressure.

1.3.8. The relationship between *hsp70* expression patterns and thermotolerance in *Drosophila*

within the genus *Drosophila*, several lines of research have investigated the inducible nature of the *hsp70* gene family and its relationship with thermotolerance defined as the ability to withstand relatively hot conditions (reviewed in Feder and Hofmann (1999)). A positive link between inducible Hsp70 protein levels and thermal resistance is likely present in *D. melanogaster* at the larval stage, while such correlation may not be so discernible in other drosophilids such as *D. mojavensis* (Krebs 1999). The *hsp70* inducible response is influenced by its biological context and shows variation attributable to a particular developmental stage (Lakhotia and Prasanth 2002; Sørensen and Loeschcke 2002), adaptation to different environments (Sørensen et al. 2001), inbreeding (Kristensen et al. 2002), larval crowding (Sørensen and Loeschcke 2001), gender (Dahlgaard et al. 1998), line (Krebs and Feder 1997a) and tissue specificity (Lakhotia and Prasanth 2002). H*sp70* expression is likely under the influence of *cis*-regulatory elements (CREs) that lie in the promoter. For instance, the insertion of TEs in the *hsp70* promoters translates into reduced expression levels, lower thermotolerance and increased fecundity in *D. melanogaster* strains (Zatsepina et al. 2001; Lerman et al. 2003; Lerman and Feder, 2004). Interestingly, certain features of the heat shock

promoters, such as an accessible chromatin conformation, seem to significantly increase the rate of *P* element insertions in these regions (Walser and Feder 2006).

There is evidence indicating that closely related strains adapted to different environments might differ in their *hsp70* stress response. In general, warm climate-associated strains tend to have a higher temperature threshold and show a rather weak *hsp70* induction response in comparison with the more temperate ones. However, the *hsp70* induction profile does not always correlate with thermal resistance, and while in some cases the more thermotolerant *Drosophila* strain shows increased Hsp70 protein and *hsp70* mRNA levels after heat stress relative to the less thermotolerant one (Feder et al. 1996; Garbuz et al. 2002, 2003; Evgen'ev et al. 2004; Bettencourt et al. 2008), the opposite trend has also been observed (Sørensen et al. 2001; Zatsepina et al. 2001). It is also worth noting that higher Hsp70 levels might result in lower fitness under moderate stressful and nonstressful conditions as a consequence of increased energetic demand (Roberts and Feder 2000; Hoekstra and Montooth 2013), which might help explain why some *Drosophila* strains perpetually exposed to high temperatures seem to have evolved decreased instead of increased Hsp70 expression. Nonetheless, the outcomes of Calabria et al. (2012) suggest that the advantages of having high basal Hsp70 levels might outnumber the costs in certain *D. subobscura* strains wearing the warm climate-associated $O_{\underline{3+4}}$ arrangement. However, this pattern is rarely observed in species with short life cycles as *Drosophila*, where perpetually increased Hsp70 protein levels have multiple negative consequences for the individual, including a longer time for full development, reduced cell growth rates, decreased survival into adulthood and reduced fecundity (Feder et al. 1992; Krebs and Feder 1997b; Williams et al. 2003; Garbuz and Evgen'ev 2017). These observations make the results of Calabria et al. (2012) even more unconventional.

1.3.9. Inversion polymorphism, thermal preference and thermotolerance in *D. subobscura*

Considering that warm climate-associated arrangements in *D. subobscura* are positively associated with temperature, it is tempting to speculate if the individuals bearing them differ in some phenotypic traits such as thermotolerance from those bearing cold climate-associated arrangements.

An association between adaptive *D. subobscura* inversions and phenotypic and behavioural traits has been explored in some studies. For instance, in summer, O_{3+4} carriers appear to be more active towards midday (Savkovic et al. 2004), and the most active flies show a lower body temperature (*Tb*) at higher latitudes and in colder seasons. This finding may help to explain the latitudinal clines observed for various traits, including inversion polymorphism (Huey and Pascual 2009). Rego et al. (2010) noticed that adult *D. subobscura* flies wearing cold climate-associated versus warm climate-associated arrangements displayed different thermal preferences congruent with their clinal patterns, with some chromosomes contributing more than others. Dolgova et al. (2010) found that O_{3+4}/O_{3+4} flies did not exhibit increased thermotolerance relative to their O_{ST} counterparts, although the former seemed to prefer warmer environments, which is in agreement with Rego et al. (2010). Regarding these outcomes, Dolgova et al. (2010) concluded that if this pattern were general to all chromosomes, any correlation between thermal resistance and thermal preference across latitudinal gradients would probably reflect a pattern of correlated selection rather than genetic correlation.

In Calabria et al. (2012) higher basal Hsp70 protein levels were reported for homokaryotypic O_{3+4}/O_{3+4} strains relative to their O_{ST}/O_{ST} and O_{3+4+8}/O_{3+4+8} counterparts. In the same work, adult females but not males of the O_{3+4}/O_{3+4} strain showed increased basal thermotolerance relative to O_{ST}/O_{ST} females under a slow ramping assay and slightly higher basal and heat-induced thermotolerance in the two O_{3+4}/O_{3+4} sexes under a fast ramping. However, the increased heat-induced thermotolerance in O_{3+4}/O_{3+4} flies cannot be entirely caused by higher Hsp70 levels, because these do not differ from those of O_{ST}/O_{ST} after heat stress. This lack of a straightforward association between increased Hsp70 protein levels and thermotolerance in adult O_{3+4}/O_{3+4} flies could be explained by three main causes according to Calabria (2012). First, it is possible that Hsp70 does not differ substantially at the nucleotide level between O_{3+4} and other *D. subobscura* arrangements like O_{ST}. Secondly, the association between Hsp70 levels and thermotolerance may be only present in the larval stage as observed in *D. melanogaster* (Jensen et al. 2010). Thirdly, the flies carrying different chromosomal arrangements may show dissimilar phenotypical plastic responses that result in unpredictable patterns between basal and induced thermotolerance (Rezende et al. 2011).

2. OBJECTIVES

The main objectives of this work are to describe and analyze the thermally inducible *hsp70* gene family at the cytological, molecular and expression levels (including mRNA and protein) in several representative *D. subobscura* arrangements O_{3+4+8}, $O_{3+4+16+2}$, O_{ST} and O_{3+4} . All these arrangements include the chromosomal region where this locus has been localized. The cold climate-associated O_{ST} and the warm climate-associated O_{3+4}, are regarded as adaptive and show opposed and marked seasonal and latitudinal trends. On this line, the characterization of the *hsp70* family at the molecular level could give hints about their adaptive character and explain the differences observed between the warm-adapted O_{3+4} and O_{ST}/O_{3+4+8} in Hsp70 protein production before and after the heat shock.

Before the experiments carried out in this thesis, it was known that the *hsp70* locus was single and lied within the I segment of the O chromosome in the *Drosophila subobscura* species cluster, although its genome organization and *hsp70* copy number remained unknown. The two closest species to *D. subobscura* with publicly available genomes, *D. pseudoobscura* and *D. persimilis*, diverged from *D. subobscura* 10-15 Mya, a split that is too old to infer intra-specific evolutionary events. In addition, it has not been possible to reconstruct the *hsp70* locus from the unassembled *D. subobscura* genome. At this point, various research lines had to be planified to characterize the *hsp70* locus in the aforementioned chromosomal arrangements.

To accomplish our main objectives, we developed several sub-objectives that have enabled us to locate all *hsp70* loci in the genomes of O_{3+4+8}, $O_{3+4+16+2}$, O_{ST} and O_{3+4} and to reconstruct a 9-10-kb long genomic region comprising two *hsp70* genes in these and in the closely related *D. madeirensis* and *D. guanche*. Concomitantly, we quantified basal *hsp70* expression levels at the mRNA and protein stages in adult males and females of the coldadapted O_{ST} and the warm-adapted O_{3+4}.

These sub-objectives are listed as follows:

1. Identify by *in situ* hybridization (ISH) the genomic location of *hsp70* loci in the polytene chromosomes of third-instar larvae of lines isogenic for the $O_{\underline{3+4+8}}$, $O_{\underline{3+4+16}}$, O_{ST} and $O_{\underline{3+4}}$ gene arrangements and detect possible differences in the number of hybridization points between them.

2. Construct a genomic library in an isogenic O_{ST} line to assemble the sequences homologous to *hsp70* and reconstruct a 9-10 kb genomic region that encompasses two virtually identical *hsp70* paralogous genes and parts of the neighboring *CG5608* and *Dmt* loci that served to distinguish them.

3. Use the aforementioned sequenced region in the O_{ST} line and contigs from the unassembled *D. subobscura* genome to reconstruct, describe, compare and analyze the equivalent genomic region in 11 additional lines isogenic for the O_{ST} (3), $O_{\underline{3+4+8}}$ (1), $O_{\underline{3+4}}$ (5) and $O_{\underline{3+4+16}}$ (2) gene arrangements.

4. Use the extant polymorphism in *D. subobscura* to reconstruct the *hsp70* gene family in the closely related *D. madeirensis* and *D. guanche* to analyze, describe and compare the genomic organization at the *Hsp70IR* locus in the *subobscura* species cluster.

5. Use the available cytological and sequence information of less closely related species such as *D. pseudoobscura* and *D. persimilis* to infer the ancestral organization of the *hsp70* gene family in the 10-15 MYO *obscura* species group.

6. Analyze, describe and quantify the extant polymorphism and annotate several conserved *cis*-regulatory elements (CREs) in the proximal promoter and coding regions of the two *hsp70* paralogous genes that appear to evolve in concert by means of ectopic gene conversion in the four considered *D. subobscura* arrangements and in *D. madeirensis*.

7. Analyze, describe and compare the evolutionary patterns of seven distinctive regions included within the 9-10 kb reconstructed *Hsp70IR* locus in 12 *D. subobscura* lines isogenic for the $O_{\underline{3+4+8}}$ (1), $O_{\underline{3+4+16}}$ (2), O_{ST} (4) and $O_{\underline{3+4}}$ (5) arrangements to explore a possible link between the extant polymorphism across the *Hsp70IR* locus and the adaptive character of O_{ST} and $O_{\underline{3+4}}$.

8. Use the extant polymorphism at the *Hsp70IR* locus to explore the phylogenetic relationships within the *subobscura* cluster and the chromosomal inversion ancestry of our 12 *D. subobscura* lines, measure the degree of genetic structure between gene arrangements and calculate the coalescence times for our $O_{\underline{3+4+16}}$ (2), $O_{\underline{3+4}}$ (5) and O_{ST} (4) haplotypes.

9. Analyze, describe and compare the proximal promoter regions of the two *hsp70* genes in six cold-adapted O_{ST} lines and six warm-adapted $O_{\underline{3+4}}$ lines used for *hsp70* mRNA and protein quantification and annotate in these several previously identified CREs, including four heat shock elements (HSEs), three GAGA factor binding sites, the TATA box and the transcription start site.

10. Analyze, quantify and compare basal *hsp70* mRNA and Hsp70 protein levels in several replicates of adult males and females belonging to six cold-adapted O_{ST} lines and six warm-adapted $O_{\underline{3+4}}$ lines.

3. RESULTS

The results section has been divided into three parts that cover the complete thesis work.

Chapter 3.1: we describe the cytogenetic location, genomic organization and nucleotide diversity patters of the *hsp70* gene family in four representative *D. subobscura* arrangements (O_{ST} , $O_{\underline{3+4}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+16}}$) and in the closely related *D. madeirensis* and *D. guanche*. The results are discussed in the context of the adaptive character of some inversions in *D. subobscura*, including O_{ST} and $O_{\underline{3+4}}$.

Chapter 3.2: we quantified *hsp70* mRNA and protein levels in basal conditions in males and females of several isogenic lines for the cold-adapted O_{ST} and the warm-adapted $O_{\underline{3+4}}$ arrangements, to detect differences that might relate to their adaptive value. The results are contrasted with the *cis*-regulatory polymorphism in the *hsp70* proximal promoter regions of these lines.

Chapter 3.3: we used an expanded data set of *D. subobscura* isogenic lines, for the aforementioned arrangements, to make more precise estimates on the nucleotide polymorphism levels and to test for departures from neutrality in the adaptive O_{ST} and $O_{\underline{3+4}}$ using a ~10 kb reconstructed region in the *Hsp70IR* locus.

3.1 Testing for differences in *hsp70* copy number and genome organization between the O_{ST}, O_{3+4}, O_{3+4+16} and O_{3+4+8} arrangements of *D. subobscura*

This section includes the research article **"**Chromosomal inversions promote genomic islands of concerted evolution of *Hsp70* genes in the *Drosophila subobscura* species subgroup" that was published by the *Molecular Ecology* journal on February 7, 2018 and can be purchased from the *Wiley Online Library* (https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.14511). Likewise, the Supplementary Material of this article can be found in Appendix 7.1.

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

Received: 21 November 2017 | Revised: 15 January 2018 | Accepted: 17 January 2018

DOI: 10.1111/mec.14511

SPECIAL ISSUE: THE ROLE OF GENOMIC STRUCTURAL VARIANTS IN ADAPTATION AND DIVERSIFICATION

WILEY MOLECULAR ECOLOGY

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

Marta Puig Giribets[1] | María Pilar García Guerreiro[1] | Mauro Santos[1] | Francisco J. Ayala[2] | Rosa Tarrío[1] | Francisco Rodríguez-Trelles[1]

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

[2]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA

Correspondence
Francisco Rodríguez-Trelles, Grup de Biologia Evolutiva (GBE), Departament de Genètica i de Microbiologia, Universitat Autonòma de Barcelona, Bellaterra, Barcelona, Spain.
Email: franciscojose.rodrigueztrelles@uab.cat

Funding information
Spanish Ministerio de Ciencia e Innovación, Grant/Award Number: CGL2013-42432P, CGL2017-89160P; Generalitat de Catalunya to the Grup de Genòmica, Bioinformàtica i Biologia Evolutiva, Grant/Award Number: 2009SGR 636, 2017SGR 1379; Universitat Autònoma de Barcelona (Spain)

Abstract

Heat-shock (HS) assays to understand the connection between standing inversion variation and evolutionary response to climate change in *Drosophila 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 *Drosophila madeirensis* and *Drosophila 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, 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.

KEYWORDS
chromosomal inversion polymorphism, climate change, concerted evolution, *Drosophila subobscura* subgroup, evolution on islands, *Hsp70*

1 | INTRODUCTION

Chromosomal inversions are a type of genomic rearrangement that is ubiquitous in nature (Hoffmann, Sgrò, & Weeks, 2004; 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 favourable 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., Ayala et al., 2011; Coluzzi, Sabatini, Della Torre, Di Deco, & Petrarca, 2002; Dobzhansky, 1970; Kapun, Fabian, Goudet, & Flatt, 2016; Levitan & Etges, 2005; Umina, Weeks, Kearney, McKechnie, & Hoffmann, 2005), with those from *Drosophila 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 *Drosophila madeirensis* and *Drosophila 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, Alvarez, & Zapata, 1996) and long-term directional trends (Balanyà, Oller, Huey, Gilchrist, & Serra, 2006; Rodríguez-Trelles & Rodríguez, 1998, 2010), which are overall consistent with expectations assuming temperature is the causative agent (Rezende et al., 2010; Rodríguez-Trelles & Rodríguez, 1998, 2010). Twenty years ago, the inversion polymorphisms of the species were found to be evolving in association with contemporary climate warming (Balanyà et al., 2006; Bradshaw & Holzapfel, 2006; Hughes, 2000, IPCC, 2001; Parmesan & Yohe, 2003; Rodríguez-Trelles & Rodríguez, 1998, 2007). Furthermore, the standing inversion variation maintained by the spatiotemporally fluctuating thermal environment enabled a rapid genomewide evolutionary response of the species to a recent heat wave (Grant et al., 2017; Rodríguez-Trelles, Tarrío, & Santos, 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 $O_{\underline{3+4}}$ are particularly interesting. They exhibit antagonistic spatiotemporal patterns, with cold-climate O_{ST} increasing in frequency with latitude and during the winter, and warm-climate $O_{\underline{3+4}}$ 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 $O_{\underline{3+4}}$ by superposition of inversion 4 on O_3 (Figure 1). The origin of $O_{\underline{3+4}}$ was followed by several independent inversions which gave rise to the $O_{\underline{3+4}}$ phylad, including, among others, the relatively rare $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+16}}$ arrangements that show less clear spatiotemporal patterns. Nucleotide variation analyses found that O_{ST} and $O_{\underline{3+4}}$ segregate as linked blocks of loci that are recombinationally isolated from each other (Munté, Rozas, Aguadé, & Segarra, 2005).

Drosophila madeirensis and *D. guanche* originated allopatrically from continental Palearctic O_3 ancestors (González, Cabrera, Larruga, & Gullón, 1983; Khadem, Munté, Camacho, Aguadé, & Segarra, 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, Segarra, Rozas, & Aguadé, 1998). *Drosophila madeirensis* remained monomorphic for O_3, whereas *D. guanche* became fixed for the Canary endemic $O_{\underline{3+8}}$. Currently, both species coexist with *D. subobscura* in their respective islands owing to independent secondary contacts after continental *D. subobscura* propagules nearly fixed for $O_{\underline{3+4}}$ 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, Matos, & Santos, 2006) which are capable of limited gene exchange in collinear genomic regions not affected by inversions (Herrig, Modrick, Brud, & Llopart, 2014). 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 $O_{\underline{3+4}}$ 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 $O_{\underline{3+4}}$ (Cuenca, Galindo, Saura, Sorsa, & de Frutos, 1998; Moltó, Pascual, Martínez-Sebastián, & de Frutos, 1992). $O_{\underline{3+4}}$ 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 $O_{\underline{3+4}}$ 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, Haney, & Feder, 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, Mortin, & Wu, 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, Xiao, & Lis, 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

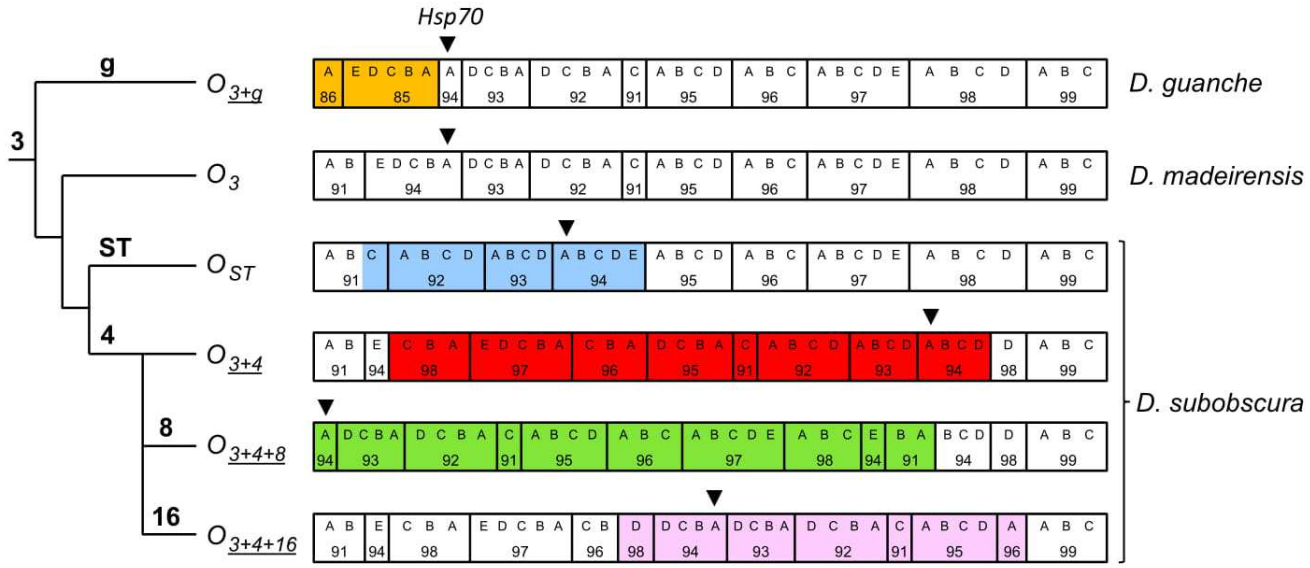


FIGURE 1 Phylogeny of O chromosomal inversions in the *Drosophila subobscura* subgroup. The cladogram on the left depicts the evolutionary history of chromosomal rearrangement on the standard cytological map of Kunze-Mühl and Müller (1958) shown on the right. The hypothetical ancestral gene order is O_3, on which inversion *g* (orange) arose in *Drosophila 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. *Drosophila madeirensis* is currently monomorphic for O_3. Arrows indicate the location of the *Hsp70* locus (subsection 94A) on the chromosomal rearrangement background

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, Mizukami, Imazu, & Sakurai, 2006). Gap-type HSEs consist of two (or more) inverted units separated by a 5 bp gap (e.g., nTTCnnGAAn[5 bp]nGAAn), and step-type HSEs consist of direct repeats of nGAAn or nTTCn units separated by 5 bp (e.g., nGAAn[5 bp]nGAAn[5 bp]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, Wilkins, Giardina, & Lis, 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.

2 | MATERIALS AND METHODS

2.1 | *Drosophila* lines

We used five strains from *Drosophila subobscura* plus one from each of *Drosophila madeirensis* and *Drosophila 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, Feuerbach-Mravlag, Lange, Michaelidis, & Pentzos-Daponte, 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 laboratory. All assayed lines were stored frozen at −20°C immediately upon obtention.

Drosophila 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: 14 May, 14 November and 7 July 2012, respectively; O_{3+4+8}, Vélez de Benaudalla, 36°50′23.66″N, 3°30′59.32″W, 16–17 April 2014; and O_{3+4+16}, Jérez del Marquesado, 37°11′6.74″N, 3°10′30.37″W, 29–30 August 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.

2.2 | In situ hybridization

Polytene chromosome preparations were performed following Labrador, Naveira, and Fontdevila (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 laboratory stock of the *ch-cu* strain, using available *Hsp70 cds* information from *Drosophila 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 × 300 MP10 assembly). One of the hit contigs included a complete *Hsp70 cds* spanning 1.9 kb, whose termini were used for nondegenerate 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 3 kb pGEM-T vector (Promega, Madison, WI, USA). Probes were labelled by random priming with digoxigenin (DIG-11dUTP) of purified PCR amplicons. Posthybridization washes were carried out 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).

2.3 | *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 × 300 MP10 assembly. The contig sequence data were 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 × 300 MP10 assembly.

2.4 | 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, Francino, Fontdevila, & Cabré, 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, USA). Phage P2-infected *Escherichia coli* were plated on NZYM culture medium (Amresco, Solon, OH, USA). Plaque lifts were carried out onto a positively charged Biodyne B Nylon membrane (Pall Corporation, Pensacola, FL, USA). 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 phosphatase, and the label developed with chromogenic alkaline phosphatase substrate NBT/BCIP (Roche, Indianapolis, IN, USA). 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.

2.5 | DNA isolation, PCR amplification, sequencing and annotation

gDNAs were isolated from 5 to 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. PCRs were performed using DFS-Taq DNA polymerase (Bioron, Ludwigshafen, Germany) on a MJ Research PTC-100 thermal cycler (MJ Research Inc., Watertown, MA, USA). 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 3 kb pGEM-T vectors (Promega, Madison, WI, USA) 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, 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.

2.6 | Multiple sequence alignment, recombination scans and phylogenetic inference

Multiple sequence alignment of the seven *Hsp70* sequences (hereafter referred to as seven OTUs data set) was conducted using the progressive guide tree-based MAFFT algorithm (version 7; http://mafft.cbrc.jp/alignment/software/) with the accuracy-oriented method "L-INS-i" (Katoh, Kuma, Toh, & Miyata, 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, Ashkenazy, Katoh, & Pupko, 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 version 1.28 alignment editor (Martin, Murrell, Golden, Khoosal, & Muhire, 2015) and checked for presence of stop codons and/or editing problems using MEGA (version 7.0.26) (Kumar, Stecher, & Tamura, 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, Rodríguez-Trelles, & Ayala, 2001). Prior to modelling 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 method implemented in the online version of IQ-Tree (W-IQ-Tree; http://iqtree.cibiv.univie.ac.at/; Trifinopoulos, Nguyen, von Haeseler, & Minh, 2016). Chi-square tests were applied both to unpartitioned and to 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 nonreticulating evolutionary processes (Sun, Evans, & Golding, 2011). The power to detect recombination in our MSA is limited because of the sparseness of intraclass, chromosomal rearrangement/species, sequence representation. On the other hand, we do not expect interclass 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 (Padidam, Sawyer, & Fauquet, 1999; Sawyer, 1989) and MaxChi (Smith, 1992; Posada & Crandall, 2001), and the phylogenetic method RDP (Martin & Rybicki, 2000) implemented in RDP4 version 4.94 with default settings (Martin et al., 2015). In addition, we used the single breakpoint phylogenetic method SBP (Kosakovsky Pond, Posada, Gravenor, Woelk, & Frost, 2006) implemented in the Datamonkey webserver (http://http://www.datamonkey.org/; Delport, Poon, Frost, & Kosakovsky Pond, 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). 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, Minh, Wong, von Haeseler, & Jermiin, 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 neighbour interchange with default perturbation strength and a stopping rule settings. Branch support was assessed using the ultrafast bootstrap approximation (UFboot; 1,000 replicates) (Minh, Nguyen, & von Haeseler, 2013), and two single-branch tests including the Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT; 1,000 replicates) (Guindon et al., 2010) and the approximate Bayes parametric test (Anisimova, Gil, Dufayard, Dessimoz, & Gascuel, 2011).

3 | RESULTS

3.1 | Localization and genomic sequence organization of the *Hsp70* genes in the subobscura subgroup

In situ hybridization of a *Drosophila 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 and Müller (1958) standard cytological map (Figure S1). 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.

Multiple sequence alignment 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 1,055 ± 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 1,929 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} (2,470 and 2,485 bp, for O_{ST1} and O_{ST2}, respectively), and longest in *Drosophila guanche* O_{3+g} (2,841 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 *Drosophila pseudoobscura* and *Drosophila 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

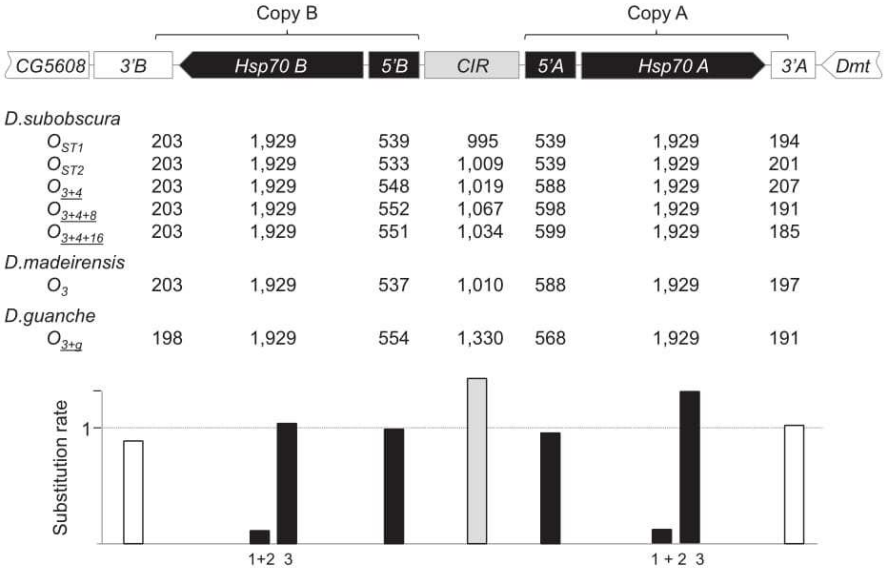


FIGURE 2 Genomic organization and rates of nucleotide substitution in the *Hsp70* locus. *Upper panel*: the boxes represent the different functional elements, including coding (*Hsp70A* and *Hsp70B*, and their upstream *CG5608* and downstream *Dmt* flanking genes) and noncoding (3′ and 5′ flanking regions 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 version 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

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, Subramanian, & Kumar, 2004). In addition, the block is partially conserved in the *melanogaster* group, which retains the pair CG5608-*Hsp70Ba*, and in *Drosophila 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).

3.2 | Molecular evolution and genealogy of the orthologous *Hsp70* genes in the subobscura species subgroup

After MSA, the final alignment matrix consisted of seven taxa and 7,042 characters of which 159 were parsimony-informative. Preliminary exploratory analysis using a chi-square approximation indicated that all sequences conform to the average character composition of their MSAs both from the unpartitioned data set 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 modelling of the substitution process. Unpartitioned and partitioned models, with a priori partitioning and best merging of a priori partitioning schemes, for

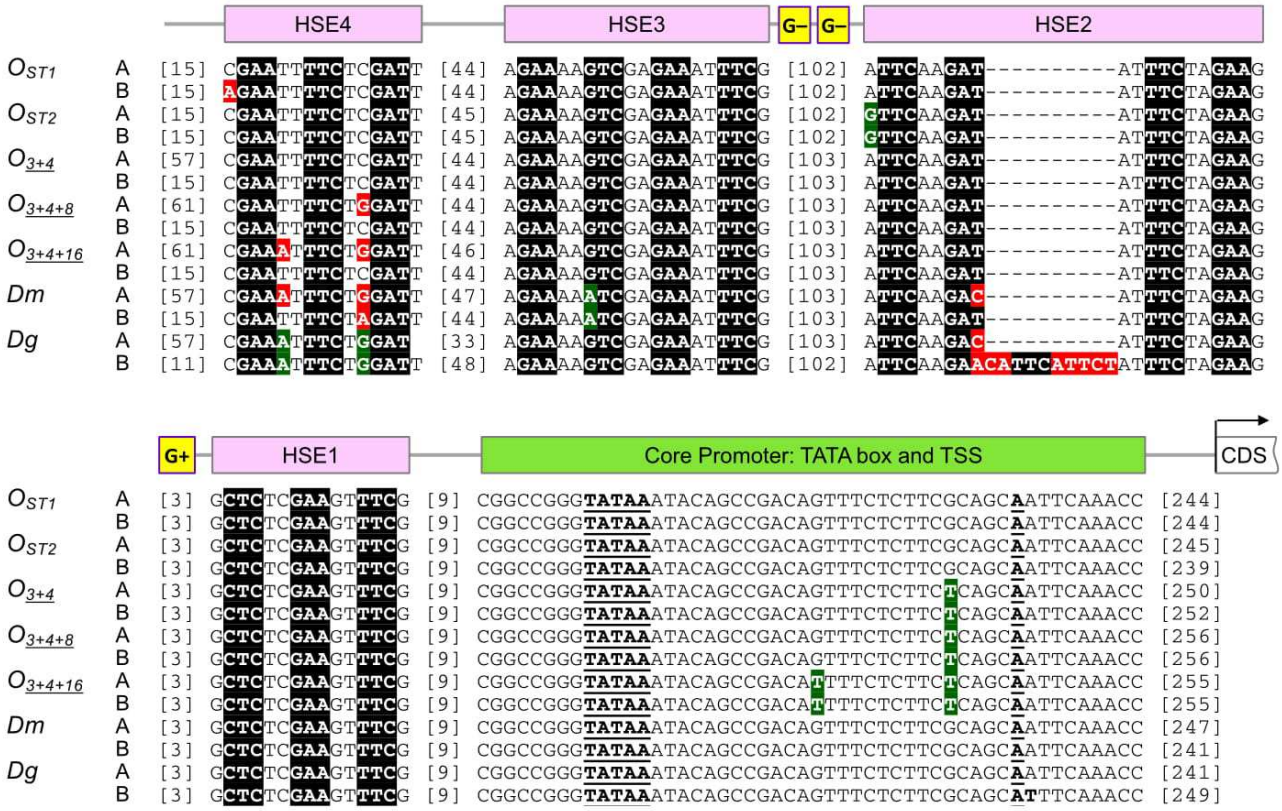


FIGURE 3 Schematic representation of a multiple sequence alignment (MSA) of the 5′ cis-regulatory region of the *Hsp70* ortholog and paralog genes from the *subobscura* subgroup. Boxes on the top of the MSA denote different cis-regulatory elements, including four HS elements (HSEs) (1–4; pink), GAGA sites (G+: GAGAGAG, G−: CTCTCTC; yellow) and the core promoter (green) with underlined TATA-box and transcription start site. Coloured 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)

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). The best BIC model is a ModelFinder best merging of an eight a priori partitions scheme of mixed noncoding nucleotide (five partitions, 2,932 characters in total) and amino acid (three partitions, 1,369 characters) 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 $O_{\underline{3+4}}$ 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 $O_{\underline{3+4}}$ 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 and S1), using BASEML from the PAML version 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.

TABLE 1 Optimal model selection for the evolution of the *Hsp70IR* nucleotide region (NUC) in the *Drosophila subobscura* species subgroup

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

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).

3.3 | 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 subgroup (<3 Mya; Herrig et al., 2014). 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 intrachromosomal gene conversion in species segregating for chromosomal arrangements, because of the interchromosomal 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

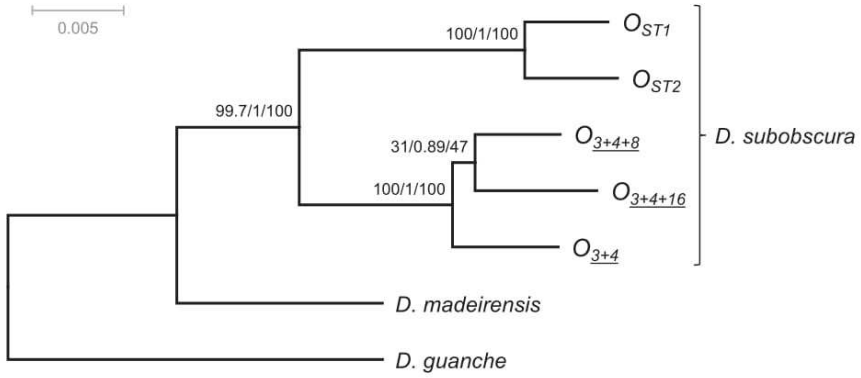


FIGURE 4 Unrooted ML tree of the phylogenetic relationships among species and chromosomal inversions of the *subobscura* subgroup based on the best Bayesian information criterion model from Table 1 (see also Table S1). The model allows for two partitions to accommodate differences in the substitution 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 1,000 replicates

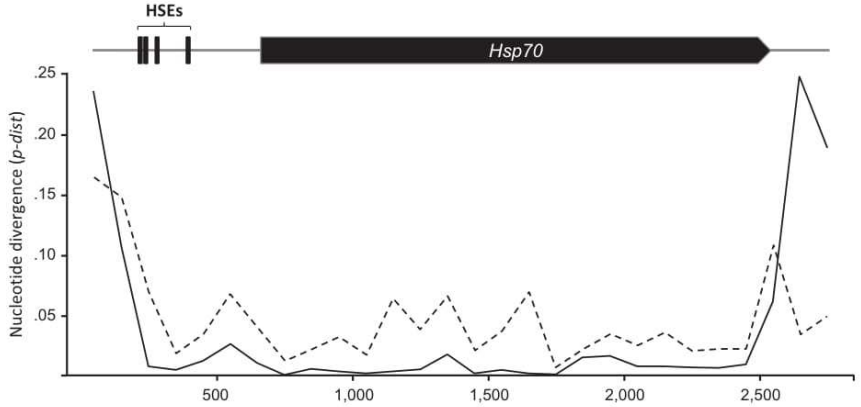


FIGURE 5 Spatial distribution patterns of the average pairwise proportion of nucleotide differences between duplicated blocks from the same chromosome (continuous line) and between orthologs (*Drosophila guanche* vs. the rest; dashed line) obtained using a nonoverlapping window analysis of size 100 bp.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

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; $p = .034$ and $p = .005$, for O_{ST1} and O_{ST2} vs. 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; $p < .001$), 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 vs. 0.0006 ± 0.0005; for *D. guanche* vs. the average across the remainder sequences; two-tailed *z*-test; $p < .000$). 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 ($-2\log\Lambda = 0.22$, $df = 1$, $p < .647$; $-2\log\Lambda = 11.30$, $df = 1$, $p = .001$; $-2\log\Lambda = 4.32$, $df = 1$, $p = .038$; $-2\log\Lambda = 0.26$, $df = 1$, $p = .610$ for the 5′ regulatory region, the third codon positions, the fourfold degenerate sites, and the amino acid sequence, respectively; conducted using PAML version 4.9d's BASEML and CODEML programs; Yang, 2007). Associated with 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, Qun, & Xia, 2013); two-tailed *z*-test; $p = .014$]. The increased Ka and Ks values, together with the two paralogs showing similar levels of codon usage deoptimization, suggest a relaxation of natural selection on *D. guanche*'s *Hsp70*. Still, the asymmetry 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* subgroup 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.

4 | DISCUSSION

4.1 | 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 (Osada & Innan, 2008; Sugino & Innan, 2006). The pattern of concerted evolution, however, is strongly structured and idiosyncratic across 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 and 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

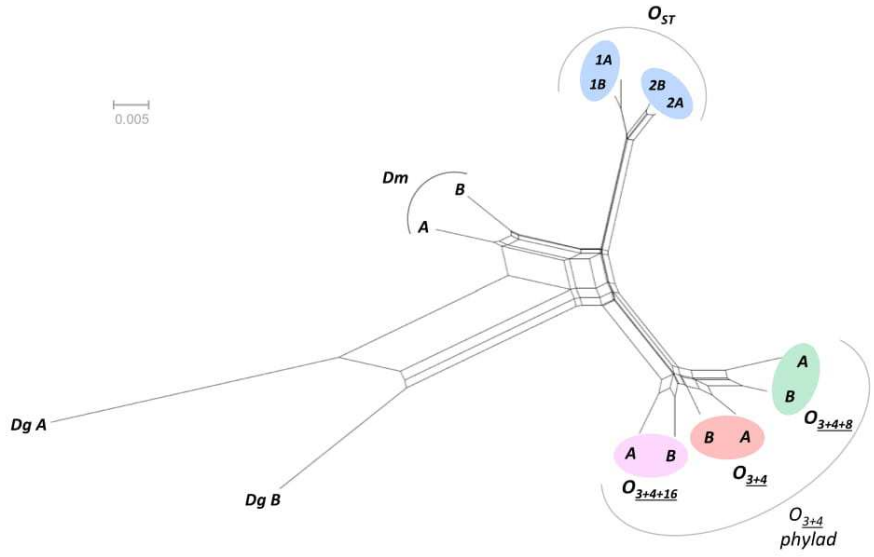


FIGURE 6 Phylogenetic network of *Hsp70* orthologous and paralogous sequences from the species and chromosomal arrangements of the *subobscura* subgroup. The split network was constructed using the NeighborNet method as implemented in SPLITSTREE version 4.14.5 (Huson & Bryant, 2006), on TN93 + I (% of invariable sites 85.1) model, which is the best-fit model from the unpartitioned data (Tables S2 and S3), 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

TABLE 2 Nei and 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* subgroup (O_{ST}, O_{3+4}, O_{3+4+8} and O_{3+4+16} gene arrangements are from *Drosophila subobscura*; Dm: *Drosophila madeirensis*; Dg: *Drosophila 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	—

in a head-to-head inverted repeat. A similar number and arrangement of genes are 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 fruit fly 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 (Garbuz & Evgen'ev, 2017; O'Brien et al., 1995; Wilkins & Lis, 1997), 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, Maine, & Schedl, 1985; Kuhn, Hart, & Geyer, 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, de Frutos, & Martínez-Sebastián, 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* (Evgen'ev, Garbuz, & Zatsepina, 2014; Lerman, Michalak, Helin, Bettencourt, & Feder, 2003; Zatsepina et al., 2001), 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, Hubley, & Green, 2013). We put particular attention on the promoter and 3′ regions, as they contain constitutively nucleosome free, open chromatin domains (Karpov, Preobrazhenskaya, & Mirzabekov, 1984; Petesch & Lis, 2008; Tsukiyama, Becker, & Wu, 1994), 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 was 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.

4.2 | 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). *Drosophila madeirensis* and *D. subobscura* are partially isolated from each other, but conclusions about the degree of the isolation are mixed (Herrig et al., 2014; Khadem & Krimbas, 1991; Krimbas & Loukas, 1984; Rego et al., 2006). 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; Khadem, Rozas, Segarra, Brehm, & Aguadé, 1998; Larruga, Cabrera, González, & Gullón, 1983).

Typically, interchromosomal recombination may occur via reciprocal crossing-over, which can be single or double, and/or nonreciprocal, so-called copy-and-paste gene conversion, which can occur either associated with 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, Betrán, Barbadilla, & Ruiz, 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% (Krimbas, 1992; Loukas, Krimbas, Mavragani-Tsipidou, & Kastritsis, 1979) of the total O chromosome length of 190.7 cM (Pegueroles, Araúz, Pascual, & Mestres, 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, Segarra, Ribó, & Aguadé, 1999); Acph-1 (Navarro-Sabaté, Aguadé, & Segarra, 1999); and Fmr1 (Pegueroles, Aquadro, Mestres, & Pascual, 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 immediately 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.

4.3 | Variation in the rate of concerted evolution between chromosomal inversions of *Drosophila 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 indicates 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 (Fawcett & Innan, 2011; Osada & Innan, 2008). 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.

4.4 | 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, Pascual, & de Frutos, 1987; Moltó et al., 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 favourable 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.

Drosophila 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., 2014). Currently, the species is endemic to the Tertiary relictual montane (600–1,200 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 counterclockwise 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 northeast (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, subhumid conditions provided by the influence of trade winds moisture during the geological time the islands showed a suitable altitudinal range (de Nascimento, Willis, Fernández-Palacios, Criado, & Whittaker, 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 towards 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, Munté, Rozas, Segarra, & Aguadé, 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, Nagel, Bachmann, & Bachmann, 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 2,500 years 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 (Balanyà et al., 2006; Rodríguez-Trelles & Rodríguez, 1998; Rodríguez-Trelles et al., 2013).

4.5 | 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 (Casola, Ganote, & Hahn, 2010; Ezawa, Oota, & Saitou, 2006). 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 Grants from Generalitat de Catalunya to the Grup de Genòmica, Bioinformàtica i Biologia Evolutiva (2009SGR 636 and 2017SGR 1379). 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 *Drosophila madeirensis* and Montserrat Peiró Navarro for help with the in situ hybridizations.

DATA ACCESSIBILITY

Nucleotide sequences are deposited in GenBank under Accession nos: 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., analysed the data. M.P.G., M.S., F.J.A., R.T. and F.R.-T. wrote the article.

ORCID

Francisco Rodríguez-Trelles http://orcid.org/0000-0001-6279-3630

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. https://doi.org/10.1093/sysbio/syr041

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. https://doi.org/10.1093/molbev/msq248

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. https://doi.org/10.1126/science.1131002

Bettencourt, B. R., & Feder, M. E. (2002). Rapid concerted evolution via gene conversion at the *Drosophila hsp70* genes. *Journal of Molecular Evolution*, *54*, 569–586. https://doi.org/10.1007/s00239-001-0044-7

Bradshaw, W. E., & Holzapfel, C. M. (2006). Evolutionary response to rapid climate change. *Science*, *312*, 1477–1478. https://doi.org/10.1126/science.1127000

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. https://doi.org/10.1111/j.1420-9101.2012.02463.x

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. https://doi.org/10.1002/jez.b.22647

Casola, C., Ganote, C. L., & Hahn, M. W. (2010). Nonallelic gene conversion in the genus *Drosophila*. *Genetics*, *185*, 95–103. https://doi.org/10.1534/genetics.110.115444

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

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. https://doi.org/10.1126/science.1077769

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. https://doi.org/10.1007/s004120050287

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. https://doi.org/10.1093/bioinformatics/btq429

Deng, W., Maust, B. S., Nickle, D. C., Learn, G. H., Liu, Y., Heath, L., … Mullins, J. I. (2010). DIVEIN: A web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *BioTechniques*, *48*, 405–408. https://doi.org/10.2144/000113370

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, NY: 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.). Berlin, Germany, New York, NY, London, UK: Springer.

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. https://doi.org/10.1093/molbev/msj093

Fawcett, J. A., & Innan, H. (2011). Neutral and non-neutral evolution of duplicated genes with gene conversion. *Genes*, *2*, 191–209. https://doi.org/10.3390/genes2010191

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. https://doi.org/10.1111/j.1365-2699.2010.02427.x

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. https://doi.org/10.1134/S1022795417010069

García Guerreiro, M. P., & Fontdevila, A. (2007). The evolutionary history of *Drosophila buzzatii*. VI. 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. https://doi.org/10.1111/j.1558-5646.1983.tb00227.x

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. https://doi.org/10.1098/rstb.2016.0146

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. https://doi.org/10.1093/sysbio/syq010

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. https://doi.org/10.1074/jbc.M510827200

Herrig, D. K., Modrick, A. J., Brud, E., & Llopart, A. (2014). Introgression in the *Drosophila subobscura—D. madeirensis* sister species: Evidence of gene flow in nuclear genes despite mitochondrial differentiation. *Evolution*, *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. https://doi.org/10.1186/1471-2148-13-68

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. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

Hoffmann, A. A., Sgrò, C. M., & Weeks, A. R. (2004). Chromosomal inversion polymorphisms and adaptation. *Trends in Ecology and Evolution*, *19*, 482–488. https://doi.org/10.1016/j.tree.2004.06.013

Hughes, L. (2000). Biological consequences of global warming: Is the signal already apparent? *Trends in Ecology and Evolution*, *15*, 56–61. https://doi.org/10.1016/S0169-5347(99)01764-4

Huson, D. H., & Bryant, I. D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*, 254–267. https://doi.org/10.1093/molbev/msj030

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, UK: Cambridge University Press.

Jedlicka, P., Mortin, M. A., & Wu, C. (1997). Multiple functions of *Drosophila* heat shock transcription factor in vivo. *EMBO Journal*, *16*, 2452–2462. https://doi.org/10.1093/emboj/16.9.2452

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. https://doi.org/10.1159/000084979

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. https://doi.org/10.1038/nmeth.4285

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. https://doi.org/10.1093/molbev/msw016

Karpov, V. L., Preobrazhenskaya, O. V., & Mirzabekov, A. D. (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. https://doi.org/10.1016/0092-8674(84)90235-6

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. https://doi.org/10.1093/nar/gki198

Kent, W. J. (2002). BLAT – The BLAST-like alignment tool. *Genome Research*, *12*, 656–664. https://doi.org/10.1101/gr.229202

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. https://doi.org/10.1038/hdy.1991.75

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. https://doi.org/10.1111/j.1420-9101.2012.02467.x

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. https://doi.org/10.1007/s000360050098

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. https://doi.org/10.1534/genetics.105.047985

Korunes, K. L., & Noor, M. A. F. (2017). Gene conversion and linkage: Effects on genome evolution and speciation. *Molecular Ecology*, *26*, 351–364. https://doi.org/10.1111/mec.13736

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution*, *23*, 1891–1901. https://doi.org/10.1093/molbev/msl051

Krimbas, C. B. (1992). The inversion polymorphism of *Drosophila subobscura*. In C. B. Krimbas, & J. R. Powell (Eds.), *Drosophila inversion polymorphism* (pp. 128–220). Boca Raton, FL: CRC Press.

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. https://doi.org/10.1038/hdy.1984.109

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. https://doi.org/10.1093/molbev/msw054

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. https://doi.org/10.1093/oxfordjournals.jhered.a110934

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. https://doi.org/10.1007/BF00122374

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. https://doi.org/10.1093/molbev/msg015

Levitan, M., & Etges, W. J. (2005). Climate change and recent genetic flux in populations of *Drosophila robusta*. *BMC Evolutionary Biology*, *5*, 4. https://doi.org/10.1186/1471-2148-5-4

Liao, D. Q. (1999). Concerted evolution: Molecular mechanism and biological implications. *American Journal of Human Genetics*, *64*, 24–30. https://doi.org/10.1086/302221

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.

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. https://doi.org/10.1093/oxfordjournals.jhered.a109181

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. https://doi.org/10.1093/bioinformatics/16.6.562

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. https://doi.org/10.1046/j.1420-9101.1992.5040625.x

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. https://doi.org/10.1093/oxfordjournals.molbev.a026259

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*, 1188–1195. https://doi.org/10.1093/molbev/mst024

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. https://doi.org/10.1007/BF00330706

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. https://doi.org/10.1007/BF01945535

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. https://doi.org/10.1139/g92-132

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.

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. https://doi.org/10.1534/genetics.104.032748

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. https://doi.org/10.1111/j.1365-2699.2008.02012.x

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. https://doi.org/10.1093/molbev/msh041

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. https://doi.org/10.1101/gad.9.9.1098

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. https://doi.org/10.1006/viro.1999.0056

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421*, 37–42. https://doi.org/10.1038/nature01286

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. https://doi.org/10.1146/annurev.ge.27.120193.002253

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. https://doi.org/10.1038/hdy.2012.118

Pegueroles, C., Araúz, P. A., Pascual, M., & Mestres, F. (2010). A recombination survey using microsatellites: The O chromosome of *Drosophila subobscura*. *Journal of Molecular Evolution*, *138*, 795–804. https://doi.org/10.1007/s10709-010-9461-0

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. https://doi.org/10.1093/nar/gkq443

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. https://doi.org/10.1093/molbev/msg199

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 5 bp recognition unit. *Cell*, *59*, 797–806. https://doi.org/10.1016/0092-8674(89)90603-X

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.

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. https://doi.org/10.1073/pnas.241370698

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. https://doi.org/10.1006/mpev.1997.0438

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. https://doi.org/10.1111/j.1558-5646.2009.00835.x

Rego, C., Matos, M., & Santos, M. (2006). Symmetry breaking in interspecific *Drosophila* hybrids is not due to developmental noise. *Evolution*, *60*, 746–761. https://doi.org/10.1111/j.0014-3820.2006.tb01153.x

Rezende, E. L., Balanyà, J., Rodríguez-Trelles, F., Rego, C., Fragata, I., Matos, M., … Santos, M. (2010). Climate change and chromosomal inversions in *Drosophila subobscura*. *Climate Research*, *43*, 103–114. https://doi.org/10.3354/cr00869

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. https://doi.org/10.1023/A:1006546616462

Rodríguez-Trelles, F., & Rodríguez, M. A. (2007). Comment on 'Global genetic change tracks global climate warming in *Drosophila subobscura*'. *Science*, *315*, 1497. https://doi.org/10.1126/science.1136298

Rodríguez-Trelles, F., & Rodríguez, M. A. (2010). Measuring evolutionary responses to global warming: Cautionary lessons from *Drosophila*. *Insect Conservation and Diversity*, *3*, 44–50. https://doi.org/10.1111/j.1752-4598.2009.00071.x

Rodríguez-Trelles, F., Tarrío, R., & Santos, M. (2013). Genome-wide evolutionary response to a heat wave in *Drosophila*. *Biology Letters*, *9*, e20130228. https://doi.org/10.1098/rsbl.2013.0228

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.

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. https://doi.org/10.1038/hdy.2010.26

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 chromosomes with a DIG-labeled DNA probe. In S. Grünewald-Janho, J. Keesey, M. Leous, R. van Miltenburg & C. Schroeder (Eds.), *Non-radioactive in situ hybridization application manual* (3rd ed., pp. 108–111). Mannheim, Germany: Roche.

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. https://doi.org/10.1093/nar/gkv318

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. https://doi.org/10.1534/genetics.105.053959

Smit, A. F. A., Hubley, R., & Green, P. (2013). *RepeatMasker Open-4.0*, 2013–2016. Retrieved from http://www.repeatmasker.org

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

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. https://doi.org/10.1016/j.tig.2006.09.014

Sun, S., Evans, B. J., & Golding, G. B. (2011). "Patchy-tachy" leads to false positives for recombination. *Molecular Biology and Evolution*, *28*, 2549–2559. https://doi.org/10.1093/molbev/msr076

Sun, X., Qun, Y., & Xia, X. (2013). An improved implementation of effective number of codons (Nc). *Molecular Biology and Evolution*, *30*, 191–196. https://doi.org/10.1093/molbev/mss201

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. https://doi.org/10.1093/oxfordjournals.molbev.a003932

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. https://doi.org/10.1093/nar/gkw256

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. https://doi.org/10.1038/367525a0

Udvardy, A., Maine, E., & Schedl, P. (1985). The 87A7 chromomere. Identification 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. https://doi.org/10.1016/0022-2836(85)90408-5

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. https://doi.org/10.1126/science.1109523

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. https://doi.org/10.1093/nar/gkm306

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. https://doi.org/10.1093/nar/25.20.3963

Woolfit, M., & Bromham, L. (2005). Population size and molecular evolution on islands. *Proceedings of the Royal Society B-Biological Sciences*, *272*, 2277–2282. https://doi.org/10.1098/rspb.2005.3217

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088

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.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Puig Giribets M, García Guerreiro MP, Santos M, Ayala FJ, Tarrío R, Rodríguez-Trelles F. Chromosomal inversions promote genomic islands of concerted evolution of *Hsp70* genes in the *Drosophila subobscura* species subgroup. *Mol Ecol*. 2018;00:1–17. https://doi.org/10.1111/mec.14511

3.2 Quantification of basal mRNA and protein levels in adult *D. subobscura* males and females of the O_{ST} and $O_{\underline{3+4}}$ adaptive arrangements.

Chapter 3.2 consists of a research article **"**Quantification of *hsp70* mRNA and Hsp70 protein levels under nonstressful conditions in the warm climate-associated $O_{\underline{3+4}}$ and the cold climate-associated O_{ST} arrangements of *Drosophila subobscura.*" that is currently in preparation. The Supplementary Material of this article is available in Appendix 7.2.

3.2.1 Quantification of *hsp70* mRNA and Hsp70 protein levels under non heat-shock conditions in the warm climate-associated $O_{\underline{3+4}}$ and the cold climate-associated O_{ST} arrangements of *Drosophila subobscura.*

3.2.1.1 Abstract

Drosophila subobscura exhibits a rich inversion polymorphism, with some adaptive inversions that co-vary with temperature and latitude. The *hsp70* gene family codes for proteins functioning under conditions of cellular stress and/or high temperatures that help other proteins fold properly, and constitutes a good candidate to explain the adaptive inversion polymorphism in this species. A recent study of Hsp70 protein quantification in homokaryotypical $O_{\underline{3+4}}$, $O_{\underline{3+4+8}}$ and O_{ST} strains showed increased basal Hsp70 protein levels in the warm climate-associated $O_{\underline{3+4}}$ strain that did not boost with temperature. To obtain a more global picture of *hsp70* gene expression patterns and their relationship with adaptive inversion polymorphism, we quantified male and female basal *hsp70* mRNA and protein levels from several cold climate-associated O_{ST} and warm climate-associated $O_{\underline{3+4}}$ isogenic lines. Contrary to previous results, no differences in basal *hsp70* mRNA or Hsp70 protein amounts were observed between $O_{\underline{3+4}}$ and O_{ST} flies when the two sexes were analyzed together. These outcomes are reinforced by the high conservation of several *cis*-regulatory elements that we identified in the *hsp70* proximal promoter regions of all the O_{ST} and $O_{\underline{3+4}}$ lines analyzed in this work. Different results are obtained when the two sexes are analyzed separately, where $O_{\underline{3+4}}$ males show significantly decreased mRNA levels relative to $O_{\underline{3+4}}$ females, while an interaction effect between arrangement and sex is observed in the protein data set. When put together, our results indicate that *hsp70* gene expression might be influenced by sex in *D. subobscura*.

Keywords: *hsp70* expression; inversion polymorphism; heat shock promoters; *Drosophila subobscura*.

3.2.1.2 Introduction

Drosophila subobscura is a temperate species native to the Palearctic region that exhibits a rich inversion polymorphism across its five acrocentric chromosomes (A, E, O, U, J) (Menozzi and Krimbas 1992). Several studies in different populations indicate that certain inversion frequencies vary clinally and are correlated to variables like temperature and latitude (Menozzi and Krimbas 1992, Balanyà et al. 2003, Balanyà et al. 2009). Interestingly, it has been noted that similar clinal patterns have independently arisen, for the same inversion, in the colonizer populations of North and South America that corroborates their adaptive character (Ayala et al. 1989; Balanyà et al. 2003; Castañeda et al. 2013). Among heat shock proteins (HSPs) the highly conserved *hsp70* gene family could be a candidate locus to explain adaptive inversion polymorphisms. Actually, Hsp70 proteins are involved in protein folding and assembly, and can be induced by several stressors as a sudden increase in temperature (Lindquist 1986; Lindquist and Craig 1988). Some studies have found a positive link between *hsp70* copy number, expression levels after heat stress and thermal resistance in artificial (Feder et al. 1996; Bettencourt et al. 2008) and natural (Evgen'ev et al. 2004) *Drosophila* strains. In contrast, a trend towards decreased heat-inducible *hsp70* (mRNA) and Hsp70 (protein) expression levels have evolved in some *Drosophila* strains perpetually exposed to high temperatures, probably to avoid the deleterious effects of constitutively expressed Hsp70 for the cell at normal temperatures (Bettencourt et al. 1999; Sørensen et al. 1999; Sørensen et al. 2001; Zatsepina et al. 2001).

This gene family has recently been studied in *D. subobscura* in an experiment where homokaryotypical individuals for the *D. subobscura* arrangements O_{ST}, $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ were heat shocked at 32ºC and the Hsp70 protein levels measured before and 1h after the heat shock (Calabria et al. 2012). The results indicated that the warm climate-associated $O_{\underline{3+4}}$ flies had higher basal Hsp70 than those submitted to a heat shock. The Hsp70 response of $O_{\underline{3+4}}$ to thermal stress is atypical of species with short life cycles as *Drosophila*, where increased basal Hsp70 levels are particularly harmful for development and have been associated with high larval mortality (Feder et al. 1992; Krebs and Feder 1997b; Garbuz and Evgen'ev 2017). Hence, the Hsp70 inducible profile of $O_{\underline{3+4}}$ flies would resemble that of certain thermotolerant species with long life cycles such as *Tabanidae* and *Stratiomyidae*. Moreover, former *in-situ* hybridization

experiments (ISH) and the sequencing of a ~10 kb region in the *Hsp70IR* locus of several *D. subobscura* isogenic lines for four representative arrangements in the SI segment of the O chromosome (O_{ST}, $O_{\underline{3+4+16}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4}}$) showed no differences in *hsp70* copy number between them (Puig-Giribets et al. 2018). In the same study, a high degree of conservation between the four arrangements was noted after comparing their 5' *cis*-regulatory regions of the two *hsp70* copies, indicating that their gene expression regulation might be similar.

Our study is focused in determining the amplitude of above mentioned work and understand the mechanisms involved. In order to disentangle the molecular mechanisms responsible for the high basal levels of Hsp70 observed in adult individuals wearing the $O_{\underline{3+4}}$ arrangement, we studied both mRNA and protein levels, at basal conditions (18ºC±1ºC), together with promoter sequences in several isogenic lines for the cold climate-associated O_{ST} and the warm climate-associated $O_{\underline{3+4}}$ arrangements. Amongst the most relevant results of this study, we observed comparable levels of mRNA between the two arrangements studied, being significantly higher in females carrying the $O_{\underline{3+4}}$ arrangement than in males. Protein and mRNA amounts are not correlated, reaching comparable levels in both sexes and in chromosomal arrangements studied with an interaction effect between mRNA and protein. Thus, our results do not agree with those of Calabria et al. (2012) in the sense that we did not find increased basal Hsp70 levels in the flies wearing the warm climate-associated $O_{\underline{3+4}}$ arrangement in the I segment and that *hsp70* mRNA and protein levels appear to be influenced by sex. At the same time, the high degree of conservation in the proximal promoter regions of the two *hsp70* genes indicates that gene expression regulation might be similar in O_{ST} and $O_{\underline{3+4}}$ lines.

3.2.1.3 Materials and Methods

***Drosophila subobscura* lines:** All *D. subobscura* lines used in this work have been obtained and provided by Drs. R. Tarrío and F. Rodríguez-Trelles. Briefly, the O chromosomes of 12 *D. subobscura* lines (six for O_{ST} and six for $O_{\underline{3+4}}$) were made isogenic for two inversions that occur in the same chromosomal segment where the *Hsp70IR* locus has been described (Moltó et al. 1992; Krimbas 1993). For O chromosome isogenisation, wild males during 2012 in Berbikiz (Gordexola, Basque Country, N. Spain; 43°11′20.31″ N, 3°5′23.74″ W) were crossed to virgin females of the

Va/Ba balanced lethal stock as described in Sperlich et al. (1977). All stocks were reared at 18ºC±1ºC on standard *Drosophila* media. To avoid high larval crowding conditions that have previously shown to trigger *hsp70* stress response (Sørensen and Loeschcke 2001), we placed a maximum of 45 eggs in a vial containing 6 mL of food as described in Dolgova et al. (2010). For each line studied, two or three biological replicates of three 7-day-old adults per sex were used for RNA analyses. Another set of three equivalent replicates per line was used for protein analyses.

RNA expression quantification: RNA was extracted from 70 samples of one-week old adult adult flies stored at -80ºC (2 sexes x 2-3 biological replicates per sample x 12 isogenic lines; Supplementary Table 1). RNA extraction was performed using the Qiazol Lysis reagent as described in Qiazol handbook (Qiagen). Isolated RNA was treated with 1 µl (2U/µl) of DNase I (Ambion) in 20 µl of reaction mixture during 4h at 37ºC. Concentration of DNA-free samples was quantified (NanoDrop-2000, Thermo Scientific) and adjusted to ~1 µg/µL before cDNA synthesis with the Transcriptor First Strand cDNA Synthesis kit (Roche). Gene expression was quantified by measuring fluorescence intensity using iQ SYBR Green Supermix (BioRad, Hercules CA, USA) on a CFX96 BioRad Real-Time LightCycler at 59ºC using *hsp70* specific primers. All assays were performed in three technical replicates and the *rp49* houseskeeping gene (expressed equally in all samples) used as endogenous control. Relative expression was then calculated with the comparative C_T method (Schmittgen and Livak 2008; Supplementary Table 1). In each of the 12 lines we amplified a conserved *hsp70* coding fragment of 226 bp, common to both *hsp70* copies described in this species (Puig-Giribets et al. 2018). The primers used for amplification were: HSP70_1L (5'CACAGTCTTTGACGCCAAGC3') and HSP70_1R (5'TGTGATGCTTTGGCCCAGAT3'). For *rp49* housekeeping the primers RP49F (5'ACATCGGTTATGGCTCCAC3') and RP49R (5'GATTTCCTTGCGCTTCTTTG3' were designed from a *D. subobscura* GenBank sequence (Accession: AJ228921) amplifying a 164 bp segment from the gene second exon. Primer efficiencies of *hsp70* and *rp49*, calculated on a standard curve slope using serially diluted templates, were 98% and 99.6% respectively.

Hsp70 protein quantification by ELISA (enzyme-linked immunosorbent assay): Hsp70 protein quantification was carried out by ELISA in 72 individual samples

equivalent to those used in mRNA quantification (Supplementary Table 2). A Hsp70-specific monoclonal primary antibody (clone 5A5, dilution 1:1000 PBS; ThermoScientific) and a HRP-conjugated secondary antibody (anti-mouse IgG; ThermoScientific) were used following the ELISA protocol described in Sørensen et al. (1999). Color reaction was measured with a spectophotometric microplate reader (PowerWave XS2, Biotek) at wavelength 490 nm. To avoid plate-specific effects, we placed randomly 12 samples in 6 ELISA plates with three replicates and two blanks per sample. The Hsp70 absorbance values per plate were obtained by substraction of the blank mean value from the Hsp70 sample means. To normalize Hsp70 levels between different ELISA plates, one plate mean was chosen randomly as reference to calculate the correction factor for the rest.

Characterization of CREs in the proximal promoter region: A genomic region expanding ~300 bp upstream the transcription start site (TSS), was sequenced in the *hsp70A* and *hsp70B* genes of our 12 isogenic lines as described in Puig-Giribets et al. (2018). The recently published sequences the closely related *D. madeirensis* and *D. guanche* are described in Puig-Giribets *et al.* (2018). To identify several CREs that are typically found in the proximal 5' flanking region, the sequences of the two *hsp70* genes were first separately aligned using the MAFFT E-INS-i algorithm (Katoh and Standley 2013; available at: https://mafft.cbrc.jp/alignment/server/). Next, we used the information available on the *hsp70* proximal promoters of *D. subobscura* (Puig-Giribets et al. 2018) and other *Drosophila* species (Tian et al. 2010) to identify four conserved heat shock elements (HSEs) in the sequences of our 12 lines (HSEs 1-4). Eukaryotic Neural Network Promoter Prediction software (http://www.fruitfly.org/seq_tools/promoter.html) was used to find other CREs such as GAGA binding sites and some core promoter elements, including the transcription start site (TSS) and the TATA box.

Statistical analyses: The software SPSS Version 22.0 (IMB SPSS Inc., NY, US) was used for the statistical analyses of protein and mRNA data. Parametric tests were used only when the prerequisites of normality and equal variances between groups were met. For mRNA analyses, the nonparametric Kruskal-Wallis rank sum test (Kruskal and Wallis 1952) was chosen to detect significant differences in *hsp70* relative expression levels between four groups represented by all O_{ST} males (n = 17), O_{ST} females (n = 17),

$O_{\underline{3+4}}$ males (n = 18) and $O_{\underline{3+4}}$ females (n = 18). Significant differences between some groups were found at P = 0.010 (data not shown). Subsequently, the nonparametric Wilcoxon rank-sum test (Wilcoxon 1992) was used to test for differences in pairwise comparisons between sexes and arrangements. A two-way nested ANOVA test was performed to detect significant differences between sexes and arrangements and/or interaction effects between the two variables on Hsp70 protein levels. The test was done using the complete dataset (n = 72). The Spearman rank-order correlation test was used to calculate the correlation coefficient between *hsp70* mRNA relative expression and protein levels in males and females.

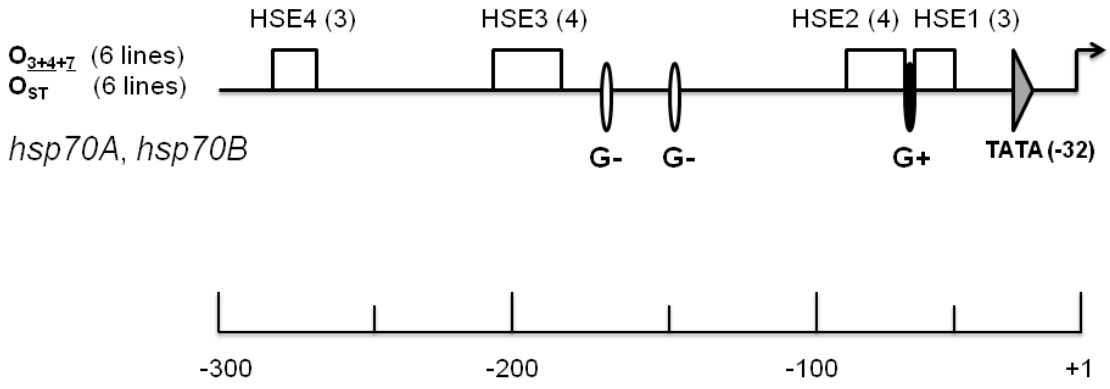
3.2.1.4. Results and Discussion

High degree of conservation in the *hsp70* proximal promoter region between O_{ST} and $O_{\underline{3+4}}$.

Multiple sequence alignments of a ~300 bp genomic region upstream of the transcription initiation site (TSS) in the two *hsp* genes (*hsp70A* and *hsp70B*) indicate high CRE conservation between our O_{ST} and $O_{\underline{3+4}}$ lines (Supplementary Figures 1 and 2). As shown in Puig-Giribets et al. (2018), four highly conserved HSEs were identified in all *hsp70* homologs and show no fixed nucleotide differences between the two aforementioned arrangements (Figure 1). Each HSE consists of at least three continuous alternate nGAAn/nTTCn pentanucleotide subunits that represent binding sites for the heat shock factor (HSF) monomer, as reported in Tian et al. (2010). Although the number of repetitions is identical in the 12 *D. subobscura* lines, it varies between HSEs: three in the most distal and proximal (HSE1, HSE4) and four in HSE2 and HSE3 (Figure 1). Similarly, the genomic organization of HSEs is virtually identical in the two arrangements and resembles that of other *Drosophila* species (Tian et al. 2010), where the two most proximal (HSE1 and HSE2) are more tightly spaced relative to the two most distal ones (Figure 1). Within each HSE, polymorphisms appear to be overrepresented in the first and fifth position of nGAAn/nTTCn pentanucleotides, which are the least constrained within the unit (Tian et al. 2010). It is worth mentioning the presence of a few insertion/deletion (INDEL) polymorphisms among the 12 lines in the presumably unconstrained regions separating the more distal HSEs. These polymorphisms tend to be shared between the two *hsp70* paralogs of the same line as

result of ectopic gene conversion (Puig-Giribets et al. 2018). Concomitantly, we studied the number of binding sites for the GAGA (GAF) transcription factor, which participates in the regulation of global gene expression. GAGA factors bind to promoter sequences of numerous genes with a consensus (GAGAG) (Omichinski et al. 1997). The study of GAGA site distribution in the *hsp70* promoters of the 12 *D. subobscura* lines showed no differences in the total number of GAGA sites in the proximal promoters. In the same way, two core proximal promoter elements, including the TATA box (TATAAA) at position -32 relative to the TSS (AATT…), are identical in all *hsp70* homologous sequences examined in this work. Overall, our results indicate that the two *hsp70* genes (*hsp70A* and *hsp70B*) must be regulated by similar mechanisms in the adaptive O_{ST} and O_{3+4} gene arrangements, examined in this work, and in O_{3+4+8} and O_{3+4+16} examined by Puig-Giribets et al. (2018).

Figure 1. Schematic representation of the 5' *cis*-regulatory elements in the two copies of *hsp70* gene (A and B) from the O_{ST} and O_{3+4} lines. White rectangles: HSEs 1-4, where numbers in brackets represent the number of nGAAn or nTTCn pentanucleotide units. Black ellipses: GAGAG (G+) repeat motifs. White ellipses: CTCTC (G-) repeat motifs. Gray arrow: TATA box. The scale represents nucleotides relative to the transcription start site (+1), indicated with a forward arrow.



Decreased mRNA levels in males relative to females of the warm climate-associated O_{3+4} arrangement.

The Wilcoxon rank-sum test was used to detect significant differences in pairwise comparisons between sexes, between arrangements, between sexes of the same arrangement and between arrangements of the same sex (Figure 1 and Table 1). Significant differences among the two sexes (P = 0.008) but not between the two arrangements (P = 0.108) were observed. Considering that no differences are detected neither between the O_{ST} male and female nor between the O_{ST} and O_{3+4} female data sets, the variation observed between the two sexes is principally explained by reduced mRNA levels in O_{3+4} males relative to their female counterparts (P = 0.006). In our opinion, the observed differences between the two sexes in O_{3+4} lines indicate that *hsp70* mRNA levels are influenced by sex and points to possible epistatic effects between the O chromosome and the sexual chromosomes in *D. subobscura*. Although not much information is available on basal *hsp70* mRNA levels in *Drosophila* males versus females, it is worth mentioning that some studies have found gender-based differences on Hsp70 protein levels in *D. melanogaster* adults after heat stress (Dahlgaard et al. 1998; Folk et al. 2006). Interestingly, certain members of the Hsp70 protein family have shown to have an important role in mouse spermatogenesis (Govin et al. 2006). In a nutshell, our results indicate that some Hsp70 family members might participate in gender-specific pathways also in *D. subobscura.*

Figure 1. Box plots representing the *hsp70* mRNA basal expression in males and females from O_{3+4} and O_{ST} arrangements. Males are represented in turquoise and females in pink. Mild outliers (1.5 to 3.5 fold the interquartile range) are represented by a cercle and a asterisk). Triangles represent extreme outliers (above 3-fold the interquartile range) that could not be represented in the same scale. *: *P* <0.05.

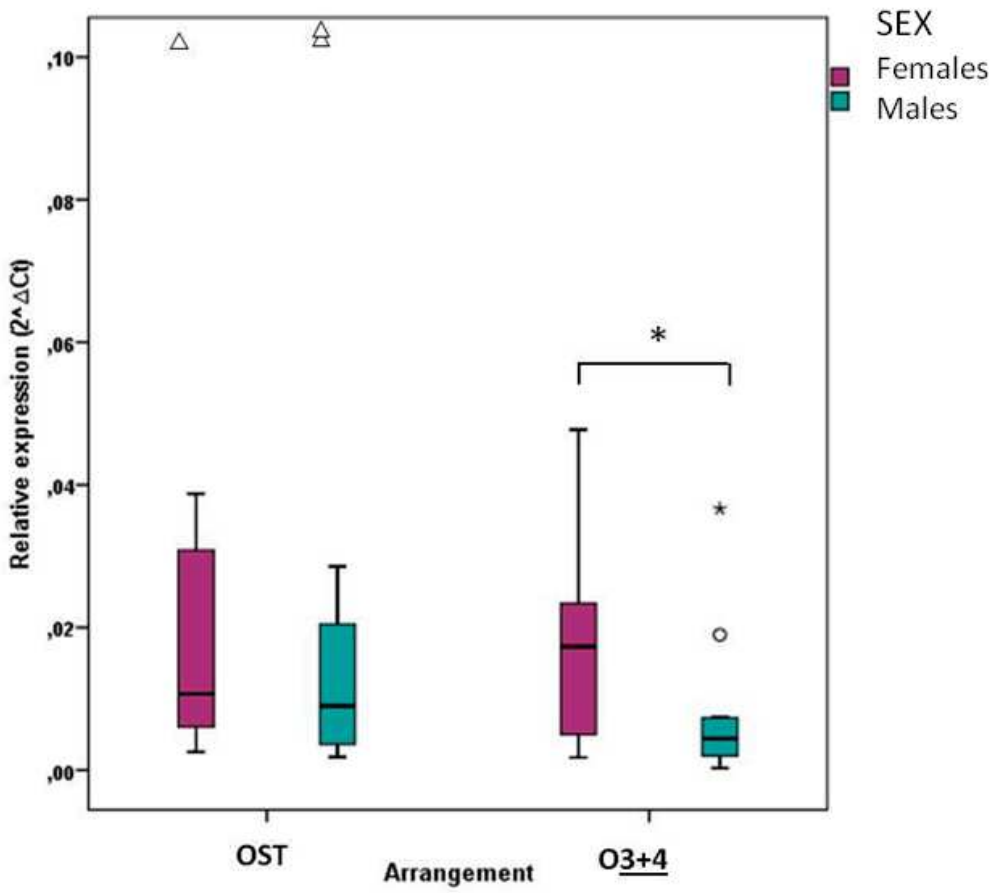


Table 1: Comparisons of *hsp70* mRNA expression rates.

Comparisons	W	P
Sex: Males vs females	1018.5	0.008 *
Arrangement: Ost vs O3+4	1070	0.108 (NS)
Arrangement Ost: Males vs females	272	0.394 (NS)
Arrangement O3+4: Males vs females	248	0.006 *
Males: Ost vs O3+4	240	0.029 (NS)
Females: Ost vs O3+4	299.5	0.832 (NS)

W: Wilcoxon rank sum statistic test; P: p-value; *significant values after Bonferroni correction (p<0.0083); NS: Non-significant.

Basal Hsp70 protein levels are influenced by sex and arrangement.

A two-way nested ANOVA test was performed in the complete data set of protein expression values (n = 72). We show no significant differences between sexes (P = 0.770) or arrangements (P = 0.880, Table 2) but a significant interaction effect between arrangement and sex (P = 0.006) is observed: O_{ST} males tended to have increased protein levels relative to O_{ST} females, whereas the opposite trend was observed for the $O_{\underline{3+4}}$ arrangement (Figure 2). Globally, these results disagree with those of Calabria et al. (2012), where basal increased Hsp70 protein levels in the $O_{\underline{3+4}}$ arrangement together with no interaction between sex and arrangement were observed.

Different factors could be at the origin of the discrepancies between the basal increased Hsp70 protein levels formely observed in $O_{\underline{3+4}}$ (Calabria et al. 2012) and the values reported in this study for the lines wearing the same arrangement. For instance, there are some notorious genetic differences between the lines used in Calabria et al. (2012) and ours. Even though both studies followed a similar isogenization protocol to obtain fully homozygous O chromosomes (Sperlich et al. 1977), the lines in Calabria et al. (2012) corresponded to F1 crosses between isogenic lines for the same arrangement. Consequently, the O chromosomes were more heterozygous in the lines in Calabria et al. (2012) than in ours and possibly more representative of what occurs in nature, where individuals with isogenic chromosomes as in our lines are rare. In addition, the $O_{\underline{3+4}}$, $O_{\underline{3+4+8}}$ and O_{ST} arrangements considered in Calabria et al. (2012) were made isogenic using the *Varicose* chromosome of the balanced lethal *Va/Ba* strain that carries the $O_{\underline{3+4}}$ inversion in the SI segment, while our $O_{\underline{3+4}}$ lines were isogenized using the *Bare* chromosome that carries the O_{ST} arrangement. It must be noted that none of the two methodologies can guarantee 100% nonrecombinant O chromosomes. In fact, gene flow has been attested between the $O_{\underline{3+4}}$ and O_{ST} segments (Rozas et al. 1999). Consequently, recombination in heterozygous females during the isogenization process may have introduced a limited amount of genetic variability in the O chromosomes of our lines and those in Calabria et al. (2012).

Figure 2. Box plots showing Hsp70 protein levels in O_{ST} and $O_{\underline{3+4}}$ males and females. Males are represented in turquoise and females in pink. The circle corresponds to a mild outlier (above 1.5-fold the interquartile range). Notice that $O_{\underline{3+4}}$ males tend to have decreased levels relative to their female counterparts, while the opposite is observed in O_{ST}.

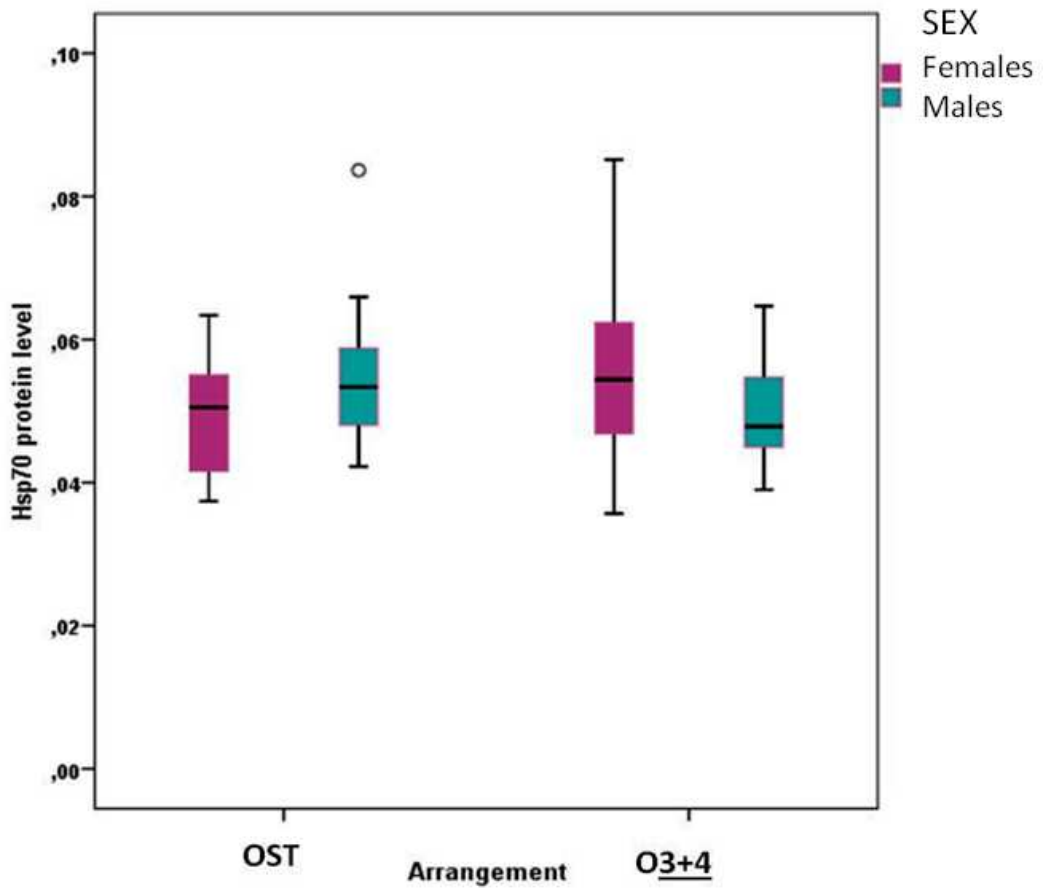


Table 2: ANOVA for the effects of sex, arrangement and line on Hsp70 protein levels.

Source of variation	d.f.	Sum of squares	F	P
Sex (S)	1	7.04E-06	0.09	0.770 (NS)
Arrangement (A)	1	3.05E-06	0.03	0.880 (NS)
Line ⊂ A	10	0	1.80	0.080 (NS)
S x A	1	1.00E-03	8.01	0.006**

⊂: nested, d.f.: degrees of freedom; **p-value<0.01; NS: Non-significant.

It must be noted that inbreeding is a potential factor that could elevate basal Hsp70 levels (Kristensen et al. 2002). Considering their length, isogenic O chromosomes may increase the total inbred load by about 20%. It is also worth mentioning that the lines used in Calabria et al. (2012) and ours stem from different locations (Barcelona and the Basque Country, respectively). Although the $O_{\underline{3+4}}$ and O_{ST} arrangements show little population structure, they could have captured multiple alleles that are individually adaptive in different locations according to the local adaptation hypothesis that might influence *hsp70* gene expression regulation (Kirkpatrick and Barton 2006; Pegueroles et al. 2013).

The sex x arrangement interaction effect on Hsp70 protein levels indicates that these are influenced by sex-linked factors and also by the inversion polymorphism in the O chromosome. Regarding the first, certain members of the Hsp70 family have shown to participate in mouse spermatogenesis (Govin et al. 2006) and Hsp70 protein levels are under hormonal influence in certain mammal cell types (Milne and Noble 2008; Bombardier et al. 2009). In addition, Hsp70 protein levels seem to differ between adult *D. melanogaster* males and females after thermal stress in some studies (Dahlgaard et al. 1998; Lerman et al. 2003; Folk et al. 2006). Then, it would not be strange that basal Hsp70 protein levels are influenced by sex. The fact that the pattern is opposite between O_{ST} and $O_{\underline{3+4}}$ indicates that the genetic content in these two arrangements has an effect on Hsp70 protein levels too. Nevertheless, the high degree of conservation in the coding regions of the two *hsp70* genes and their proximal promoters in O_{ST} and $O_{\underline{3+4}}$ suggests that the extant polymorphism at the *Hsp70IR* locus is unlikely to explain the interaction effect (Puig-Giribets et al. 2018 and this study). On this line, epistatic effects that influence basal Hsp70 protein levels might take place between unknown loci in the two considered O-chromosome arrangements and the sexual chromosomes. However, it is worth noting that no association between inversions in the A and O chromosomes has been found in *D. subobscura* (Zivanovic et al. 2016) (there are no described inversions in the Y chromosome; Krimbas 1993), and that all individuals in our experiment are expected to carry genetically indistinguishable A chromosomes derived from the homokaryotypic *ch-cu* strain (Koske and Smith 1954). In contrast, Y chromosomes are inherited from wild males and may therefore differ genetically between lines. However, Y chromosomes in *Drosophila* seem to contain very few genes that are principally involved in male fertility (Charlesworth 2001). When put together, these observations are difficult to reconcile with our results, and further replication is warranted.

In summary, our data concerning basal *hsp70* mRNA and Hsp70 protein levels does not back up the results of Calabria et al. (2012) and indicates that *hsp70* expression patterns do not explain the adaptive value of certain warm climate-associated arrangements in *D. subobscura* such as $O_{\underline{3+4}}$. Measuring *hsp70* mRNA and Hsp70 protein levels during and/or after heat stress in lines equivalent to the ones used in this work might help to validate our preliminary results.

Lack of correlation between mRNA and protein levels.

To test if there is an association between *hsp70* mRNA and Hsp70 protein levels within the same line, we applied the Spearman rank-sum correlation test using the average mRNA relative expression levels and protein amounts per line considering males and females separately. No significant correlation was observed in any comparison (Table 3). This is unexpected considering that a positive and relatively high correlation between mRNA and protein abundance has been noted for Hsp70 and other Hsps in *D. melanogaster* females during heat stress (Colinet et al. 2013; Sørensen et al. 2013) that is even more apparent than in other lineages such as springtails (Collembola) due to a delay in Hsp70 protein production in the latter (Bahrndorff et al. 2009). Interestingly, a lack of correlation is observed in *D. melanogaster* females during exposure at cold temperatures, which could result from insufficient mRNA levels to cause a significant increase in Hsp70 protein abundance, differences in temporal dynamics between mRNA and protein production or absence of expression at 11ºC (Colinet et al. 2013).

Table 3: Spearman's Rho correlation coefficient values calculated using the average values of relative *hsp70* mRNA expression and protein levels in the 12 lines. The correlation coefficient has been obtained for protein and mRNA amounts in males (MAL) and for protein and mRNA amounts in females (FEM).

Correlations	Number of lines	Spearman's Rho	Sig. (two-tailed)
mRNA FEM vs. protein FEM	12	0.140	0.665 (NS)
mRNA MAL vs. protein MAL	12	-0.112	0.729 (NS)

NS: Non-significant.

For instance, differences in the genetic content between individuals of the same line could explain the lack of correlation between *hsp70* mRNA and protein values. Indeed, Hsp70 protein levels have been shown to differ between genetically diverse *D. melanogaster* strains after heat shock (Krebs and Feder 1997a; Krebs et al. 2001). However, in our study the flies that belong to the same line are expected to carry genetically identical O chromosomes and the genetic background of the *ch-cu* strain in the remaining four acrocentric chromosomes. The high diversity observed within the same line for both mRNA and protein levels might indicate that some or most of our lines are more genetically diverse than assumed.

Besides genetic diversity, there are other explanations for the lack of an association between *hsp70* mRNA and protein amounts. For example, negative associations

between transcript and protein amounts in some heat-shock genes can result from inhibition of intron splicing (Arbona et al. 1993). However, the correlation is not expected to be lowered in the two intronless *hsp70* copies in *D. subobscura* (Puig-Giribets et al. 2018). Provided that the targeted coding regions are virtually identical in the two recently described *hsp70* paralogs of the 12 lines as a result of ectopic gene conversion (Puig-Giribets et al. 2018), the reported mRNA and protein values per sample likely correspond to the sum of two isoforms. Interestingly, the 3'UTRs appear to be considerably more divergent between the two *hsp70* paralogs than the coding and proximal promoter regions (Puig-Giribets et al., in preparation), which might point to half-life differences between the two *hsp70* mRNA isoforms. In fact, it is known that *hsp70* transcript half-life is regulated by AU-rich sequences in the 3'UTR (Petersen and Lindquist 1989; Balakrishnan and De Maio 2006). More precisely, the messenger is rapidly degraded after deadenylation of its 3'UTR under normal conditions, yet these mechanisms are inactivated by heat shock, resulting in an extended half-life of *hsp70* mRNA transcripts (Bönisch et al. 2007). Considering the harmful effects that Hsp70 protein levels have for the cell at normal conditions (Feder et al. 1992), one might conclude that basal *hsp70* transcript half-life is kept short to produce minimal Hsp70 protein amounts under non heat-shock conditions. This trend might also help explain the apparent lack of association between *hsp70* mRNA and protein amounts in our lines.

3.2.1.5 Acknowledgments

This work was supported by research grant CGL2013-42432-P from the Ministerio de Economía y Competitividad (Spain) and grant 2014 SGR 1346 from Generalitat de Catalunya to the Grup de Genòmica, Bioinformàtica i Biologia Evolutiva (GBBE). M.P.G. was supported by a PIF PhD fellowship from the Universitat Autònoma de Barcelona (Spain) and an *ERASMUS+* grant during a three-month stay at Aarhus Universitet (Denmark). We would like to thank Dr. Mauro Santos for his collaboration in the statistical treatment of the data, the laboratory technician Doth Andersen for teaching us the ELISA protocol and Sergi Martínez and Alba Vilella for their help in fly maintenance.

3.3 Analysis of the extant polymorphism in several distinctive regions of the *Hsp70IR* locus in O_{ST} and in the $O_{\underline{3+4}}$ phylad (including $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$).

Chapter 3.3 consists of a research article **"**A complex interplay between stability, convergence and divergence drives the evolution of the *Hsp70IR* locus in *Drosophila subobscura*" that is currently in preparation. The Supplementary Material of this article can be found in Appendix 7.3.

3.3.1 A complex interplay between stability, convergence and divergence drives the evolution of the *Hsp70IR* locus in *Drosophila subobscura*

3.3.1.1 Abstract

The *Hsp70IR* locus shows a similar genomic organization in the closely related *D. guanche*, *D. madeirensis* species and also in four *D. subobscura* arrangements (O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+16}}$). This locus consists in two head-to-head oriented intronless *hsp70* genes separated by an intergenic, non-duplicated region. Polymorphism analyses indicate that the high similarity between the two *hsp70* genes in *D. subobscura* and *D. madeirensis* is enhanced by gene conversion and purifying selection in coding and promoter DNA regions. In contrast, the intergenic region separating the two *hsp70* genes is characterized by numerous repetitive DNA sequences alongside abundant INDEL polymorphism. In former work it was found that the divergence between the two paralogous *hsp70* genes was lower in the cold climate-associated O_{ST} inversion, a trend that might be enhanced by natural selection. To find evidence of selection in the *hsp70* gene family of *D. subobscura*, in this work we explore the nucleotide diversity patterns and test for deviations from neutrality across a ~6 kb region at the *Hsp70IR* locus using several haplotypes of O_{ST} and $O_{\underline{3+4}}$. Our results confirm the high degree of conservation at the nucleotide and organizational levels within *Hsp70IR* in these two adaptive arrangements. Nucleotide diversity is substantially lower in O_{ST} than in $O_{\underline{3+4}}$, which might indicate that the former has a smaller *Ne* assuming that the genetic diversity of our haplotypes is representative enough. Under neutrality, a possibility is that this difference is due to stochastic changes in the inversion frequencies of the source population.

3.3.1.2 Introduction

Chromosomal inversions are thought to play an important role in processes such as the evolution of sex chromosomes, speciation and local adaptation through a variety of molecular mechanisms (Hoffmann and Rieseberg 2008). For instance, the inhibited recombination in the heterokaryotypes could maintain allelic combinations that might become advantageous in a new environment (Dobzhansky 1970, Kirkpatrick and Barton 2006). Besides suppression of recombination, the inversion process itself can disrupt the expression patterns of flanking genes, creating new phenotypes (Puerma et al. 2016).

The adaptive value of chromosomal inversion polymorphism is widely attested in the genus *Drosophila*, where certain inversions show an association with phenotypic traits such as body size, wing loading and plant toxin resistance (Fernández-Iriarte et al. 2003;

Kennington et al. 2007; Pegueroles et al. 2016). *Drosophila subobscura*, a member of the *obscura* group of *Drosophila*, is a temperate species native to the Palearctic region that colonized some parts of the American continent in late 1970s (Prevosti et al. 1988). This species exhibits a rich inversion polymorphism in its 5 acrocentric chromosomes (A, J, U, E and O, plus a small dot–like). Surprisingly, some of these arrangements have evolved latitudinal clines in the colonizer populations that mirror those already present in the original Palearctic populations, which indicates they are adaptive (Prevosti et al. 1988; Balanyà et al. 2006; Castañeda et al. 2013). The O chromosome is the most polymorphic with 24 described inversions that form complex chromosomal arrangements of overlapping and non-overlapping inversions (Krimbas 1993). It is conventionally divided into the distal (SI) and the proximal (SII) segments relative to the centromere. Several alternative inversions have been described in the SI, including O_{ST} and $O_{\underline{3+4}}$. These two inversions show an opposite latitudinal component, being the cold climate-associated O_{ST} inversion very frequent in Northern Europe and relatively rare around the Mediterranean, while the warm climate-associated $O_{\underline{3+4}}$ shows the opposite distribution. In addition to a latitudinal component, cyclic seasonal shifts of some chromosomal arrangements, such as the warm climate-associated $O_{\underline{3+4}}$ and the cold climate-associated O_{ST}, have been detected in long-term seasonal collections in the same population accompanied by a gradual expansion of low latitude genotypes poleward (Fontdevila *et al.*, 1983; Rodríguez-Trelles *et al.*, 1996; Rodríguez-Trelles and Rodríguez, 1998), a trend that has been interpreted as a genome response to climate warming (Balanyà *et al.*, 2009; Rezende *et al.*, 2010).

The ample evidence on the adaptive value of certain *D. subobscura* inversions contrasts with the lack of knowledge about the genetic basis responsible of their maintenance in natural populations. For instance, heat shock resistance experiments carried out by Quintana and Prevosti (1991) indicated the presence of possible heat shock factors in the A chromosome and in the $O_{\underline{3+4}}$ inverted region in the SI of the O chromosome. Shortly after, it was determined by *in situ* hybridization (ISH) that the heat shock-inducible *hsp70* gene family was localized in the 94A band within the $O_{\underline{3+4}}$ inverted region (Moltó et al. 1992), which shows high genetic differentiation between the cold climate-associated O_{ST} and the warm climate-associated $O_{\underline{3+4}}$ arrangements (Munté et al. 2005; Calabria 2012; Pegueroles et al. 2013). Interestingly, the heat shock experiments of Calabria et al. (2012) revealed that homokaryotypic flies for the warm climate-

associated O_{3+4} arrangement showed increased Hsp70 protein levels under basal conditions that did not boost with heat stress. These results are difficult to interpret considering that the Hsp70 inducible profile of O_{3+4} is atypical of species with short life cycles like *Drosophila,* where high basal Hsp70 levels have negative consequences for development (reviewed in Garbuz and Evgen'ev 2017).

A possibility to explain the outcomes of Calabria et al. (2012) is that warm-adapted arrangements such as O_{3+4} have an extra number of *hsp70* copies that may enhance higher inducible Hsp70 expression levels and heat resistance, as it has been observed for some natural and artificial *Drosophila* strains with increased *hsp70* copy number (Feder et al. 1996; Evgen'ev et al. 2004; Bettencourt et al. 2008). However, in former work we demostrated that a similar genomic organization within *Hsp70IR* prevails in the four studied inversions (O_{ST}, O_{3+4}, O_{3+4+16} and O_{3+4+8}) that consists in two ~2.5 kb head-to-head oriented *hsp70* genes (conventionally labeled as *hsp70A* and *hsp70B*) separated by a ~1 kb intergenic region (Puig-Giribets et al. 2018). The *hsp70A* duplicate is flanked by the *Dalmatian* (*Dmt*) gene, while *hsp70B* is flanked by *CG5608*, forming a *CG5608–Hsp70IR–Dmt* synteny block that is preserved in *D. subobscura* and *D. madeirensis*. Besides extra copies, increased *hsp70* transcript levels can also be achieved by changes in *cis*-regulatory DNA regions (Bettencourt et al. 2002; Lerman and Feder 2004). Nonetheless, we failed to find any conspicuous inversion-associated polymorphisms in the *hsp70* proximal promoters of O_{ST}, O_{3+4+16}, O_{3+4+8} and O_{3+4} that contain each four highly conserved heat shock elements (HSEs 1-4) in the *hsp70* promoters of *Drosophila* annotated by Tian et al. (2010) and three GAGA (GAF) binding sites described in Puig-Giribets et al. (2018).

Intriguingly, phylogenetic analyses using the nucleotide variation at the *Hsp70IR* locus in *D. subobscura* reveal higher identity between *hsp70* gene paralogs of the same inversion than between orthologs of different inversions (Puig-Giribets et al. 2018). Considering that the most parsimonious scenario is that the duplication that led to the *hsp70* inverted pair predates the divergence of the considered arrangements and species, these outcomes indicate that the two *hsp70* genes evolve in concert by means of a homologous recombination mechanism akin to gene conversion (GC). GC is described as a common form of double strand break repair by which one DNA sequence replaces a homologous sequence in the same or in another chromosome such that the two

segments become identical after the conversion event. Although the molecular mechanisms responsible of GC are not precisely known, it has been detected in the sequences of *hsp70* genes from several dipterans (Brown and Ish-Horowicz 1981; Benedict et al. 1993; Bettencourt and Feder 2001; Bettencourt and Feder 2002; Evgen'ev et al. 2004). The evolutionary consequences of GC are many and diverse (reviewed in Fawcett and Innan 2011). For instance, recurrent GC among gene duplicates emulates an increase in the effective population size, because it allows a faster spread of beneficial mutations while deleterious ones are rapidly eliminated. Surprisingly, the identity between *hsp70* duplicates seem to vary between lineages, being higher in O_{ST} and lower in *D. guanche* compared to that in the $O_{\underline{3+4}}$ phylad (including $O_{\underline{3+4+8}}$, $O_{\underline{3+4}}$ and $O_{\underline{3+4+16}}$). This indicates that differences in the rate of GC might exist between different inversions and closely related species, and that these could be somehow adaptive, for example, by enhancing a more efficient heat shock response in the case of O_{ST} (Puig-Giribets et al. 2018).

Taking into account the information available on the *Hsp70IR* locus in *D. subobscura* provided by Puig-Giribets et al. (2018), in the present work we have reconstructed an equivalent ~10 kb genomic region in seven additional *D. subobscura* lines with homozygous O chromosomes that we added to the ones previously analyzed making a total of 12 (O_{ST} (4), $O_{\underline{3+4+16}}$ (2), $O_{\underline{3+4+8}}$ (1) and $O_{\underline{3+4}}$ (5)) to pursue the following objectives: i) to analyze, describe and compare the evolutionary trends of several distinctive regions within the *Hsp70IR* locus in the four aforementioned O-chromosome arrangements of *D. subobscura* and in the closely related *D. madeirensis* and *D. guanche*, ii) to explore an association between the extant polymorphism through the *Hsp70IR* locus and the adaptive value of O_{ST} and $O_{\underline{3+4}}$ and iii) to quantify and compare the dynamics of GC in the O_{ST} and $O_{\underline{3+4}}$ arrangements using the extant polymorphism of the two *hsp70* duplicates that evolve in concert in *D. subobscura* and *D. madeirensis* as noted in Puig-Giribets et al. (2018). As expected, our results evidence purifying selection in the nonsynonymous polymorphism of the two *hsp70* CDSs and comparable evolutionary rates among the two highly conserved *hsp70* genes in O_{ST} and $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad (including $O_{\underline{3+4+16}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4}}$). In agreement with Puig-Giribets et al. (2018), we confirm that the high identity between the two *hsp70* duplicates in the proximal promoter and coding regions is maintained by purifying selection enhanced by GC events that periodically erase the accumulated divergence. In contrast, we did not

found a significantly higher nucleotide identity between the two paralogous CDSs in O_{ST} than in $O_{\underline{3+4}}$, which indicates that GC events might take place by similar mechanisms in the two arrangements assuming constant mutation rates.

3.3.1.3 Materials and Methods

***Drosophila* lines.**

All *D. subobscura* stocks used in this work were obtained and provided by Drs. R. Tarrío and F. Rodríguez-Trelles. These include 12 *D. subobscura* lines isogenic for 4 representative combined arrangements that occur in the SI ($O_{\underline{3+4+8}}$, $O_{\underline{3+4}}$, O_{ST}, and $O_{\underline{3+4+16}}$) (nomenclature according to Krimbas 1993). To obtain homozygous O chromosomes, Drs. R. Tarrío and F. Rodríguez-Trelles captured wild males seasonally in 2012 and 2014 in five localities from Iberian Peninsula (Berbikiz (2012): 43°11′20.31″ N, 3°5′23.74″W; Jérez del Marquesado (2014): 37º11'6.74"N, 3º10'30.37"W; Vélez de Benaudalla (2014): 36º50'23.66"N, 3º30'59.32"W and Juviles (2014): 43°38'19.39"N, 116°14'28.86"W; see Supplementary table S1 for a more detailed list) and crossed them to virgin females of the *Varicose/Bare* (*Va/Ba*) balanced lethal stock selected for strong *Bare* penetrance and of known karyotype (Sperlich et al. 1977). The chromosomal arrangement of wild O chromosomes was determined by direct observation of polytene chromosomes of salivary glands of third instar larvae. All stocks were reared at 18ºC on standard *Drosophila* medium.

Reconstructing a 9-10 kb genomic region at the *Hsp70IR* locus in 12 lines and three species.

Using the sequence information of the five *D. subobscura* lines isogenic for O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+16}}$ analyzed in Puig-Giribets et al. (2018) available in Genbank (Accession numbers: MG780233-MG780237), we sequenced an equivalent ~10-kb long region in the *Hsp70IR* locus in seven additional isogenic lines (O_{ST} (2), $O_{\underline{3+4}}$ (4) and $O_{\underline{3+4+16}}$ (1)) making a total of 12 (see Figures 1A-1B for a representation). The name of each line consists in its arrangement followed by a number (ex: O3+4_1). We performed several separate PCR reactions of ~1.5-2.0-kb long regions with overlapping segments big enough to diminish sequencing errors. The list of primers and their sequences, primer combinations used for PCR amplification and their location within the *Hsp70IR* locus are available at Supplementary table S2 and Supplementary figure

S1. Amplicons were sequenced on one strand by Next-Generation Sequencing (NGS) using Macrogen's sequencing service (Macrogen Europe, Amsterdam, The Netherlands), and the homologous sequences of the different lines were aligned together and manually checked for sequencing errors. The previously reconstructed 9-10 kb *Hsp70IR* genomic regions in the closely related *D. madeirensis* and *D. guanche* served as outgroup sequences (Accession: MG780238-MG780239).

PCR amplifications.

DNA for PCR reactions was extracted from 5-10 whole adult individuals per line using phenol-chloroform and isopropanol precipitation as described in Puig-Giribets et al. (2018). Primers for PCR amplification were designed with the online tool Primer3plus (Utergasser et al. 2007; available at: http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). PCR reactions were carried out in a final volume of 25 µl, including 2.5 KCL buffer (Bioron, Ludwigshafen, Germany), 0.2 mM of each dNTP (Roche Diagnostics, Indianapolis, IN, US), 0.4 µM primer (Sigma-Aldrich, Gillingham, Dorset, UK), template DNA (50-100 ng) and 0.04 units of DFS-Taq DNA polymerase (Bioron, Ludwigshafen, Germany). Amplifications were run in a MJ Research Inc. Thermal Cycler (MJ Research Inc., Watertown, MA, US) programmed as follows: 2 min preliminary denaturation at 94ºC, 30 cycles of 10 sec at 94ºC (denaturation), 20 sec at specific PCR annealing temperatures, and 3 min at 72ºC (extension), and a final extension for 10 min at 72ºC. PCR products were gel purified with the PCR Clean-up kit (Macherey-Nagel, Düren, Germany), quantified (NanoDrop-2000, Thermo Scientific) and checked by agarose gel electrophoresis before sequencing.

Characterization of the *Hsp70IR* genomic region and protein prediction.

To make solid inferences based on nucleotide diversity, we first performed BLASTN searches at the NCBI platform to delimit the two ~2.5-kb long *hsp70A* and *hsp70B* paralogous genes in the 14 lines including *D. madeirensis* and *D. guanche* (http://www.ncbi.nlm.nih.gov/BLAST). Similar to Puig-Giribets et al. (2018), all examined *hsp70* duplicates included a unique 1929 bp intronless and highly conserved coding sequence strand (CDS). In contrast with the CDSs, the limits of the palindrome-like structure of the two *hsp70* duplicates remained unclear due to the numerous poorly aligned sites in the non-functional regions. For instance, the highest nucleotide identity between the two paralogs is observed in a region that includes the complete CDS and

the first ~250 bp upstream the start codon, where several *cis*-regulatory elements have been previously described (Tian et al. 2010; Puig-Giribets et al. 2018). The identity between the two paralogs ranges from 75% to 95% and stretches another further ~300 bp upstream the proximal promoter and ~200 bp downstream the stop codon until a conserved polyadenylation sequence (AATAAA). Using this information and their distinctive molecular function, we delimited three regions within each hsp70 duplicate that we labeled as follows: 3_B: 3'UTR of the hsp70B copy, CDS_B: coding region in the hsp70B copy; 5_B: 5'UTR of the hsp70B copy; 5_A: 5'UTR of the hsp70A copy; CDS_A: coding region of the hsp70A copy; 3_A: 3'UTR of the hsp70A copy. Lastly, we identified a ~1-kb long intergenic, non-duplicated region between the two hsp70 genes with no significant homology to any other locus that we labeled as 'IGR' (see Figures 1A-1B for a representation). The flanking genomic regions downstream hsp70A and hsp70B copies that included parts of the *CG5608* and *Dmt* genes were not considered for analysis. The *hsp70* amino acid sequences were predicted with AUGUSTUS software (Stanke and Morgenstern 2005; available at: http://augustus.gobics.de) and aligned with BioEdit software (Hall 1999; available at: www.mbio.ncsu.edu/BioEdit/BioEdit.html).

Estimates on DNA polymorphism: neutrality tests.

The software DnaSP v.5.10.01 (Librado and Rozas 2009; available at: http://www.ub.edu/DnaSP/) was used to estimate nucleotide polymorphism and divergence, the genetic differentiation between populations, and to detect recombination tracts. Deviations from neutrality were checked using several neutrality tests implemented in the same software. Most neutrality tests were performed in the different arrangements separately (O_{3+4}, represented by three non-recombinant lines, O_{ST}, represented by four lines and the O_{3+4} phylad, represented by eight lines of the closely related O_{3+4}, O_{3+4+8} and O_{3+4+16} arrangements). Specifically, we used the following neutrality tests: the Tajima's *D* test that compares the average number of pairwise differences with the number of segregating sites (Tajima 1989) and the Fu and Li's *D** and *F** tests that compare the number of derived singletons to the total number of mutations (η) (Fu and Li's *D**) and to the average number of nucleotide differences between pairs of sequences (*k*) (Fu and Li's *F**) (Fu and Li 1993). Fu and Li's *D** and *F** statistics were also estimated using *D. guanche* as outgroup to infer the ancestral and derived states of polymorphic sites. Finally, we computed the normalized version of the

Fay and Wu's H-test that looks for derived alleles that are significantly more frequent than expected by the neutral theory (Fay and Wu 2000), where the H statistic values are scaled by the variance of the statistic as described in Zeng et al. (2006). These neutrality tests were chosen because they are widely used in population genetics and it is interesting to compare their results since they rely on different attributes of DNA sequence data as previously detailed. Some tests required a minimum of four sequences per population (including Tajima's *D* and Fu and Li's *D** and *F**) and could therefore not be performed for O_{3+4} and O_{3+4+16}.

Estimates on DNA polymorphism: multiple sequence alignment.

Two types of polymorphism were analyzed separately: single-nucleotide polymorphism (SNP) and insertion-deletion (INDEL) polymorphism. For SNP analyses, the complete 9-10 kb *Hsp70IR* sequences of the 14 lines were first aligned with the MAFFT 7.243 E-INS-i algorithm optimized for multiple conserved domains and long gaps (Katoh and Standley 2013; available at: https://mafft.cbrc.jp/alignment/server/index.html). To get rid of poorly aligned positions, the alignment was cured by the software G-blocks v. 0.91b (Castresana 2000; available at: http://molevol.cmima.csic.es/castresana/Gblocks_server.html) with the most stringent selection option that does not allow many contiguous nonconserved positions (Parameters used: minimum number of sequences for a conserved position: 7; minimum number of sequences for a flanking position: 10; maximum number of contiguous nonconserved positions: 4; minimum length of a block: 10; allowed gap positions: none). Using these procedures, we obtained our final multiple sequence alignment (MSA) that consisted in 5978 bps. To analyze INDEL polymorphism, two separate analysis have been performed after the initial alignment with MAFFT 7.243 E-INS-I, one with no curation to retain all INDELs in our *D. subobscura* data set and another using curation in Gblocks with the least stringent criteria to preserve the more conserved INDELs (Parameters used: minimum number of sequences for a conserved position: 7; minimum number of sequences for a flanking position: 7; maximum number of contiguous nonconserved positions: 8; minimum length of a block: 5; allowed gap positions: with half). To explore polymorphism and divergence between orthologous and paralogous *hsp70* genes, the two ~2.5 kb palindromic duplicates in the 14 lines were first aligned with MAFFT 7.243 E-INS-i and subsequently cured with G-blocks using the most stringent selection option (Parameters used: minimum number of

sequences for a conserved position: 7; minimum number of sequences for a flanking position: 10; maximum number of contiguous nonconserved positions: 4; minimum length of a block: 10; allowed gap positions: none), obtaining a 2533 bp MSA. Statistical tests were performed with the SPSS v.22.0 software (IMB SPSS Inc., NY, US) and a *P* value of 0.05 or below indicated a significant difference. Parametric tests were chosen over non-parametric ones when the criteria of normality and equal variances among groups were met.

Phylogenetic analyses.

The software TREE-PUZZLE v.5.2 (Schmidt et al. 2002; available at: http://www.tree-puzzle.de/) was used to test for differences in the ratio of the consituent monomer units between the 14 lines. As expected, no significant differences were noted, indicating that the nucleotide composition at the *Hsp70IR* locus is similar in these closely related lineages. On the other hand, although inversions do inhibit recombination, a proportion of recombinant tracts in homologous regions may still exist in our lines. This could mask the evolutionary history among the considered inversions and species. To identify recombination breakpoints, we employed a variety of methods implemented in the RDP4 v.4.95 software (Martin et al. 2015; available at: http://web.cbio.uct.ac.za/~darren/rdp.html) and the Psi (Ψ) parameter (Betrán et al. 1997) implemented in DnaSP. A ~1.2 kb recombinant tract from an O_{ST} line was detected in the intergenic region separating the two *hsp70* homologs of one O_{3+4} line (O3+4_1) that showed significant *P*-values in the MaxChi, GENECONV, 3Seq and Bootscan tests. Since there is a strong genetic structure between the O_{ST} and O_{3+4} arrangements, we performed separate phylogenetic analyses including and excluding the O3+4_1 intergenotype to infer the chromosomal inversion ancestry. We followed the procedures described in Puig-Giribets et al. (2018) to carry out the phylogenetic analyses with and without O3+4_1. Briefly, the genomic region including the *hsp70A* and *hsp70B* copies and the IGR was first aligned with the MAFFT 7.243 L-INS-I algorithm (Katoh et al. 2005). MSA uncertainty was assessed with Guidance2 (Penn et al. 2010; Sela et al. 2015), and columns scoring below the guidance default value of 0.935 were removed. The base MSA obtained a guidance confidence score of 0.982, where a score of 1 indicates 100% robustness of the MSA to 100 bootstrap perturbations in the guide tree. To reconstruct phylogenetic trees, we perfomed a maximum likelihood analysis using the latest version of IQ-TREE (W-IQ-TREE) with

default parameters (Trifinopoulos et al. 2016; available at: http://iqtree.cibiv.univie.ac.at/). The ultrafast bootstrap approximation was used to determine branch support (UFboot; 1000 replicates) (Minh et al. 2013). The same procedure has been followed to carry out phylogenetic reconstructions of the paralogous *hsp70* regions in the 14 lines. Phylogenetic trees in Newick format were drawn with the Interative Tree of Life (ITOL) online tool (Letunic and Bork, 2016; available at: https://itol.embl.de/).

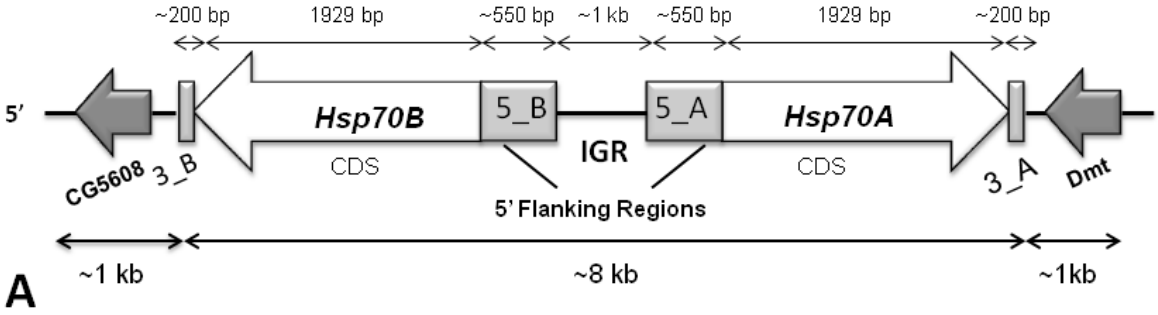
3.3.1.4 Results

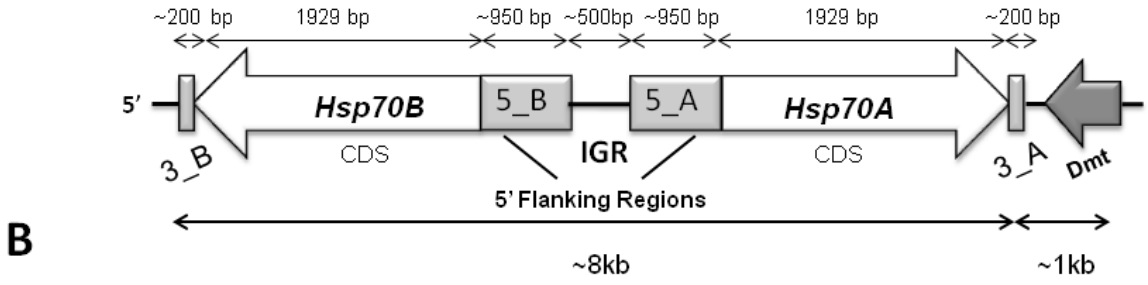
Preserved synteny at the *Hsp70IR* locus in the O_{ST}, O_{3+4+16}, O_{3+4+8} and O_{3+4} arrangements.

Previous experiments have indicated that adult carriers of the warm climate-associated O_{3+4} arrangement exhibit increased thermal preference, heat resistance (Rego et al. 2010; however, see Dolgova et al. 2010) and higher basal Hsp70 protein levels in comparison with their cold climate-associated O_{ST} counterparts (Calabria et al. 2012). The differences in *hsp70* gene expression could be due to changes in *hsp70* copy number, gene content and/or *cis*-regulatory DNA. In the present work we characterized a ~10-kb long genomic region at the Hsp70IR locus in four representative arrangements in the SI segment of the O chromosome represented by 12 isogenic lines (O_{ST} (4) O_{3+4} (5), O_{3+4+8} (1) and O_{3+4+16} (2) and in the closely related *D. madeirensis* and *D. guanche*. A similar genomic organization was found in the four *D. subobscura* arrangements and the three species of the subobscura cluster that consisted in two ~2.5-kb long hsp70 duplicates including each a 1929 bp intronless CDS arranged as an inverted pair and separated by a ~1 kb intergenic, non-duplicated region (Figures 1A-1B), which is in agreement with former work (Puig-Giribets et al. 2018). Interestingly, the organization of *hsp70* genes in the *subobscura* species cluster that resembles the ancestral condition of Diptera (Benedict et al. 1993) is also found in more divergent species of the *obscura* group of *Drosophila* such as *D. pseudoobscura* (Bettencourt and Feder 2001). This might indicate that the synteny in this locus has been preserved since the *obscura* radiation, 10-15 Myr ago (Russo et al. 1995). Thus, in agreement with former results, our data indicates that the four analyzed *D. subobscura* arrangements, including the

adaptive O_{ST} and $O_{\underline{3+4}}$, do not differ in *hsp70* copy number and show a similar genomic organization within the *Hsp70IR* locus.

Figure 1: Schemes showing several delimited regions in a ~9-10 kb reconstructed genomic region at the *Hsp70IR* locus of *D. madeirensis* and *D. subobscura* lines (**A**) and in *D. guanche* (**B**). CG5608: region downstream the *Hsp70B* copy that includes a fragment of the *CG5608* gene. 3_B: palindromic untranslated region downstream the coding sequence of the *hsp70B* copy; CDS_B: coding region of the *hsp70B* copy; 5_B: palindromic untranslated region ustream the coding sequence of the *hsp70B* copy; IGR: intergenic nonduplicated region; 5_A: palindromic untranslated region ustream the coding sequence of the *hsp70A* copy; CDS_A: coding region of the *hsp70A* copy; 3_A**:** palindromic untranslated region downstream the coding sequence of the *hsp70A* copy. Dmt**:** region downstream the *hsp70A* copy that includes a fragment of the *Dmt* gene. Notice that the intergenic region between the two *hsp70* duplicates differs a little between *D. guanche* and *D. subobscura*/*D. madeirensis*, being the IGR shorter and the two palindromic 5' flanking regions longer in the former relative to the latter.





The *Hsp70IR* locus evolves close to the neutral expectations in O_{ST} and $O_{\underline{3+4}}$.

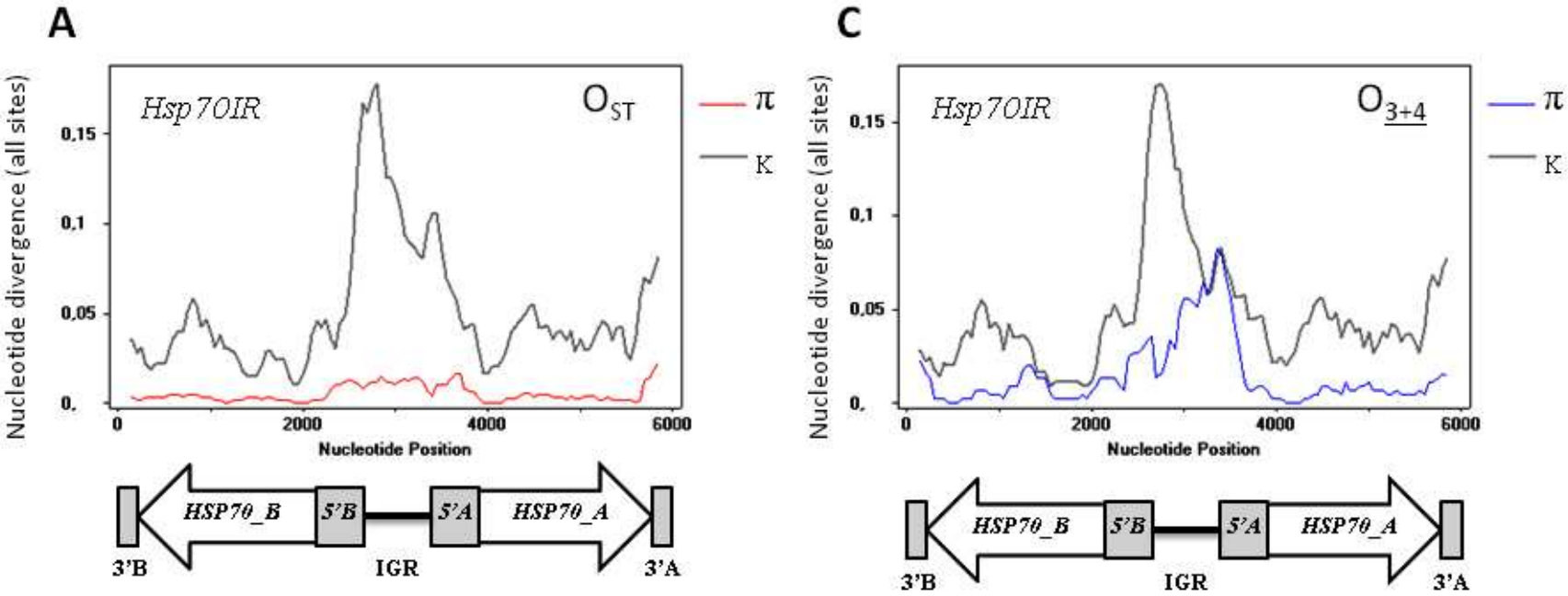
A worth exploring possibility is whether the adaptive value of the cold climate-associated O_{ST} and the warm climate-associated $O_{\underline{3+4}}$ arrangements can be explained by certain genetic variants in the *Hsp70IR* locus that are exclusive to either of the two arrangements. We used our data set of 12 *D. subobscura* lines to test for deviations from neutrality across the *Hsp70IR* locus in those arrangements represented by multiple lines, that is, O_{ST} and $O_{\underline{3+4}}$**.** Several widely used neutrality tests and population genetic parameters were estimated in seven a priori delimited regions within Hsp70IR (the two hsp70 CDSs, their respective 5' and 3' flanking regions and the IGR; see Materials and Methods for more details) of a 5978 bp MSA (Table 1 and Supplementary table S3).

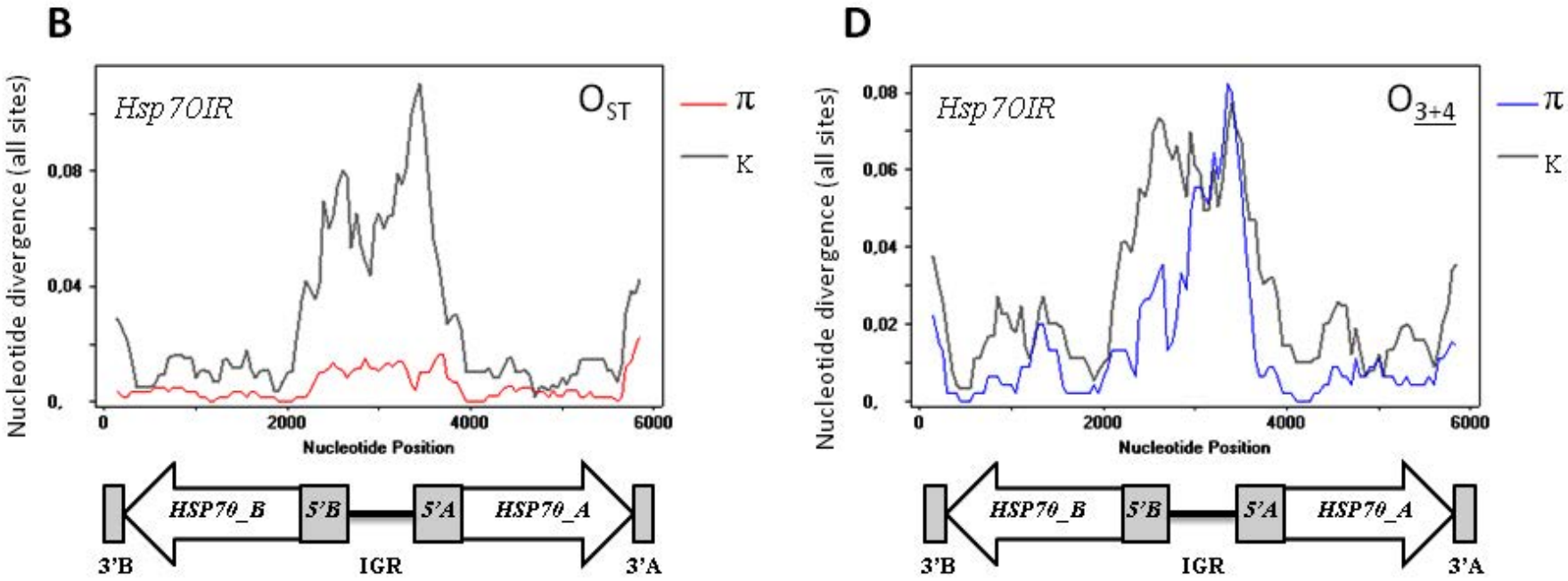
These include the number of segregating sites (S), the total number of mutations (η) and nucleotide diversity in nonsynonymous (∏a) and synonymous sites (∏s). No departure from neutrality was observed in any comparison, which indicates that the *Hsp70IR* locus evolves close to the neutral expectations in O_{ST} and in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad. As expected, the nucleotide diversity in O_{ST} and $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad is highest in non-coding DNA, especially in the intergenic, non-duplicated region between the two copies (IGR). Surprisingly, the diversity in the IGR is even higher than that in the synonymous sites, which might indicate that polymorphism in the latter could be somewhat restrained as well, for example, by codon usage bias or by a homogenization process. Conversely, the increased polymorphism in the IGR could be caused by the lack of homology in highly repetitive motifs. In the Tajima's *D* test (Tajima 1989), statistically negative values are observed in the nonsynonymous sites of the two *hsp70* paralogs (Supplementary table S3), which is probably the result of purifying selection acting on this type of polymorphism. At the same time, the nucleotide diversity is higher at synonymous sites than in the 5'UTRs, indicating that the latter might be also constrained; this is reinforced by the presence of four highly conserved cis-regulatory heat shock elements (HSEs) and three GAGA sites in the proximal promoter region of the two hsp70 genes that control their gene expression (Tian et al. 2010; Puig-Giribets et al. 2018). As expected, we could identify these conserved cis-regulatory elements in the proximal promoter regions of the seven remaining D. subobscura lines used in this work (data not shown). Curiously, we noticed that the nucleotide diversity in most noncoding regions is substantially lower in O_{ST} than in $O_{\underline{3+4}}$. Considering that the patterns of variability across *Hsp70IR* are similar in the two arrangements (in equilibrium with no departure from the standard neutral model (SNM) of evolution (Kimura 1983) our estimates of nucleotide diversity levels at putatively neutral sites indicate that O_{ST} might have a lower effective population size (Table 1). The difference in nucleotide variability between these two arrangements can be appreciated in plots of the estimated polymorphism levels across the *Hsp70IR* locus (Figures 2A-2 D).

Table 1: No departure from neutrality is observed in any of the seven delimited regions at the *Hsp70IR* locus. The estimates of nucleotide diversity and several neutrality tests have been performed on a 5978 bp cured MSA in three non-recombinant O_{3+4} lines and four O_{ST} lines. *S*, number of segregating sites. η (eta), total number of mutations. ∏, nucleotide diversity; ∏(a), nonsynonymous nucleotide diversity; ∏, silent nucleotide diversity; **3'B**: 3'UTR of the *hsp70B* copy, **CDS_B**: coding region of the *hsp70B* copy. **5'B**: 5'UTR of the *hsp70B* copy. **IGR:** intergenic nonduplicated region. **5'A**: 5'UTR of the *hsp70A* copy, **CDS_A**: coding region of the *hsp70A* copy. **3'A**: 3'UTR of the *hsp70A* copy. ***Hsp70IR***: the genomic region comprising the two *hsp70* copies and the IGR (see Figures 1A-1B for more details). JC: values after Jukes-Cantor distance correction (Jukes and Cantor 1969). Na = not applicable due to insufficient samples. SD: standard deviation of ∏.

Genomic region	**3'B**		**CDS_B**		**5'B**		**IGR**		**5'A**		**CDS_A**		**3'A**		***Hsp70IR***	
Arrangement	**O_{ST}**	**O_{3+4}**	**O_{ST}**	**O_{3+4}**	**O_{ST}**	**O_{3+4}**	**O_{ST}**	**O_{3+4}**	**O_{ST}**	**O_{3+4}**	**O_{ST}**	**O_{3+4}**	**O_{ST}**	**O_{3+4}**	**O_{ST}**	**O_{3+4}**
Numer of lines	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3
Nonsynonymous sites	0	0	1466.21	1466.89	0	0	0	0	0	0	1466.38	1466.67	0	0	2932.58	2933.56
Synonymous and non-coding sites	201	201	459.79	459.11	225	225	1300	1300	219	219	459.63	459.33	175	175	3045.42	3044.44
S (η)	1 (1)	9 (9)	10 (10)	17 (17)	1 (1)	5 (5)	28 (28)	82 (82)	4 (4)	3 (3)	9 (9)	18 (18)	10 (10)	2 (2)	64 (64)	136 (136)
∏	0.0024	0.0298	0.0026	0.0058	0.0022	0.0148	0.0120	0.0420	0.0091	0.0091	0.0025	0.0062	0.0304	0.0076	0.0057	0.0151
± SD	±0.0013	±0.0089	±0.0004	±0.0018	±0.0011	±0.0046	±0.0023	±0.0123	±0.0024	±0.0032	±0.0006	±0.0020	±0.0084	±0.0025	±0.0011	±0.0043
∏(a) (JC)	na	na	0.0006	0.0009	na	na	na	na	na	na	0	0.0009	na	na	0.0003	0.0009
∏(s) (JC)	0.0024	0.0305	0.0087	0.0221	0.0022	0.0149	0.0120	0.0433	0.0091	0.0092	0.0106	0.0236	0.0304	0.0076	0.0096	0.0228
Tajima's *D* (NonSyn)	na	na	-0.7099	na	na	na	na	na	na	na	na	na	na	na	-0.7099	na
Tajima's *D* (Syn)	-0.6123	na	-0.8240	na	-0.6123	na	0.2663	na	-0.7801	na	-0.1536	na	-0.2223	na	-0.4836	na
Fu and Li's *D*	-0.6123	na	-0.8337	na	-0.6123	na	0.2663	na	-0.7801	na	-0.1536	na	-0.2223	na	-0.1720	na
Fu and Li's *F*	-0.4787	na	-0.8337	na	-0.4787	na	0.2791	na	-0.7205	na	-0.1524	na	-0.2223	na	-0.1834	na
Fu and Li's *D* + outgroup	-0.9128	1.6922	-1.8615	-0.2343	-0.9128	-1.3944	0.0323	-0.7309	-1.3803	-1.1456	-0.3123	0.2878	-0.4870	0.4830	-0.7271	-0.4198
Fu and Li's *F* + outgroup	-0.9756	1.6922	-1.9598	-0.2343	-0.9756	-1.3944	0.1129	-0.7309	-1.4661	-1.1456	-0.3136	0.2878	-0.4988	0.4830	-0.7177	-0.4198
Normalized Fay and Wu's FW-Hn	0.6770	-1.6922	1.5250	0.2343	0.6770	1.3944	0.5524	0.7309	1.0635	1.1456	0.3613	-0.2878	0.4960	-0.4830	0.9489	0.4198

Figure 2. Visualizing the distribution of nucleotide diversity across the *Hsp70IR* locus in O_{ST} and $O_{\underline{3+4}}$. Sliding window plots showing polymorphism, π, and divergence, *K*, in our 5978 bp cured MSA that includes the two complete *hsp70* copies and the IGR. **A**: Polymorphism and divergence levels calculated for the O_{ST} arrangement that is here represented by four lines using *D. guanche* as outgroup. **B**: Polymorphism and divergence calculated for the O_{ST} arrangement using *D. madeirensis* as outgroup. **C**: Polymorphism and divergence calculated for the $O_{\underline{3+4}}$ arrangement represented by three non-recombinant lines and using *D. guanche* as outgroup. **D**: Polymorphism and divergence calculated for the $O_{\underline{3+4}}$ arrangement represented by three lines and using *D. madeirensis* as outgroup. As expected, divergence levels are higher when *D. guanche* is used as outgroup instead of *D. madeirensis* that is the sister species of *D. subobscura*. Polymorphism levels are highest in the IGR of $O_{\underline{3+4}}$ and $O_{ST,}$ although they are substantially lower in the latter. The scheme below each window represents the approximate nucleotide position of the seven delimited regions. Window length: 300 nts; step size: 50 nts.




B
Hsp70IR
O_{ST}
π
K
Nucleotide divergence (all sites)
0,08
0,04
0,
0
2000
4000
6000
Nucleotide Position
HSP70_B
5'B
5'A
HSP70_A
3'B
IGR
3'A
D
Hsp70IR
O_{3+4}
π
K
Nucleotide divergence (all sites)
0,08
0,06
0,04
0,02
0,
0
2000
4000
6000
Nucleotide Position
HSP70_B
5'B
5'A
HSP70_A
3'B
IGR
3'A

Gene conversion fuels a significant amount of diversity in the gene pool of an inversion.

Despite the recombination suppression effects of inversions, several GC tracts have been detected in the *rp49* gene region of various O_{ST} and $O_{\underline{3+4}}$ lines, indicating that a limited amount of gene flow between these two arrangements does occur periodically (Rozas and Aguadé 1994; Rozas et al. 1999). To find evidence of gene flow among the 14 lines considered in this work, we used a 5956 bp MSA of the genomic region that comprises the two *hsp70* copies and the nonduplicated segment between them after eliminating all poorly aligned positions with G-blocks (Parameters used: minimum number of sequences for a conserved position: 7; minimum number of sequences for a flanking position: 10; maximum number of contiguous nonconserved positions: 4; minimum length of a block: 10; allowed gap positions: none). Several putative GC tracts obtained statistically significant values in various methods used for detection of recombination tracts, including the Bootscan, GENECONV, 3Seq and MaxChi algorithms, and high scores for the Psi (Ψ) parameter (Betrán et al. 1997) that is commonly used to identify GC tracts from two divergent populations here represented by O_{ST} and $O_{\underline{3+4}}$ (Table 2). Among these, we highlight a ~1.2 kb GC tract in the intergenic region of the O3+4_1 line, originally classified as $O_{\underline{3+4}}$, that was likely transferred from an O_{ST} genotype. Curiously, the remaining genomic regions of O3+4_1 that are presumably descended from an $O_{\underline{3+4}}$ genotype showed the highest similarity to an $O_{\underline{3+4+16}}$ line. In our opinion, this can be possibly explained by the low genetic structure between the $O_{\underline{3+4}}$ and $O_{\underline{3+4+16}}$ lines used in this work as noted in the section below. Although the number of GC tracts in O3+4_1 differ depending on the method used for recombination detection, it is worth noting that they overlap for the most part and cover a genomic area of similar length (~1.2-1.3 kb). Concomitantly, a ~830 bp recombinant tract showing the highest similarity to the only $O_{\underline{3+4+8}}$ line in our data set was identified in the intergenic region of O3+4_5, also an $O_{\underline{3+4}}$ line (Table 6). To our surprise, the outcomes indicate that two out of the five $O_{\underline{3+4}}$ lines contain relatively long GC tracts from a different chromosomal arrangement. Consequently, the O3+4_1 and O3+4_5 lines were excluded from polymorphism analyses concerning $O_{\underline{3+4}}$ and also from the $O_{\underline{3+4}}$ phylad in the case of O3+4_1.

Table 2: Evidence of recombination between different chromosomal arrangements in our data set. Shown are *P*-values for various statistical tests implemented in RDP4 to detect GC tracts in a 5956 bp MSA of the *Hsp70IR* locus, including MaxChi, Bootscan, 3Seq and GENECONV. Two sizeable GC tracts shown in the 'Minor Parent' column that match a line from another chromosomal arrangement were detected in the intergenic region of the O3+4_1 and O3+4_5 lines originally classified as O_{3+4}. Concomitantly, two GC tracts derived from the O_{ST} clade that cover most of the intergenic region of O3+4_1 received high scores in the Ψ (Psi) parameter implemented in DnaSP (Betrán et al. 1997, equation A4). This parameter could not be applied to detect GC tracts between the three arrangements of the O_{3+4} phylad (O_{3+4+8}, O_{3+4} and O_{3+4+16}) because not enough lines were available for each of them.

Recombinant line (arrangement)	MSA length (bps)	Gene conversion tract (length in bps)	Covered genomic region (*Hsp70IR*)	Minor Parent (arrangement)	Major Parent (arrangement)	Method	Av. *P*-Val
O3+4_1 (O3+4)	5956	1845-3050 (1205)	CDS_B, 5'B, IGR	OST_3 (OST)	O3+4+16_2 (O3+4+16)	GENECONV	7.89E-07
O3+4_1 (O3+4)	5956	1845-3050 (1205)	CDS_B, 5'B, IGR	OST_3 (OST)	O3+4+16_2 (O3+4+16)	MaxChi	4.59E-04
O3+4_1 (O3+4)	5956	1845-3050 (1205)	CDS_B, 5'B, IGR	OST_3 (OST)	O3+4+16_2 (O3+4+16)	Bootscan	8.43E-08
O3+4_1 (O3+4)	5956	1845-3050 (1205)	CDS_B, 5'B, IGR	OST_3 (OST)	O3+4+16_2 (O3+4+16)	3Seq	1.43E-16
O3+4_5 (O3+4)	5956	2112-2946 (834)	5'B, IGR	O3+4+8 (O3+4+8)	O3+4_2 (O3+4)	GENECONV	4.55E-03
O3+4_5 (O3+4)	5956	2112-2946 (834)	5'B, IGR	O3+4+8 (O3+4+8)	O3+4_2 (O3+4)	MaxChi	1.93E-04
O3+4_5 (O3+4)	5956	2112-2946 (834)	5'B, IGR	O3+4+8 (O3+4+8)	O3+4_2 (O3+4)	Bootscan	1.42E-04
O3+4_5 (O3+4)	5956	2112-2946 (834)	5'B, IGR	O3+4+8 (O3+4+8)	O3+4_2 (O3+4)	3Seq	2.64E-04
Recombinant line (arrangement)	**MSA length (bps)**	**Gene conversion tract (length in bps)**	**Covered genomic regions (*Hsp70IR*)**	**Population 1 (num. lines)**	**Population 2 (num. lines)**	**Num. informative sites**	**Psi (ψ)**
O3+4_1 (O3+4) (population 1)	5956	**1:** 2233-3192 (960); **2:** 3524-3905 (382).	5'B, IGR, 5'A, CDS_A	O3+4 phylad (8)	OST (4)	73	0.00562

Analyzing the phylogenetic relationship between arrangements and species.

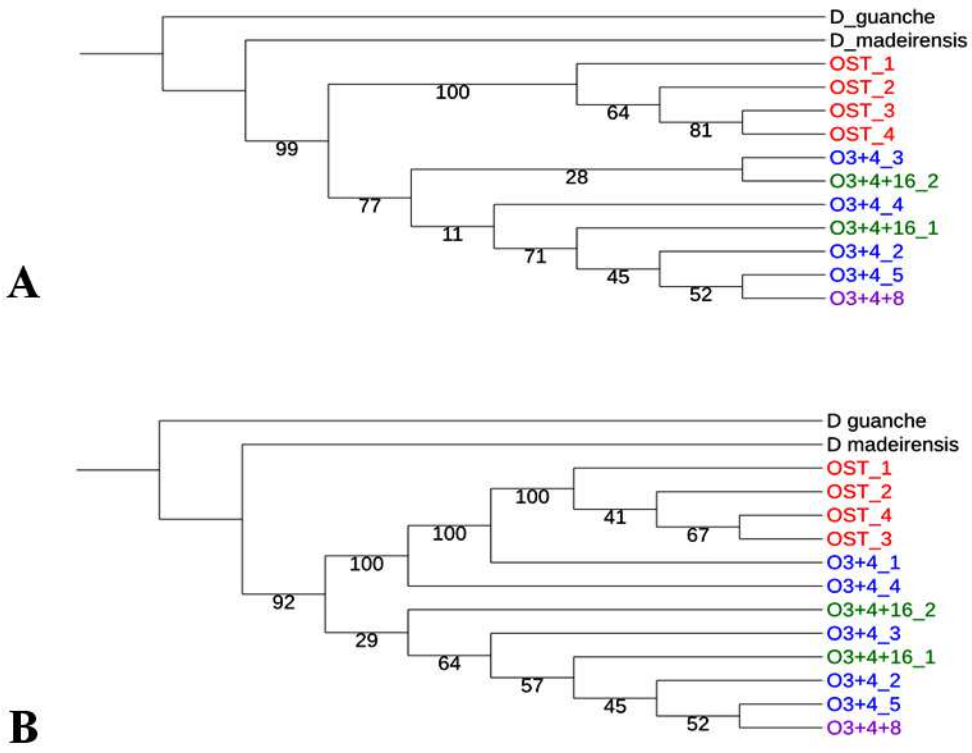
The most widely accepted phylogenetic model postulates that the O_{ST} and O_{3+4} arrangements originated independently from the ancestral O_3 inversion that is now extinct in *D. subobscura* populations but fixed in *D. madeirensis* (Ramos-Onsins et al. 1998). Later on, the O_{3+4} clade gave rise to the O_{3+4+8} and O_{3+4+16} arrangements (; Rozas et al. 1999; Khadem et al. 2001). To explore the phylogenetic relationship between the four considered gene arrangements and the three species of the *subobscura* cluster, we used several measures of population genetic differentiation and phylogenetic analyses to infer the chromosomal inversion ancestry. The extant SNP polymorphism within the *Hsp70IR* locus points to significant genetic differentiation between O_{ST} vs. O_{3+4} and the O_{3+4} phylad as indicated by the permutation test proposed by Hudson et al. (1992) using the Ks^* statistic of population genetic differentiation (P<0.01 after 1000 permutations; Table 3) and the relatively high values for the F_{ST} fixation index (F_{ST} = 0.5860 and 0.6382, respectively; Table 3), which is in agreement with former work using other loci within the inverted segment (Munté et al. 2005; Pegueroles et al. 2013). In contrast, we did not detect any significant genetic structure between O_{3+4} and O_{3+4+16} with the limited number of lines used here. This last outcome is at odds with former work using the

extant polymorphism in the *rp49* gene region, where low but significant F_{ST} values were found between O_{3+4+16} (labeled as O_{3+4+23} in the study) and O_{3+4}, indicating some genetic structure (Rozas et al. 1999). The F_{ST} values between the different arrangements are reflected in a phylogenetic analysis of the complete *Hsp70IR* locus including all *D. subobscura* lines but O3+4_1 (11) and the two outgroup species (Figure 3A). In agreement with accepted phylogenetic relationships, the tree recovers a well-supported OST clade (BS = 100) and moderately supported O3+4 phylad (BS = 77). *D. madeirensis* is placed as an outgroup to all *D. subobscura* lines that form a well-supported cluster (BS = 99). Within the O_{3+4} phylad, we failed to recover a monophyletic O_{3+4+16} clade, which is in line with the reported negative F_{ST} values between this arrangement and O_{3+4} (Table 3). A separate phylogenetic analysis using the same procedures was made including the O3+4_1 intergenotype that contains a 1.2-kb long O_{ST} recombination tract in its intergenic region (Table 2). The inclusion of this line results in an incorrect tree for the inversions considered, where the O3+4_1 and O3+4_4 lines are grouped with the O_{ST} clade with very high support (BS = 100) forming the sister group of a poorly supported O_{3+4} phylad (BS = 29; Figure 3B).

Table 3: Strong genetic differentiation between the O_{ST} and $O_{\underline{3+4}}$ arrangements. Measures of population genetic differentiation have been calculated for different arrangement combinations, including $O_{ST}/O_{\underline{3+4}}$, $O_{ST}/O_{\underline{3+4}}$ phylad and $O_{\underline{3+4}}/O_{\underline{3+4+16}}$. *S*, number of segregating sites; η, total number of mutations; Dxy, average number of nucleotide substitutions per site between populations; Da, number of net nucleotide substitutions per site between populations; F_{ST}, fixation index of population genetic structure. *P* values for the *Ks** statistic (Hudson 1992, equation 11) were obtained after performing a permutation test with 1000 replicates. JC: values with Jukes-Cantor distance correction (Jukes and Cantor 1969). * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; (two-tailed). [§:] Group does not include the recombinant O3+4_1 line.

	Pop 1 = O_{ST} (n=4) *vs.* Pop 2 = $O_{\underline{3+4}}$ (n=3)	**Pop 1 = $O_{\underline{3+4+16}}$ (n=2) *vs.* Pop 2 = $O_{\underline{3+4}}$ (n=3)**	**Pop 1 = $O_{\underline{ST}}$ (n=4) *vs.* Pop 2 = $O_{\underline{3+4}}$ phylad (n=7)[§]**
Number of sites	5978	5978	5978
***S* (η)**	265 (266)	136 (136)	321 (326)
∏ (JC). pop 1	0.0057	0.0089	0.0057
∏ (JC). pop 2	0.0153	0.0153	0.0130
Shared mutations	2	33	10
Fixed mutations	68	6	59
Av. number of nucleotide differences. Kt	109.047	75.200	111.636
Polymorphic mutations (Pop 1)	62	20	54
Polymorphic mutations (Pop 2)	134	103	203
Dxy (JC)	0.0257	0.0120	0.0262
Da (JC)	0.0151	-0.0001	0.0168
F_{ST}	0.5860	-0.0093	0.6382
***P*-value (Ks*)**	0.029*	1	0.002**

Figure 3: Phylogenetic relationships between the taxa used in this work. **A:** Cladogram showing a phylogenetic analysis of 11 *D. subobscura* lines (excluding the recombinant O3+4_1) and the *D. madeirensis* and *D. guanche* lines using a filtered alignment of 5982 sites of the complete *Hsp70IR* gene region. **B**: Cladogram showing a phylogenetic analysis inferred from a filtered alignment of 5956 sites of the complete *Hsp70IR* gene region after including the O3+4_1 intergenotype. Maximum likelihood analyses were performed for phylogenetic tree inference using W-IQ-TREE. Trees were roo ted with the more divergent *D. guanche* species. Branch support was assessed with UFBoot after 1000 bootstrap replicates (see Materials and Methods for more details). Bootstrap values are shown at each node. O_{ST}, O_{3+4}, O_{3+4+16} and O_{3+4+8} lines of *D. subobscura* are, respectively, colored in red, blue, green and purple. The two outgroup species are colored in black.



The two *hsp70* genes evolve at a similar rate in *D. subobscura* but not in *D. guanche*.

The McDonald and Kreitman test (MKT; McDonald and Kreitman 1991) was used to detect signs of positive selection in the coding regions of the two *hsp70* genes in the O_{ST}, O_{3+4}, O_{3+4+16} and arrangements and in the O_{3+4} phylad using their respective orthologs in *D. guanche* as outgroup sequences. The MKT compares the amount of variation within a species to the divergence between species at two types of sites, one of which is putatively neutral (synonymous sites) and used as the reference to detect selection at the other type of site (nonsynonymous sites). No significant *P*-values were observed for any comparison, indicating that polymorphism evolves according to the neutral expectations in the two *hsp70* genes in the considered arrangements (Table 4).

Concomitantly, the Hudson-Kreitman-Aguadé test (HKA; Hudson et al. 1987) was used to compare whether polymorphism and divergence evolve at equivalent rates between the *hsp70A* and *hsp70B* genes in O_{ST} (4 lines), $O_{\underline{3+4}}$ (three lines; excluded: O3+4_1 and O3+4_5 that contain long tracts from a different inversion), in the $O_{\underline{3+4}}$ phylad (7 lines after excluding the recombinant O3+4_1) and in the complete *D. subobscura* data set (12 lines) using *D. guanche* as outgroup. The HKA test is based on assumptions of the SNM (Kimura 1983) and pressumes that polymorphism and divergence evolve at the same rate; differences between these rates at individual loci will be unexpected under the neutral model and will show a significant result. We detected no significant deviation from the neutral model in any comparison, which supports the results of the MKT in that the polymorphism/divergence ratio is similar between the *hsp70A* and *hsp70B* loci in the different *D. subobscura* arrangements (data not shown). Still, we caution on these results because the number of haplotypes used in this work is rather low. In addition, the haplotypes of the inversions considered were captured in different seasons of the same year (see Suplementary table S1), which might introduce some bias in the data considering that the gene pool of certain *D. subobscura* inversions such as O_{ST} has been shown to vary considerably between seasons in the same population (Trelles 2003). That being said, these variables are not expected to alter substantially the polymorphism/divergence ratio when using *D. guanche* as outgroup because it is reproductively isolated from *D. subobscura*. Interestingly, a slightly higher number of fixed differences at synonymous and nonsynonymous sites between *D. subobscura* arrangements and *D. guanche* seem to exist in the *hsp70A* gene while *hsp70B* appears more conservative as previously noted in Puig-Giribets et al. (2018). In fact, when we looked at the total number of fixed nucleotide changes between *D. guanche hsp70A*/*hsp70B* and their orthologs in *D. subobscura*, we found 63 substitutions (12 of which are nonsynonymous) in the first and only 37 (10 of which nonsynonymous) in the latter. This difference is reflected in the synonymous divergence (Ks) (Nei and Gojobori 1986) between the two species that is substantially higher in *hsp70A* than in *hsp70B* (Ks = 0.12860 and 0.08169, respectively). Nonsynonymous divergence (Ka) (Nei and Gojobori 1986) is also higher in *hsp70A* than in *hsp70B*, although the difference is not as marked as in Ks (Ka = 0.00986 and 0.00818, respectively). Thus, it seems that most nucleotide substitutions have accumulated in the *hsp70A copy* after *D. guanche* and *D. subobscura* split. The reasons for this are unclear and more *D. guanche* haplotypes are needed to confirm this trend.

Table 4: No evidence of selection in the coding region of the two *hsp70* genes in the different arrangements. The MKT test was performed on the synonymous and nonsynonymous polymorphism of the complete 1926 bp long *hsp70A* and *hsp70B* coding regions separately in the O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4}}$ phylad lines. The *D. guanche hsp70A* and *hsp70B* genes were used as outgroup sequences to their corresponding orthologs in the different *D. subobscura* arrangements to calculate the number of fixed and polymorphic sites in the two types of polymorphism. *P*-values for the Fisher's exact test are shown in those comparisons where the *G*-test could not be computed. No significant *P*-values were observed for any statistical test.

hsp70A	**Fixed**	**Polymorphic**		***hsp70B***	**Fixed**	**Polymorphic**	
O_{ST} (n = 4)			*G* = na	**O_{ST} (n = 4)**			*G* = 0.1950
Synonymous	54	9	*NI* = 0	Synonymous	36	8	*NI* = 0.6920
Nonsynonymous	15	0	*Pf* = 0.1933$^{\neq}$	Nonsynonymous	13	2	*P* = 0.6589
$O_{\underline{3+4}}$ (n = 3)			*G* = 0.6900	**$O_{\underline{3+4}}$ (n = 3)**			*G* = 1.4750
Synonymous	55	16	*NI* = 0.5290	Synonymous	29	15	*NI* = 0.3870
Nonsynonymous	13	2	*P* = 0.4061	Nonsynonymous	10	2	*P* = 0.2245
$O_{\underline{3+4+16}}$ (n = 2)			*G* = na	**$O_{\underline{3+4+16}}$ (n = 2)**			*G* = na
Synonymous	54	11	*NI* = 0	Synonymous	32	9	*NI* = 0
Nonsynonymous	13	0	*Pf* = 0.1947$^{\neq}$	Nonsynonymous	11	0	*Pf* = 0.1765$^{\neq}$
$O_{\underline{3+4}}$ phylad (n = 8)			*G* = 1.5140	**$O_{\underline{3+4}}$ phylad (n = 8)**			*G* = 2.2680
Synonymous	52	33	*NI* = 0.4850	Synonymous	28	34	*NI* = 0.4120
Nonsynonymous	13	4	*P* = 0.2185	Nonsynonymous	10	5	*P* = 0.1321

N = number of lines. NI = neutrality index. $^{\neq}$*Pf*-value for the Fisher's exact test (two-tailed). Na = the *G*-test could not be applied.

Estimates of coalescence (*Tc*) times differ greatly between chromosomal arrangements.

Assuming a monophyletic origin of inversions (Powell 1997), the different O_{ST} and $O_{\underline{3+4}}$ haplotypes can be used to estimate a coalescence time (*Tc*). To calculate *Tc*, we chose the synonymous polymorphism $\prod_S$ that is supposed to evolve close to the expectations of the neutral theory. Coalescence times (*Tc*) were calculated for the two *hsp70* genes separately in the O_{ST} and $O_{\underline{3+4}}$ arrangements using the formula $Tc = \theta \sum_{i=2}^{n} (1/\, i(i-1))/\mu$ deduced from Hudson (1990). The formula assumes equilibrium conditions in the population, where Watterson's θ estimator *theta* represents the nucleotide variability in the synonymous sites at this region, i equals to number of lines per inversion. We chose a mutation rate of $\mu = 2.8^{-09}$ (95% confidence interval (CI) = 1.0^{-09} - 6.1^{-09}) mutations per site per generation estimated for *D. melanogaster* (Keightley et al. 2014). Although $O_{\underline{3+4}}$ and O_{ST} show strong latitudinal and seasonal components that indicate that these inversion populations are not at equilibrium, we detected no departure from neutrality using the extant polymorphism in the *Hsp70IR*

locus (Table 1 and Table S3). It is possible that selection is operating at loci other than *Hsp70IR* and/or that the number of samples used here is too low to detect evidence of selection. In any case, we considered pertinent to make the estimates to have an idea of how closely related the different haplotypes are, and to compare them between the two *hsp70* paralogous genes. Note that coalescent times are proportional to the variability level. Unsurprisingly, our *Tc* estimates are comparable between the two *hsp70* genes within the same inversion with some variations probably explained by the high stochasticity in our small data set (Table 5). As expected, younger coalescence times were observed for the O_{ST} inversion (0.25-0.28 Myr) that are only about half to one third of those of $O_{\underline{3+4}}$ (0.55-0.51 or 0.63-0.66 Myr, see Table 5) assuming a generation time of 24 days and 8 reproductively active months that equal to ~10 generations per year in the considered arrangements. A possibility to explain the discrepancy is that our figures for $O_{\underline{3+4}}$ are slightly increased due to the presence of undetected GC tracts from different chromosomal arrangements as noted in Rozas et al. (1999). Another possibility is that, although they were captured in the same population and year (see Supplementary table S1), the $O_{\underline{3+4}}$ haplotypes have a more heterogeneous origin relative to their O_{ST} counterparts. This would not come as unexpected considering that our *D. subobscura* inversion populations are substructured in time, and that there is evidence showing that their genetic content varies from one season to another in the same population (Trelles 2003). Taking this into account, the nucleotide diversity and the coalescence times deduced from the few haplotypes used in this work might not be representative of the inversions considered. To confirm or refute our outcomes, an analysis of a higher number of haplotypes captured in different seasons of the year in the same population is warranted.

Table 5: Coalescence times inferred from the synonymous polymorphism indicate a younger origin for O_{ST}. Coalescence times (*Tc*) for the different chromosomal arrangements using the nucleotide variation in the synonymous sites, $\theta_{(SYN)}$, were calculated with the equation $Tc = \theta \sum_{i=2}^{n} (1/\, i(i-1))/\mu$ deduced by Hudson (1990) assuming equilibrium conditions in the population and the same mutation rate per generation of $\mu = 2.8 \times 10^{-9}$(95% CI = 1.0^{-09} - 6.1^{-09}) for the different gene arrangements. No estimates are available for the $O_{\underline{3+4+8}}$ arrangement that is represented by a single sequence. Times are given in millions of generations (Mg) and in millions of years (Myr). To make a rough estimation of the latter, we considered a generation time of 24 days under standard developmental conditions (Krimbas 1993) and 8 reproductively active months per year that represent ~10 generations (*D. subobscura* flies are inactive in winter-like conditions, which in Iberia typically last from November to February).The 95% confidence interval (CI) of *Tc* (Mg) and *Tc* (Myr) estimates has been inferred after computing the standard deviation (SD) from the number of pairwise nucleotide differences between lines of the same arrangement. #: To obtain more reliable estimates of coalescence times, the O3+4_5 line has been included in the $O_{\underline{3+4}}$ data set because no gene conversion tracts from a different arrangement have been detected in its parlaogous *hsp70* CDSs (see Table 2 for further information).

hsp70A	**O_{ST}**	**$O_{\underline{3+4}}$**
Number of alleles (n)	4	3 (4)#
$\theta_{(SYN)}$	0.0106	0.0232 (0.0237)#
Mutation rate per generation	$2.8E^{-09}$	$2.8E^{-09}$
sum_i	0.75	0.667 (0.75)#
Tc (Mg)	0.284±0.113 (95% CI)	0.552±0.281 (95% CI) 0.634±0.233# (95% CI)
Tc (10g/y) (Myr)	0.028±0.011 (95% CI)	0.055±0.028 (95% CI) 0.063±0.023# (95% CI)
hsp70B	**O_{ST}**	**$O_{\underline{3+4}}$**
Number of alleles (n)	4	3 (4)#
$\theta_{(SYN)}$	0.0094	0.0217 (0.0249)#
Mutation rate per generation	$2.8E^{-09}$	$2.8E^{-09}$
sum_i	0.75	0.667 (0.75)#
Tc (Mg)	0.251±0.089 (95% CI)	0.516±0.160 (95% CI) 0.666±0.164# (95% CI)
Tc (10g/y) (Myr)	0.025±0.008 (95% CI)	0.051±0.016 (95% CI) 0.066±0.016# (95% CI)

Abundant INDEL polymorphism characterizes the intergenic, non-duplicated region (IGR) that is unaffected by gene conversion.

In contrast with the intronless *hsp70* CDSs, numerous insertion-deletion (INDEL) polymorphisms exist in the regions surrounding them that makes sequence alignment difficult. To obtain some estimates on such polymorphism, we chose the Multiallelic Model that considers all INDELs (Librado and Rozas 2009). Five genomic regions at the *Hsp70IR* locus (including the 5'and 3'UTRs and the IGR) were separately analyzed in O_{ST}, $O_{\underline{3+4}}$ and in the $O_{\underline{3+4}}$ phylad. INDEL polymorphism estimates were obtained

using two types of procedures. In the first case, we performed no curation after a MSA with the MAFFT E-INS-i algorithm that is recommended for sequences that contain long gaps to preserve all INDELs (see Materials and Methods). Unfortunately, the quality of this MSA was likely far from optimal due to uncertain positional homology (results in Table 6.1). To minimize this problem, the initial MSA was cured with Gblocks using the less stringent criteria (see Materials and Methods) to retain only the more conserved INDELs and improve the reliability of our MSA (results in Table 6.2). Our results concerning the unfiltered MSA indicate that INDELs are relatively more abundant and longer in the nonduplicated IGR of the $O_{\underline{3+4}}$ lines than in its neighboring 5'UTRs that undergo GC (Table 6.1). This is somehow expected considering that the IGR, being a nonfunctional region, is probably the least constrained among the five considered. In addition, the IGR shows the highest number of INDEL haplotypes and INDEL haplotype diversity per site (∏i), which parallels SNP polymorphism levels that are also highest in this region in the different gene arrangements (Table 1 and Supplementary table S3). The relatively inflated polymorphism levels in the IGR that are even higher than those of the synonymous sites can perhaps relate to the numerous non-homologous repetitive motifs that exist in this region. Also in agreement with estimates on SNP polymorphism, we notice that the total number of INDEL sites and ∏i tend to be lower in O_{ST} than in $O_{\underline{3+4}}$. The discrepancies between the two arrangements are most pronounced in the IGR that contains the highest number of INDELs and INDEL variability, and probably relate to differences in the nucleotide diversity levels between the two arrangements, being lower in O_{ST} (Table 1). Similar outcomes are observed after analyzing the filtered alignment (Table 6.2), in which a significant part of INDELs have been removed by the Gblocks algorithm. Curiously, in the filtered and unfiltered MSA we observe that INDELs are significantly more numerous and longer in the 3'A than in its paralogous 3'B region, which is practically depleted of such polymorphism. This is a worth mentioning difference considering that the two regions are presumably involved in transcript regulation. In light of these differences, a more rigurous and detailed analysis of INDEL polymorphism is warranted to corroborate our findings, and to identify putatively functional INDELs that might exist in the 5' and 3' UTRs.

Tables 6.1-6.2. INDEL polymorphism is most abundant in the intergenic nonduplicated region (IGR). Various parameters of INDEL polymorphism have been calculated in the five delimited regions of the *Hsp70IR* locus where such polymorphism exists in the O_{ST}, $O_{\underline{3+4}}$ arrangements and in the $O_{\underline{3+4}}$ phylad (phy) separately with the Multiallelic Model (Librado and Rozas 2009). **3'B**: 3'UTR of the *hsp70B* copy. **5'B**: 5'UTR of the *hsp70B* copy. **IGR**: intergenic nonduplicated region. **5'A**: 5'UTR of the *hsp70A* copy. **3'A**: 3'UTR of the *hsp70A* copy. Na= not applicable (the Tajima's *D* test requires a minimum of four sequences). [§:] Group does not include the recombinant O3+4_1 line. Results are presented for the unfiltered MSA (Table 6.1) and after filtering the MSA with Gblocks to remove the INDELs that are less conserved (Table 6.2).

6.1	3'B			5'B			IGR			5'A			3'A		
Number of sites	203			543			1425			547			212		
Arrangement	OST	O3+4	O3+4 phy[§]	OST	O3+4	O3+4 phy[§]	OST	O3+4	O3+4 phy[§]	OST	O3+4	O3+4 phy[§]	OST	O3+4	O3+4 phy[§]
Number of sequences	4	3	7	4	3	7	4	3	7	4	3	7	4	3	7
INDEL sites analyzed (total)	4	0	0	9	11	24	33	187	451	2	15	24	8	4	40
INDEL events analyzed (total)	2	0	0	4	5	9	9	23	36	2	7	10	2	1	5
Average INDEL length (bp)	2	na	na	1.625	2.600	4.250	3.941	13.706	21.758	1	2.889	2.840	12	4	7.273
Number of INDEL haplotypes	3	na	na	3	3	6	4	3	7	2	3	6	2	2	5
INDEL haplotype diversity	0.8330	na	na	0.8330	1	0.9520	1	1	1	0.5	1	0.9520	0.5	0.6670	0.9050
InDel diversity per site. ∏(i)	0.0049	na	na	0.0037	0.0061	0.0052	0.0046	0.0132	0.0094	0.0018	0.0085	0.0059	0.0049	0.0031	0.0094
Tajima's *D*	-0.7099	na	na	-0.7801	na	-1.1810	-0.1536	na	-0.5683	-0.7099	na	-1.1089	-0.7099	na	-0.0990

6.2	3'B			5'B			IGR			5'A			3'A		
Number of sites	203			526			1020			525			202		
Arrangement	OST	O3+4	O3+4 phy[§]	OST	O3+4	O3+4 phy[§]	OST	O3+4	O3+4 phy[§]	OST	O3+4	O3+4 phy[§]	OST	O3+4	O3+4 phy[§]
Number of sequences	4	3	7	4	3	7	4	3	7	4	3	7	4	3	7
INDEL sites analyzed (total)	4	0	0	7	3	10	12	44	70	1	4	5	8	0	30
INDEL events analyzed (total)	2	0	0	2	3	5	3	9	12	1	4	5	1	0	4
Average INDEL length (bp)	2	na	na	3.5	1	2	3.250	5.571	4.964	1	1	1	8	na	7
Number of INDEL haplotypes	3	na	na	2	2	4	4	3	7	2	2	3	2	na	4
INDEL haplotype diversity	0.8330	na	na	0.5	0.6670	0.7140	0.8330	1	1	0.5	0.6670	0.5240	0.5	na	0.8100
InDel diversity per site. ∏(i)	0.0049	na	na	0.0019	0.0038	0.0027	0.0016	0.0059	0.0043	0.0009	0.0050	0.0027	0.0024	na	0.0075
Tajima's *D*	0.7099	na	na	-0.7099	na	-1.4861	0.1676	na	-0.5747	-0.6123	na	-1.4861	-0.6123	na	-0.3187

Gene conversion is restricted to the coding regions and their promoters.

In agreement with Puig-Giribets et al. (2018), our estimates of nucleotide diversity in the *Hsp70IR* locus indicate that in some *D. subobscura* arrangements such as O_{ST} and $O_{\underline{3+4}}$, the extant polymorphism in the CDS and 5'UTRs is more similar between paralogs of the same arrangement than among orthologs of different arrangements because of GC (Table 7.1). In contrast, the 3'UTRs are significantly more similar at the nucleotide level between orthologs of the same duplicate, which suggests that these regions do not urdergo GC (Table 7.2). To visualize the effects of GC on the sequence similarity between *hsp70* paralogs, we made a phylogenetic analysis with W-IQ-TREE for the paralogous 5'UTR and CDS regions and another for the more divergent 3'UTRs in the 14 lines using a MSA filtered with Guidance2 and default settings (see Materials and Methods). The base MSA of the paralogous 5'UTR and CDS regions received a relatively high confidence score (0.992) while moderately low scores were obtained for the MSA of the more divergent 3'UTRs (0.645). To improve the quality of the latter, we used G-blocks with either less or more stringent parameters as an alternative MSA curating method. Similar phylogenetic trees were obtained after using any of these procedures, indicating that the conclusions derived from the original MSA were robust to the use of G-blocks. As expected, the 5'UTR and CDS paralogous sequences grouped according to the arrangement/species they belonged to (Figure 4A). However, in the MSA of the short (< 200 bp) 3'UTRs that apparently do not undergo GC we observed that the paralogous sequences no longer cluster together (Figure 4B) but consistently cluster (100% bootstrap) within homologous copies, confirming the differential pattern in both regions and supporting the GC effect on 5'UTR and CDS regions. In addition to being more similar within the same arrangement, the CDSs and 5'UTRs of the two *hsp70* copies show in some cases an increased similarity within the same line, indicating that multiple independent GC events might have taken place since the origin of the four arrangements analyzed in this work (Supplementary tables S4.1-S4.4). This can also be appreciated in Figure 4A, where some paralogous duplicates of the same line form well-supported clusters (BS $\geq$ 95). Interestingly, the effects of GC are also obvious in the nonsynonymous variation, as evidenced by the predicted amino acid (aa) sequences from the two *hsp70* genes in the 12 *D. subobscura* lines. An alignment of all homologous *hsp70* aa sequences in our data set indicates that only seven positions are polymorphic in *D. subobscura*, and that in two of them the rarest amino acid only occurs in the two *hsp70* genes of the same line, which is probably the result of GC (one

in O3+4_4 and one in O3+4+8; Supplementary figures S4.1-S4.3). Perhaps unexpectedly, four polymorphic sites contain amino acids that are fixed between *hsp70* homologs in *D. subobscura/D. madeirensis* and the two *D. guanche hsp70* genes (Supplementary figures S4.1-S4.3) that presumably do not undergo GC (Puig-Giribets et al. 2018). The pattern described so far suggests that the region affected by GC in *D. subobscura* is somehow delimited. However, although there is a sharp increase in the divergence between the two *hsp70* genes of the same line towards the last ~100 bp of their CDSs, we have not been able to identify any candidate boundary sequence. Considering that the rate of GC events is positively linked with high (>80%) sequence identity (Mansai et al. 2011), it is possible that GC mechanisms act efficiently through the highly conserved CDSs and their proximal promoter regions, while they rapidly decay upon entering in the much more divergent 3'UTRs and beyond the first ~250 bp upstream the start codon that contain several conserved *cis*-regulatory elements.

Tables 7.1-7.2. Higher nucleotide divergence between orthologous than paralogous sequences as a result of GC. Table 7.1. Polymorphism and divergence levels between *hsp70* orthologous regions in different gene arrangements, including the 5'UTRs, CDSs and 3'UTRs. Π(a), nonsynonymous nucleotide diversity. Π(s), synonymous nucleotide diversity. K(a) nonsynonymous nucleotide divergence. K(s) synonymous nucleotide divergence. **Table 7.2.** Polymorphism and divergence levels between *hsp70* paralogous regions of the same arrangement, including the 5'UTRs, CDSs and 3'UTRs. Π(a), nonsynonymous nucleotide diversity; Π(s), synonymous nucleotide diversity. K(a) nonsynonymous nucleotide divergence. K(s) synonymous nucleotide divergence. F_{ST}, fixation index of population genetic structure. JC: values with Jukes-Cantor distance correction (Jukes and Cantor 1969).*: O_{3+4} phylad (including the O_{3+4} (5), O_{3+4+16} (2) and O_{3+4+8} (1) lines).

Table 7.1	***hsp70A***						***hsp70B***					
Genomic region	**3'UTR**		**CDS**		**5'UTR**		**3'UTR**		**CDS**		**5'UTR**	
Number of sites	143		1926		461		143		1926		461	
Arrangements	OST / O3+4	OST / O3+4*	OST / O3+4	OST / O3+4*	OST / O3+4	OST / O3+4 *	OST / O3+4	OST / O3+4 *	OST / O3+4	OST / O3+4 *	OST / O3+4	OST / O3+4 *
∏ (a) OST (JC)	0	0	0	0	0	0	0	0	0.0006	0.0006	0	0
∏ (s) OST (JC)	0.0260	0.0261	0.0105	0.0105	0.0109	0.0109	0.0035	0.0035	0.0087	0.0087	0.0120	0.0120
∏ (a) O3+4* (JC)	0	0	0.0009	0.0006	0	0	0	0	0.0009	0.0009	0	0
∏ (s) O3+4* (JC)	0.0093	0.0170	0.0235	0.0230	0.0279	0.0255	0.0382	0.0260	0.0220	0.0255	0.0205	0.0204
K (a) (JC)	0	0	0.0018	0.0017	0	0	0	0	0.0021	0.0022	0	0
K (s) (JC)	0.0543	0.0488	0.0450	0.0455	0.0416	0.0393	0.0206	0.0178	0.0402	0.0448	0.0391	0.0379

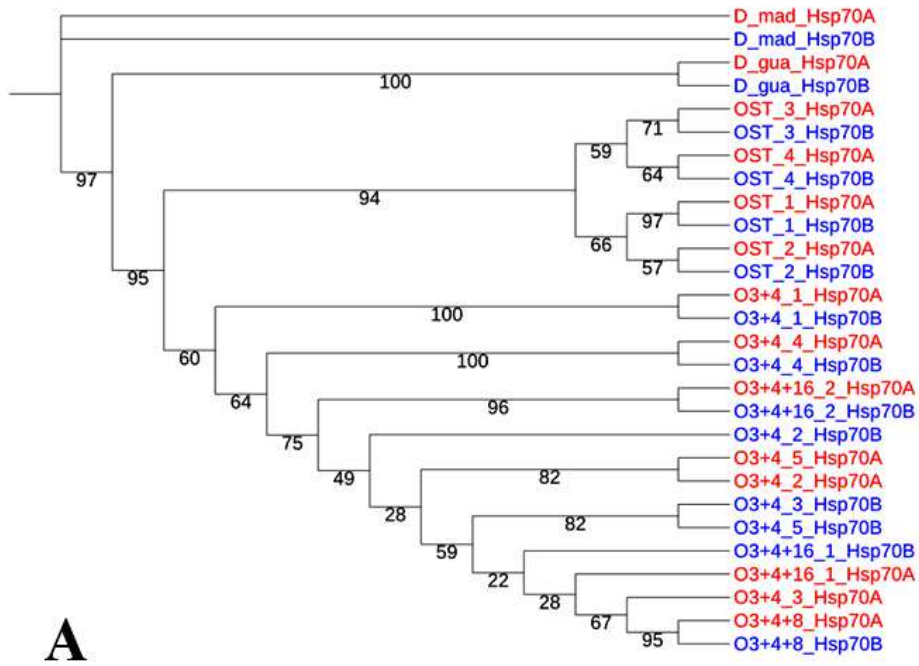
Table 7.2	*hsp70A / hsp70B*								
Genomic region	**3'UTR**			**CDS**			**5'UTR**		
Number of sites	143			1926			461		
Arrangements	OST	O3+4	O3+4*	OST	O3+4	O3+4*	OST	O3+4	O3+4*
Number of sequences	4	3	8	4	3	8	4	3	8
∏ (a) A (JC)	0	0	0	0	0.0009	0.0006	0	0	0
∏(s) A (JC)	0.0261	0.0093	0.0169	0.0105	0.0235	0.0233	0.0109	0.0279	0.0254
∏ (a) B (JC)	0	0	0	0.0006	0.0009	0.0009	0	0	0
∏(s) B (JC)	0.0035	0.0382	0.0259	0.0087	0.0221	0.0260	0.0120	0.0205	0.0202
K (a) AB (JC)	0	0	0	0.0003	0.0006	0.0007	0	0	0
K (s) AB (JC)	0.2645	0.2624	0.2658	0.0095	0.0193	0.0229	0.0092	0.0180	0.0218
F_{ST}	0.9346	0.8947	0.9056	-0.0085	-0.2069	-0.0760	-0.2352	-0.3378	-0.0473

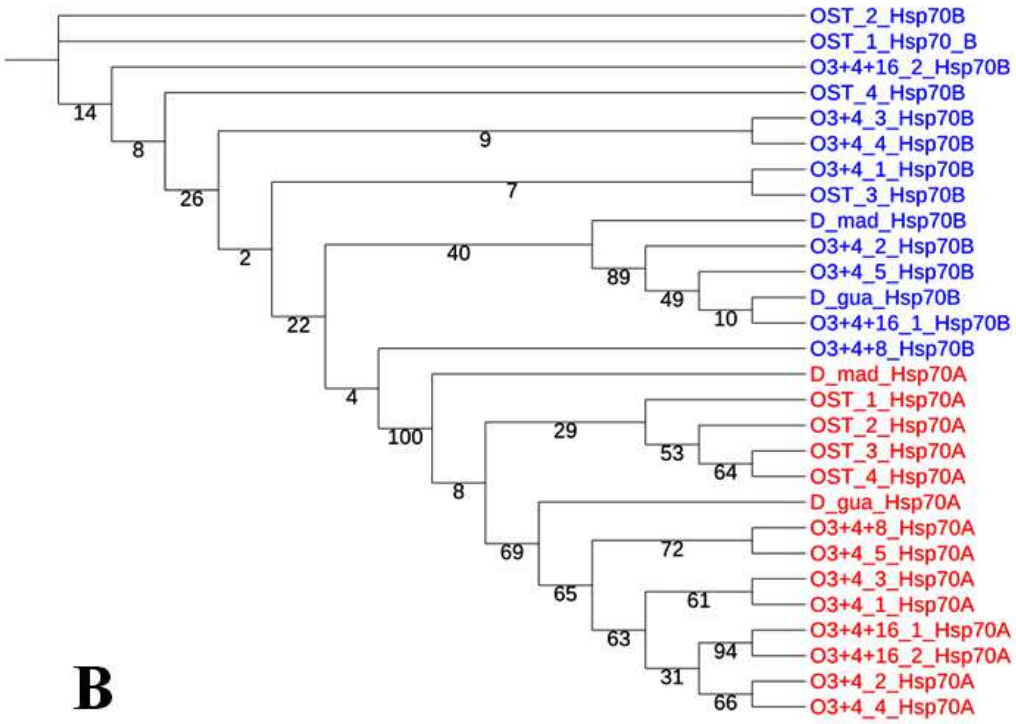
Quantifying the effects of GC on the similarity between *hsp70* paralogs in the different arrangements.

In the small data set used by Puig-Giribets et al. (2018), the similarity between the two *hsp70* paralogs was found to be increased in O_{ST} relative to the $O_{\underline{3+4}}$ phylad possibly as a result of higher GC rates in the former. In addition, the physical distance between the two *hsp70* paralogs was noted to be shorter in O_{ST}, which might have consequences on the rate of GC events in this arrangement. To further explore this fascinating question, we took advantage of our limited albeit expanded data set to quantify and study these polymorphisms in the different gene arrangements (Table S6.1). In agreement with Puig-Giribetset al. (2018), we found that the average physical distance separating the two *hsp70* CDSs is reduced in the O_{ST} arrangement by about 66-122 bps ($\bar{x}$: 2134 bps in $O_{\underline{3+4}}$, $\bar{x}$: 2192 bp in the $O_{\underline{3+4}}$ phylad and $\bar{x}$: 2069 bp in $O_{ST;}$ Supplementary table S5), the differences being statistically significant (Mann–Whitney-U test, $p < 0.05$; Supplementary figure S2). In addition, the variability in length is higher in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad than O_{ST}, which is congruent with lower diversity levels through *Hsp70IR* in the latter (SD: 24.73 and 107.21 vs. 4.34 bps, respectively, Table S5). Next, we calculated the median values of silent divergence (K_{SIL}, which includes all synonymous sites plus the stop codon, althoug in the latter only a small part of mutations are synonymous) between the two *hsp70* duplicates in the O_{ST} arrangement (here represented by four lines) and in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad (represented by five and eight lines, respectively). We noticed that median K_{SIL} values are slightly higher in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad than in O_{ST} (K_{SIL} = 0.0109/0.0120 vs. 0.0076, respectively) with the two former arrangements showing a

higher variability as reflected by standard deviation (SD: 0.0084/0.0095 vs. 0.0038, respectively). The difference in K_{SIL} values between O_{ST} and O_{3+4}/O_{3+4} phylad, however, was not found to be significant using the parametric t-Test for independent samples (p = 0.437 and p = 0.378, respectively; Supplementary tables S6.1-S6.2). Thus, although higher levels of compactness are observed in O_{ST} than in O_{3+4}/O_{3+4} phylad, the role that physical distance exerts on the nucleotide identity between *hsp70* paralogs remains elusive. A higher number of lines per arrangement are needed to add more light on this issue.

Figures 4A-4B. Evidence of concerted evolution in the CDSs and 5'UTRs of the two *hsp70* duplicates. 4A: Unrooted cladogram showing a phylogenetic analysis of a 2390 bp filtered alignment including the 5'UTR and CDSs regions of both *hsp70A* and *hsp70B* palindromic copies in the 14 lines. **4B**: Unrooted cladogram showing a phylogenetic analysis of a 140 bp filtered alignment of the *hsp70A* and *hsp70B* palindromic 3'UTRs in the 14 lines. Maximum likelihood analyses were performed for phylogenetic tree inference using W-IQ-TREE. Branch support was assessed with UFBoot after 1000 bootstrap replicates (see Materials and Methods for more details). Bootstrap values are shown at each node. The *hsp70A* orthologs in the 14 lines have been colored in red while their *hsp70B* counterparts are shown in blue. D_mad: *D. madeirensis.* D_gua: *D. guanche.* Notice that the tree topology differs markedly between the 5'UTRs/CDSs and their adjacent 3'UTRs because the latter do not undergo GC.





3.3.1.5 Discussion

The 10 kb reconstructed *Hsp70IR* locus is composed of several distinctive DNA regions.

The preserved synteny among the three species of the *subobscura* cluster within the ~10 kb *Hsp70IR* locus contrasts with the distribution of nucleotide polymorphism levels that vary sharply between the seven delimited regions within it. Amongst the most salient features, we highlight the high degree of nucleotide conservation in the coding regions of the two *hsp70* genes and their proximal promoters that in the case of *D. subobscura* and *D. madeirensis* is enhanced by purifying selection and periodic intrachromosomal GC events. No signs of selection are detected in the nucleotide diversity of any of the seven delimited regions within *Hsp70IR* that appear to evolve according to the expectations of the neutral theory in the O_{ST}, $O_{\underline{3+4}}$ and the $O_{\underline{3+4}}$ phylad. In addition, the two *hsp70* paralogous genes appear to evolve at a similar rate in the different arrangements as indicated by the HKA test. These outcomes suggest that the extant

polymorphism at *Hsp70IR* might not explain the adaptive character of some arrangements, including O_{ST} and $O_{\underline{3+4}}$. In line with this observation, it is worth mentioning that a quantification of basal Hsp70 protein levels in several homozygous lines for the warm climate-associated $O_{\underline{3+4}}$ and the cold climate-associated O_{ST} did not reveal increased levels in the former (our unpublished data), which calls into question the outcomes of Calabria et al. (2012). Unsurprisingly, a relatively increased number of nonshared SNPs in nonsynonymous sites as evidenced by negative Tajima's *D* values suggests that the two paralogous *hsp70* CDSs are under the effects of purifying selection. Polymorphism levels are highest in the IGR separating the two *hsp70* genes that is probably the least constrained of the seven regions. The numerous INDEL polymorphism in this region is in sharp contrast with the neighboring *hsp70* paralogs that in the considered lineages consist in a single exon and are almost indistinguishable from each other as a result of recurrent GC events. As typical of other heat-inducible loci, *hsp70* genes tend to lack introns to eliminate splicing, a mechanism that is temperature-sensitive and time-consuming (Feder and Krebs 1998). The strong and fascinating disparity in the quantity and quality of polymorphism between such physically close regions is explained on the one hand by the essential and numerous cellular roles of the ancient *hsp70* gene family that is conserved from bacteria to man and, on the other hand, by the rapid divergence of a non-functional region that varies noticeably between our closely related lines.

A possible link between rich INDEL polymorphism and repetitive DNA in the IGR.

The length of the complete intergenic region including the IGR and the proximal promoter regions varies notably among the 12 closely related *D. subobscura* lines, ranging from 2062 to 2423 bps (Supplementary table S5). These discrepancies are explained by INDEL polymorphism that is more numerous in the IGR than in other more constrained regions (Table 6). Of all INDEL polymorphisms, we highlight a 224 bp duplication that occurs uniquely in the IGR of the O3+4+16_2 line (data not shown). It is known that transposable elements (TEs) may generate duplications of the same genetic material via ectopic recombination between two members of the same family (Kidwell and Lisch 2001). In fact, some authors have proposed that the variation in copy number and other gross rearrangements within *hsp70* gene clusters of certain dipteran lineages could be due to unequal homologous recombination between

duplicated TE sequences (Evgen'ev et al. 2004; Garbuz et al. 2011) and retrotransposition events (Bettencourt and Feder 2001). Thus, TEs appear to have played an important role in the evolution and organization of the *hsp70* gene family in several *Drosophila* lineages. To explore the role of TEs in the configuration of the *Hsp70IR* locus in the *subobscura* species cluster, we searched for traces of repetitive DNA sequences in the complete reconstructed region in our 14 lines using the CENSOR software with default values and hosted in the GIRI web server (Kohany et al. 2006; available at: http://www.girinst.org/repbase/). Interestingly, numerous ~50-200 bp long segments distributed across the *Hsp70IR* locus in the 14 lines matched repetitive DNA sequences of diverse TE families, although the alignment scores were relatively low in most cases (data not shown). The highest scores corresponded to the first ~200 bp of the *hsp70* CDS, a segment that exhibited similarity to the *Copia* retroelement family. Curiously, *Copia* retroelements have shown to be responsive to environmental stress, and sequences similar to those of heat shock gene promoters have been described in their LTRs (Strand and Macdonald 1985; Pietzenuk et al. 2016). Aside from contributing to copy number variation, INDEL polymorphism and genome restructuring, TE insertions in the heat shock promoters of certain species have been linked with changes in phenotypic traits including gene expression, thermal resistance and female fecundity (Lerman and Feder 2004; Chen at al. 2008). However, we could not identify the presence of any novel TE insertion in the highly preserved *hsp70* proximal promoter regions of the 14 lines, which in our opinion argues against drastic changes in their gene expression regulation.

Unexpectedly low polymorphism levels are observed in the O_{ST} inversion relative to O_{3+4}.

Significantly lower polymorphism levels are observed across the *Hsp70IR* locus in our O_{ST} data set when compared with its O_{3+4}/O_{3+4} phylad counterparts. Polymorphism levels in the IGR of O_{ST} are ½ to ⅓ of those in O_{3+4}, a trend also observed for the synonymous polymorphism of both *hsp70A* and *hsp70B* copies (Table 1 and Table 5). A somewhat lower nucleotide diversity of O_{ST} relative to that of O_{3+4} has been noted in some studies of loci distributed along the inverted segment, including *Acph-1* (Navarro-Sabaté et al. 1999) and *rp49* (Rozas and Aguadé 1994; Rozas et al. 1999), while comparable levels of diversity have been observed in others such as *Fmr1* (Pegueroles et al. 2013). In contrast with these studies, our *Tc* estimates suggest that the time to the

common ancestor of our O_{ST} haplotypes is only half of that of $O_{\underline{3+4}}$, indicating a much lower effective population size for the former arrangement. It is possible that the lower diversity levels in O_{ST} are related to the more recent origin of this inversion relative to $O_{\underline{3+4}}$. However, the fact that no deviation from neutrality is observed across *Hsp70IR* in any of the two arrangements indicates that these are no longer in the rapid population expansion phase that occurs after the origin of an inversion that leaves a footprint in the nucleotide diversity patterns that can be used to infer its age. Consequently, other scenarios must be considered to explain the reduced effective population size of O_{ST}. A possibility is that a selective sweep at the *Hsp70IR* locus has trimmed the nucleotide diversity levels of this arrangement. Interestingly, selection has been proposed as a mechanism to explain the seasonal fluctuations of allelic frequencies of loci located within O_{ST} (Fontdevila et al. 1983) and the low diversity values displayed by certain microsatellite loci within this inversion (Calabria 2012). Nonetheless, the nucleotide polymorphism at the *Hsp70IR* locus fits better into a scenario of weak negative selection on coding DNA as evidenced by negative Tajima's *D* values, where new mutations in nonsynonymous sites persist as polymorphic because they are slightly deleterious and cannot therefore become fixed. Considering that the nucleotide diversity patterns across *Hsp70IR* are similar in O_{ST} and the three other considered arrangements, it is likely that the lower effective population size of O_{ST} is due to stochastic changes in inversion frequencies in the Berbikiz population that have resulted in an increase of the effective population size of some arrangements like $O_{\underline{3+4}}$ at the expense of others such as O_{ST}. It is possible that these changes are not stochastic but rather dictated by environmental factors such as temperature (Rodríguez-Trelles et al. 1996; Rodríguez-Trelles et al. 2013). In any case, our results indicate that *Hsp70IR* does not seem to count among the loci that are under natural selection. Alternatively, the extant polymorphism levels in our $O_{\underline{3+4}}$ lines could be somewhat inflated as a result of undetected GC tracts from different arrangements as indicated in Rozas et al. (1999) or by more diverse origins of our $O_{\underline{3+4}}$ haplotypes relative to the O_{ST} ones. The latter is supported by evidence showing that the gene pool of certain *D. subobscura* inversions like O_{ST} varies significantly along the year in the same population (Trelles 2003), which indicates that foreign alleles are periodically introduced in the population, and that there is some degree of genetic structure between different populations. An analysis of the nucleotide diversity at *Hsp70IR* using a higher number of O_{ST} and $O_{\underline{3+4}}$ haplotypes from unrelated populations is warranted to add more light to this question.

Concerted evolution enhances purifying selection in coding DNA.

The promoter and CDS regions of the two *hsp70* duplicates have shown to evolve in concert in our *D. subobscura* data set here represented by four arrangements and also in *D. madeirensis* (Puig-Giribets et al. 2018). In these two species, the genomic region affected by GC is restricted to the promoter and CDS regions, while the paralogous 3'UTRs seem to evolve independently from each other (Table 7.2 and Figure 4B). The evolutionary significance of concerted evolution among members of the same gene family is a hotly debated topic in evolutionary biology. An obvious advantage of GC among gene duplicates is that it parallels the effects of an increase in the effective population size. On this line, GC allows a more efficient elimination of deleterious mutations while it enhances the spreading of favorable ones (Mano and Innan 2008). Retention of sequence similarity among gene duplicates may also be advantageous when there is selection to produce more of the same protein. This is the case of several gene families, such as human Y-chromosome genes (Rozen et al. 2003), ribosomal genes (Brown et al. 1972) and histone genes (Galtier 2003). Thus, in the case of *hsp70* duplicates in *D. subobscura* and its sister species *D. madeirensis* as well as in other *Drosophila* (Bettencourt and Feder 2002), their high identity to produce more of the same protein might be maintained by purifying selection enhanced by recurrent GC. Curiously, the two copies of *D. guanche* appear to be significantly more divergent between each other than those of *D. subobscura*/*D. madeirensis*, suggesting that they might not undergo GC (K_{SIL} = 0.0790; Supplementary table S5). It is possible that the low effective population size of *D. guanche* has resulted in a relaxation of the selective pressures that maintain GC mechanisms active in the other two species (Puig-Giribets et al. 2018). A worth exploring question is whether the suppression of GC in *D. guanche* is related to the lack of puffing activity at the 94A locus in third instar larvae of this species after heat stress that contrasts with *D. subobscura* (Moltó et al. 1988). Surprisingly, more nucleotide substitutions seem to have accumulated in the *hsp70A* than in the *hsp70B* CDS of *D. guanche* when compared with their respective orthologs in *D. subobscura,* which is difficult to interpret. The sequencing of more *hsp70A* and *hsp70B* alleles is necessary to clarify the evolutionary trends of this gene family in *D. guanche.*

Similar gene conversion rates between *D. subobscura* arrangements.

The increased identity between *hsp70* duplicates in O_{ST} pointed to a higher frequency of GC events in this inversion than in the $O_{\underline{3+4}}$ phylad (Puig-Giribets et al. 2018). An evolutionary explanation is that selection to produce a more efficient heat shock response is stronger in O_{ST} to compensate for its increased sensitivity to heat stress relative to the $O_{\underline{3+4}}$ phylad. On this line, a higher GC rate in O_{ST} would allow a more efficient removal of deleterious mutations in this inversion comparable to the effect of an increase in its effective population size. Higher GC rates in O_{ST} could be enhanced by several mechanisms including a more compact disposition of the two *hsp70* copies. Indeed, the two *hsp70* CDS are separated on average by ~2134 bps in $O_{\underline{3+4}}$, 2192 bps in the $O_{\underline{3+4}}$ phylad and 2069 bps in O_{ST}, and the differences between O_{ST} and $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad are statistically significant. For instance, it has been noted that physical distance in the genome is negatively associated with the GC rate (Lichten and Haber 1989; Evgen'ev et al. 2004; Fawcett and Innan 2011; Garbuz et al. 2011). Nonetheless, it is unclear how an average difference of ~60-120 bps might cause a qualitative increase in the GC rate of O_{ST}. Also in contrast with O_{ST}, higher levels of compactness and homogenized coding sequences between *hsp70* genes that may be necessary to enable a faster and more efficient transcriptional response have been described in some heat-resistant strains such as *S. singularior and D. virilis*, whereas greater physical distance between *hsp70* copies and pseudogenization are noted for species from temperate climates like *O. pardalina* and *D. lummei* (Evgen'ev et al. 2004; Garbuz et al. 2011). Despite this discrepancy, it is possible that the higher sensitivity to heat stress in O_{ST} depends on other loci linked to *Hsp70IR*, and that stronger selection in the latter somehow makes up for this situation enabling a more efficient heat shock response (Puig-Giribets et al. 2018). A way to test for differences in the strength of selection is by analyzing codon usage bias. For instance, genes that are highly expressed under stressful conditions such as *hsp70* show bias towards more efficiently transcribed codons (Sharp et al. 1993). Therefore, if selection to mount a more effective heat shock response is more efficient in O_{ST} than in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad, we would expect to find stronger bias in the former. To test for bias in codon usage, we used the Effective Number of Codons (ENC) index implemented in DnaSP. Significant differences in ENC values were observed between O_{ST} and $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad (Mann–Whitney-U test, $p = 0.021$ and $p = 0.014$, respectively); Supplementary figure S3). However, contrary to our initial hypothesis, bias is stronger in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad than in O_{ST}, which together

with estimates on nucleotide diversity indicate that O_{3+4} has a higher effective population size that enables more efficient selection. In addition, our estimates using the silent divergence (K_{SIL}) between *hsp70* duplicates indicate that the differences between O_{ST} and O_{3+4}/O_{3+4} phylad are not statistically significant (Supplementary tables S6.1-S6.2). In conclusion, our results suggest that the shorter distance between the two *hsp70* genes in O_{ST} can in principle be attributed to neutral or near-neutral INDEL polymorphism that tends to be shared between the four O_{ST} haplotypes as a consequence of their relatively recent origin. A higher number of lines per inversion and/or more sensitive methods than those used here may help to confirm or refute these findings.

3.3.1.6 Acknowledgments:

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 and grant AGL2016-78709-R (MEC, Spain) to S.E.R.-O. M .P. G. was supported by a PIF PhD fellowship from the Universitat Autònoma de Barcelona. We thank Dr. Julio Rozas for his help in the polymorphism analyses, Raquel Navarro Fernández for her contribution in the sequencing part and Víctor Gámez Visairas for his advice in the graphic designs.

4. DISCUSSION

The main objectives of this thesis are to characterize the thermally inducible *hsp70* gene family at the molecular and cytogenetic levels in four representative arrangements of *D. subobscura* (O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$) and to quantify *hsp70* mRNA and protein levels under basal conditions in the two adaptive arrangements O_{ST} and $O_{\underline{3+4}}$. Since these arrangements show opposed north-south clinal distribution patterns we intend to shed light on the molecular mechanisms responsible of the adaptive value of these arrangements.

The rich chromosomal inversion polymorphism of *D. subobscura* has been studied by many evolutionary biologists during the last four decades, and the adaptive character of some of its chromosomal arrangements has been repeatedly proven by multiple lines of research (Prevosti et al. 1988, Rodríguez-Trelles et al. 1996; Balanyà et al. 2006; Rodríguez-Trelles et al. 2013; Orengo et al. 2016). Unfortunately, the molecular mechanisms responsible of their adaptive value remain to date unknown. For instance, the heat shock experiments of Calabria et al. (2012) p indicated that flies wearing the warm-adapted $O_{\underline{3+4}}$ arrangement showed increased basal levels of the heat-inducible Hsp70 protein relative to those of $O_{\underline{3+4+8}}$ and the cold-adapted O_{ST} that did not boost with heat stress in contrast with the other two arrangements. The authors postulated that increased basal Hsp70 levels could outnumber the cellular costs in the warm Mediterranean areas where $O_{\underline{3+4}}$ is most abundant (Calabria et al. 2012).

The increased basal Hsp70 protein levels in the $O_{\underline{3+4}}$ arrangement could be caused by several molecular mechanisms. For instance, higher *hsp70* mRNA and/or protein levels after heat shock have been observed in artificial and natural *Drosophila* strains with increased thermotolerance and an extra number of *hsp70* copies (Feder et al. 1996; Evgen'ev et al. 2004; Bettencourt et al. 2008; reviewed in Garbuz and Evgen'ev 2017). Besides a higher number of copies, other possibilities to explain the differences in *hsp70* gene expression levels are changes in *cis*-regulatory DNA regions, for example, by the insertion of transposable elements (TEs) in the promoters that might confer an advantage to certain climate conditions (Zatsepina et al. 2001; Bettencourt et al. 2002; Lerman and Feder 2004). To examine if any of these hypotheses can relate to the

adaptive character of some *D. subobscura* inversions, we developed a comprehensive approach that enabled us to describe the genomic organization and cytogenetic location of the *hsp70* gene family in four representative arrangements of the O chromosome of this species (O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$) and in the closely related *D. madeirensis* and *D. guanche*, and to study basal *hsp70* mRNA and protein levels in two adaptive arrangements that show opposed north-south clinal patterns, the cold-adapted O_{ST} and the warm-adapted $O_{\underline{3+4}}$.

The strategies employed to characterize the *hsp70* gene family at the molecular and cytogenetic levels in the O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$ arrangements of *D. subobscura*, *D. madeirensis* and *D. guanche* are detailed in the research article "Chromosomal inversions promote genomic islands of concerted evolution of Hsp70 genes in the *Drosophila subobscura* species subgroup", and can be divided into two parts. In the first part, we determined the location of the *Hsp70IR* locus in the karyotype of O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$ via ISH using a 1.9 kb DIG-labeled probe designed from the *hsp70* coding region. In the second part, we constructed a genomic library for the O_{ST} inversion and scanned for positive clones with an equivalent probe to the one used for ISH experiments. We sequenced the ~15 kb insert of a positive clone that contained a single *hsp70* gene that, together with contigs of the unassembled *D. subobscura* genome, enabled us to reconstruct a ~10 kb region in a O_{ST} line that comprised two ~2.7-kb long head-to-head oriented *hsp70* copies that we called '*Hsp70IR* locus'. The sequencing of the *Hsp70IR* locus was extended to several *D. subobscura* lines isogenic for the O_{ST} (2), $O_{\underline{3+4}}$ (1), $O_{\underline{3+4+16}}$ (1) and $O_{\underline{3+4+8}}$ (1) arrangements and to the closely related *D. madeirensis* and *D. guanche* species.

In the second chapter "Quantification of *hsp70* mRNA and Hsp70 protein levels under non heat-shock conditions in the warm climate-associated $O_{\underline{3+4}}$ and the cold climate-associated O_{ST} arrangements of *Drosophila subobscura*", we determined mRNA and protein levels under non heat-shock conditions in several replicates of adults isogenic for the cold climate-associated O_{ST} and the warm climate-associated $O_{\underline{3+4}}$ arrangements, males and females separately. For mRNA and protein analyses, we used the quantitative PCR (qRT-PCR) and the enzyme-linked immunosorbent assay (ELISA) techniques, respectively. The results have been contrasted with the extant polymorphism in the *cis*-regulatory proximal promoter regions of the two *hsp70* copies in the different lines. The main objective of this part was to test for sex and/or arrangement-influenced differences

in basal *hsp70* mRNA and/or protein levels in the context of *hsp70* proximal promoter polymorphism.

Finally, in the third chapter "A complex interplay between stability, convergence and divergence drives the evolution of the *Hsp70IR* locus in *Drosophila subobscura*", we used a limited albeit expanded data set of 12 *D. subobscura* lines isogenic for O_{ST} (4), $O_{\underline{3+4}}$ (5), $O_{\underline{3+4+16}}$ (2) and $O_{\underline{3+4+8}}$ (1) to quantify and describe the nucleotide polymorphism patterns across a ~10 kb reconstructed region that comprises the two *hsp70* copies in these four arrangements and to test for departures from neutrality in the cold climate-associated O_{ST} and the warm climate-associated $O_{\underline{3+4}}$. Concomitantly, we explored the degree of genetic differentiation between O_{ST}, $O_{\underline{3+4}}$ and $O_{\underline{3+4+16}}$ and quantified and compared the extant polymorphism in the coding and proximal promoter regions of the two *hsp70* duplicates that appear to evolve in concert in the four studied *D. subobscura* arrangements and in *D. madeirensis* as a result of recurrent gene conversion (GC) events.

In the next lines of this section, I will discuss the most relevant results obtained in the different chapters to provide a global picture of the evolution of the heat-inducible *hsp70* gene family in the context of adaptive inversion polymorphism in *D. subobscura*. The methods developed in this work have proven useful to determine the location of this gene family in the genome of four representative *D. subobscura* arrangements, to describe its genomic organization and nucleotide diversity in these and in the closely related *D. madeirensis* and *D. guanche*, and to quantify its expression levels at the pre- and post-translational stages.

4.1 On *hsp70* genomic organitzation in *D. subobscura* and its closest relatives.

4.1.1 A *CG5608—Hsp70IR—Dmt* syntenic cluster is conserved in *D. subobscura* and *D. madeirensis*.

The reported differences in thermal preference, thermotolerance and basal Hsp70 protein levels between adult homokaryotypes of O_{ST} and $O_{\underline{3+4}}$ (Rego et al. 2010; Calabria et al. 2012) could be explained by differences in *hsp70* copy number and/or genome organization within *hsp70* gene clusters. For instance, some *Drosophila* strains with enhanced thermotolerance show a higher number of *hsp70* copies arranged in more compact clusters relative to less thermotolerant ones that might enhance their transcription activation by cooperative interaction among their neighboring regulatory regions (Garbuz et al. 2003; Evgen'ev et al. 2004; Garbuz et al. 2011; reviewed in Garbuz and Evgen'ev (2017)). Nonetheless, other highly thermotolerant *Drosophila* strains have evolved decreased instead of increased *hsp70* expression levels relative to the wild type strain, indicating that the former trends are not universal (Sørensen et al. 2001; Zatsepina et al. 2001; Lerman et al. 2003). For instance, in **chapter 3.1** we describe the ISH experiments performed on the polytene chromosomes of third-instar larvae isogenic for the O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$ arrangements using a 1.9 kb DIG-labeled probe designed from the *hsp70* coding region. The results indicate that all *hsp70* homologs lie within a single cluster in the 94A band in the segment I of the O chromosome in the four considered arrangements (Figure S1 in **Appendix 7.1.1**). Nonetheless, the resolution of the ISH technique is not high enough to appreciate differences in the number nor arrangement of *hsp70* genes within the *Hsp70IR* cluster that can only be determined by direct sequencing.

To clarify this point, we used the insert of a positive clone of the O_{ST} genomic library and several contigs from unassembled *D. subobscura* genomes to we sequenced a 9-10 kb region in the *Hsp70IR* locus of several O_{ST} (2), $O_{\underline{3+4}}$ (1), $O_{\underline{3+4+16}}$ (1) and $O_{\underline{3+4+8}}$ (1) lines and in the closely related *D. madeirensis* and *D. guanche*. The sequenced region in the aforementioned lineages consisted in a palindrome-like region comprising two ~2.7 kb *hsp70* gene duplicates in divergent orientation separated by a ~1 kb intergenic, nonduplicated region. Each of the two *hsp70* genes encompassed a highly conserved

1.9-kb long coding DNA sequence (CDS) flanked by its respective 5' *cis*-regulatory region. The two *hsp70* genes, conventionally labeled as *hsp70A* and *hsp70B*, were flanked, respectively, by the *Dalmatian* (*Dmt*) and *CG5608* genes that allowed us to distinguish these two highly identical *hsp70* paralogs and to know their orientation in the genome (**Figure 2 in section 3.1.1)**. Together, these loci form a *CG5608—Hsp70IR—Dmt* syntenic cluster that is conserved in all *D. subobscura* lines used in our experiments and in the sister species of *D. subobscura*, *D. madeirensis*. In contrast, the aforementioned syntenic block could not be confirmed in *D. guanche* because we have been unable to PCR-amplify any segment of the *CG5608* gene in this more divergent species.

The preserved synteny and identical copy number in all sequenced *D. subobscura* lines of the different chromosomal arrangements indicate that any differences in gene expression and/or protein production might be caused by molecular mechanisms other than copy number variation, such as changes in *cis*-regulatory DNA regions. However, a close examination of the *hsp70* proximal promoter regions of our O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$ lines and in *D. madeirensis* and *D. guanche* revealed a high degree of conservation in the number, type, location and nucleotide polymorphism of the different *cis*-regulatory elements, including four heat shock elements (HSEs 1-4) described in Tian et al. (2010) and three GAGA (GAF)-motif binding factors (**Figure 3 in section 3.1.1**). Thus, the nucleotide sequence information indicated that the warm climate-associated $O_{\underline{3+4}}$ arrangement did not differ from the other three *D. subobscura* arrangements in the number and organization of *hsp70* genes nor in *cis*-regulatory polymorphism. In our opinion, a similar organization of the *Hsp70IR* locus in the four arrangements is not unexpected considering that $O_{\underline{3+4}}$ is paraphyletic relative to $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$ (Rozas et al. 1999) whose frequency distribution patterns show no positive association with temperature. The high degree of conservation between the four studied *D. subobscura* arrangements indicates that the increased basal Hsp70 protein levels reported for $O_{\underline{3+4}}$ in Calabria et al. (2012) could be caused by the extant polymorphism in other loci and/or changes in post-transcriptional regulation. Alternatively, it is possible that the results of Calabria et al. (2012) are a false positive and warrant further replication.

4.1.2 Concerted evolution takes place in the *Hsp70IR* locus of *D. subobscura* and *D. madeirensis*.

During the sequencing of the *Hsp70IR* locus in the aforementioned lines of the three *subobscura* species that we detailed in **Chapter 3.1**, we noticed that the proximal promoter and coding paralogous regions of the two *hsp70* genes tended to be more similar within the same inversion (**Table 2 and Figure 6 in section 3.1.1)**. This pattern is unexpected considering that the two *hsp70* genes originated from a duplication that likely predated the last common ancestor of all dipterans (Benedict et al. 1993). One possible explanation is that the two *hsp70* duplicates evolve in concert by means of recurrent exchange of DNA fragments that takes place by a mechanism akin to GC. Interestingly, GC between *hsp70* duplicates is not a rare phenomenon and it has been cited numerous times to explain the high similarity between *hsp70* duplicates that is observed in several ditperan species (Brown and Ish-Horowicz 1981; Benedict et al. 1993; Bettencourt and Feder 2001; Bettencourt and Feder 2002).

Surprisingly, the sequence identity between the *hsp70A* and *hsp70B* paralogs (including the promoter and coding regions) was found to be higher in the two O_{ST} lines than in the $O_{\underline{3+4}}$ phylad, represented here by three lines and three arrangements (including $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$) (**Table 2 and Figure 6 in section 3.1.1)**. This might indicate that the GC rate is not constant across all gene arrangements. The evolutionary consequences of GC are diverse (reviewed in Fawcett and Innan 2011). For instance, GC has been postulated to have an effect akin to an increase in the effective population size in that it enhaces a faster spread of benefic mutations and also a faster elimination of deleterious ones (Mano and Innan 2008). Other lines of research indicate that GC often occurs among highly expressed gene duplicates, which results in higher amounts of the same protein product (Fawcett and Innan 2011) that is perhaps useful in genes that must be expressed at high rates under stressful conditions such as *hsp70*. Considering their effects on genetic isolation, particular rates of GC could have been selected for in the different inversions. In this line, we postulated in **chapter 3.1** that an increase in the GC rate might have been favored for in O_{ST} to mount a more efficent heat shock response that would counteract the increased sensitivity to thermal stress in its bearers. Considering that GC rates between homologs appear to be enhanced by high sequence identity and a close physical distance in the genome (Lichten and Haber 1989; Datta et

al. 1997; Evgen'ev et al. 2004; Fawcett and Innan 2011; Garbuz et al. 2011), the higher frequency of GC events in O_{ST} might somehow be enhanced by the disposition of the two *hsp70* copies within the *Hsp70IR* locus that in this arrangement was found to be more compact than in the $O_{\underline{3+4}}$ phylad (**Figure 2 in section 3.1.1**).

Interestingly, GC apparently takes place in *D. madeirensis* at a similar rate than in the $O_{\underline{3+4}}$ phylad but not in the more divergent *D. guanche*, where it seems that the two *hsp70* copies have escaped GC considering their increased divergence relative to that in *D. subobscura*/*D. madeirensis*. This is reinforced by the presence of a modified HSE2 in the promoter of its *Hsp70B* gene that includes two extra pentanucleotide units (**Figure 3 in section 3.1.1**). Although the reasons for this are unclear, it is possible that the cessation of GC events is the consequence of relaxed selection due to high rates of genetic drift in a long-term isolated species with a small population size as *D. guanche*.

Considering that no differences were detected in the organization and copy number of *hsp70* genes nor in their *cis*-regulatory polymorphism between the adaptive O_{ST} and $O_{\underline{3+4}}$ arrangements, in **chapter 3.2** we quantified basal *hsp70* mRNA and protein levels in several lines isogenic for the cold climate-associated O_{ST} and the warm climate-associated $O_{\underline{3+4}}$ arrangements to gain further insight into the differences in the Hsp70 induction profile detailed in Calabria et al. (2012).

4.2 On basal *Hsp70* gene expression quantification at mRNA and protein levels.

4.2.1 Basal *hsp70* mRNA levels do not differ between O_{ST} and $O_{\underline{3+4}}$, but might be influenced by sex

The outcomes of Calabria et al. (2012) indicated that the expression patterns of the heat-inducible *Hsp70* gene family differed between O_{ST} and $O_{\underline{3+4}}$. More precisely, basal Hsp70 protein levels were significantly higher in adult males and females of the warm climate-associated $O_{\underline{3+4}}$ inversion than in their $O_{\underline{3+4+8}}$ and O_{ST} counterparts. Surprisingly, Hsp70 levels measured 2h after heat shock at 32ºC for 1h did not differ between these three arrangements, indicating that *hsp70* genes are responsive to heat stress in $O_{\underline{3+4+8}}$ and O_{ST} but not in $O_{\underline{3+4}}$. An analogous inducible profile to that of $O_{\underline{3+4}}$, with relatively high constitutive Hsp70 expression levels and a weak inducibility after heat shock is usually observed in thermotolerant *Tabanidae* and *Stratiomyidae* species that have a long life cycle (reviewed in Garbuz and Evgen'ev. (2017)). In contrast, the profile of $O_{\underline{3+4}}$ is atypical of species with short life cycles as *Drosophila*, where high basal Hsp70 levels have shown to cause several negative effects during development (Feder et al. 1992; Feder and Krebs. 1998). In our opinion, this makes the results of Calabria et al. (2012) even more unexpected.

To corroborate the results of Calabria et al. (2012) regarding the increased basal Hsp70 levels in the homokaryotipic $O_{\underline{3+4}}$ strains, we quantified basal *hsp70* mRNA and protein levels in adult male and female individuals isogenic for the cold climate-associated O_{ST} and the warm climate-associated $O_{\underline{3+4}}$ arrangements raised under non heat-shock conditions (18ºC). Our outcomes concerning the mRNA data indicate that there are no arrangement-specific differences when the two sexes are analyzed together, yet $O_{\underline{3+4}}$ males show decreased mRNA levels relative to females (**Table 1 in section 3.2.1.4**. The difference between $O_{\underline{3+4}}$ males versus females might indicate that *hsp70* gene expression is influenced by sex. This is not unexpected considering that certain members of the *hsp70* gene family have shown to participate in spermatogenesis in mice and humans (Son et al. 2000; Govin et al. 2006) and that a subset of constitutively expressed Hsp70 proteins has been detected in dividing spermatogonial cells of third-instar larvae and adults of *D. melanogaster* (Lakhotia and Prasanth 2002). Then, it

would not be strange that certain Hsp70 isoforms are also implicated in the spermatogenesis of adult *D. subobscura* males.

A similar *hsp70* gene expression regulation among the four studied *D. subobscura* inversions is reinforced by the extant polymorphism in the *hsp70A* and *hsp70B* proximal promoter sequences. A detailed polymorphism analysis of the first ~250 nucleotide positions upstream of the transcription start site (TSS) in the two *hsp70* genes evidences a high degree of conservation among the 12 lines used for the mRNA and protein expression experiments. As expected, all of the examined promoters contain the same set of conserved *cis*-regulatory elements described in section **3.1.1**, including four HSEs and three GAGA (GAF)-factor binding motifs (Figure 1 in section 3.2.1.4 and Figures S1 and S2 in **Appendices 7.2.3 and 7.2.4**). The *hsp70* proximal promoter regions are in some cases even more similar within the same line probably as a result of recurrent GC events as detailed in **chapter 3.1**. In conclusion, although *in silico* experiments are needed to confirm it, our results indicate that the two *hsp70* genes are regulated in a similar way in O_{ST} and $O_{\underline{3+4}}$, and also in $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$ according to their available promoter sequences (**Figure 3 in section 3.1.1**).

4.2.2 Basal Hsp70 protein levels are influenced by sex and arrangement

The quantification of basal Hsp70 protein levels in our O_{ST} and $O_{\underline{3+4}}$ lines shows that these do not differ between the two sexes nor the two arrangements yet it indicates that there is an interaction effect between the variables sex and arrangement, where O_{ST} males show increased protein levels relative to females of the same arrangement, and the opposite trend is observed in $O_{\underline{3+4}}$ flies (**Figure 2 in section 3.2.1.4)**. These results contrast with those of Calabria et al. (2012) in that they indicate that Hsp70 levels are influenced by sex, and that this influence varies among the two considered arrangements. There are several possibilities to explain these discrepancies, including separate geographical origins of the strains used in the two studies (Barcelona vs. Berbikiz), different warm climate-associated chromosomal arrangements ($O_{\underline{3+4}}$ vs. $O_{\underline{3+4}}$) and dissimilar degrees of genomic inbreeding (homokaryotypic but heterozygous vs. fully homozygous O chromosomes). Although these differences could partly explain the lack of reproducibility between our experiments and those of Calabria et al. (2012), we are fairly confident in our results that show no differences in basal Hsp70 protein levels between $O_{\underline{3+4}}$ and O_{ST} when the two sexes are analyzed together, and that are in turn

supported by the DNA sequence information and mRNA expression quantification. Thus, the unorthodox Hsp70 induction profile of $O_{\underline{3+4}}$ flies reported in Calabria et al. (2012) cannot be extrapolated to all *D. subobscura* warm climate-associated arrangements.

Interestingly, the interaction effect between arrangement and sex indicates that loci in the autosomal O chromosome and the sex chromosomes influence global Hsp70 protein levels in *D. subobscura*. Indeed, some previous studies have found differences in Hsp70 levels between the two sexes after heat stress in *D. melanogaster* (Dahlgaard et al. 1998; Folk et al. 2006), yet the sex-linked factors presumably implicated in their regulation remain unknown. In addition, certain Hsp70 protein isoforms have shown to participate in the spermatogenesis of various organisms including mice, humans and *D. melanogaster* (Son et al. 2000; Lakhotia and Prasanth 2002; Lerman et al. 2003; Govin et al. 2006), and our *hsp70* mRNA expression quantification results indicate that $O_{\underline{3+4}}$ males have significantly lower levels than their female counterparts, which reinforces the idea that Hsp70 protein production is somehow influenced by sex also in *D. subobscura*. Nonetheless, the regulatory mechanisms might switch between the pre- and post-translational stages because O_{ST} males tend to have increased Hsp70 protein levels relative to their female counterparts, yet the opposite trend is observed for mRNA data.

For instance, the interaction effect is opposite between O_{ST} and $O_{\underline{3+4}}$, indicating that there is some linkage between these autosomal O-chromosome arrangements and the sexual A chromosome of *D. subobscura* (there are no described inversions in the Y chromosome). Nonetheless, no association between inversions in the five acrocentric chromosomes of *D. subobscura* has been detected so far, which indicates that these combine at random (Zivanovik et al. 2016). The interaction effect is even more unexpected considering that the 12 lines used in our expression quantification experiments are supposed to carry the same genetic content as the laboratory *ch-cu* strain used for the isogenization protocol detailed in Sperlich et al. (1977) in all but in their O and Y chromosomes that are inherited from wild-type males. Nonetheless, very few functional loci have been described in the Y chromosomes of *Drosophila* that are mostly made up of heterochromatic DNA (Charlesworth 2001), which indicates that this chromosome is unlikely to cause the interaction effect. On the other hand, the high conservation of the *hsp70* gene family in the O_{ST} and $O_{\underline{3+4}}$ arrangements at the nucleotide and organizational levels indicates that the polymorphism in the *Hsp70IR*

locus alone is insufficient to explain the reported differences between the two arrangements. Thus, other unknown loci, presumably located within the inverted regions between O_{ST} and $O_{\underline{3+4}}$ that allegedly interact with sex-linked factors might influence basal Hsp70 protein amounts aside from the *Hsp70IR* locus itself. Alternatively, our results are a false positive and further replication is warranted.

Finally, it is worth mentioning that, unlike in Calabria et al. (2012), we did not perform any heat shock experiment and that our measurements of Hsp70 protein amounts are limited to the basal conditions. On this line, we cannot rule out that the two considered arrangements might differ in their *hsp70* induction profile, either during or shortly after heat stress. Consequently, future heat shock experiments that measure *hsp70* mRNA and protein levels before and after exposure to heat stress are necessary to confirm or refute if there exists any arrangement-specific and/or sex-specific differences in the Hsp70 upregulation in response to temperature increase.

4.2.3 No correlation is observed between *hsp70* mRNA and protein levels within the same line.

In the last part of the expression quantification experiments, we tested whether there is a correlation between basal mRNA and protein amounts for the same line. Considering that these appear to be sex-influenced traits, we made the comparisons separately in the two sexes using the average values calculated from three biological replicates in the 12 lines, which indicated a lack of correlation (**Table 3 in section 3.2.1.4**). This was somehow expected considering that the values of the three biological replicates for the same line in the mRNA and protein quantification experiments show a relatively high variability within the same line regardless of the arrangement. Such a high variability is difficult to explain considering that the individuals belonging to the same line are expected to carry genetically identical O chromosomes. Consequently, the lack of correlation might be explained by different regulatory mechanisms in the post-transcriptional and post-translational stages.

Aside from high genetic diversity within the same line, there are other explanations for a lack of a correlation between *hsp70* mRNA and protein amounts. For instance, it is difficult to compare our data to former work, because most studies in *Drosophila* and other taxa have studied the *hsp70* induction profile during or after exposure to heat stress, when the *hsp70* mRNA and protein levels become substantially elevated. In fact,

Hsp70 protein amounts are expected to be low to negligible in *Drosophila* cells at normal temperatures due to their numerous detrimental effects (Feder et al. 1992). That being said, a good correlation has been observed between mRNA and protein during heat stress in *D. melanogaster* females (Sørensen et al. 2013) that is even stronger than in other lineages (Bahrndorff et al. 2009). However, the correlation seems to exist only after exposure to heat stress, and is not observed at cold temperatures, a possibility being that *hsp70* mRNA levels are too low to cause a substantial increase in Hsp70 protein amounts (Colinet et al. 2013). In fact, *hsp70* mRNA amounts in our samples are roughly 100 to 1000 times less abundant that those of the *rp49* reference gene as calculated using the comparative C_T method (Schmittgen and Livak 2008; Table S1 in **Appendix 7.2.1**). The relatively low levels of the *hsp70* messenger are explained by its high instability at normal temperatures (Bönisch et al. 2007), which probably keeps Hsp70 protein amount to a minimum in the cell to avoid the harmful effects.

Overall, the evidence presented so far indicates that *hsp70* gene expression is not a good predictor of Hsp70 protein amounts in *D. subobscura* at normal temperatures. It is possible that this discrepancy is partly explained by the high variability within the same line. However, the mRNA and protein profiles still differ when all male and female samples of the same arrangement are grouped together, indicating that there are other variables at play. For instance, it is possible that different temporal dynamics between mRNA and protein production exist in *D. subobscura* and/or that mRNA levels in the cell at normal temperatures are too low to substantially increase Hsp70 protein amounts.

Considering that gene expression patterns cannot explain the adaptive value of O_{ST} and $O_{\underline{3+4}}$, in chapter 3.3 we quantified the polymorphism levels in the O_{ST}, $O_{\underline{3+4}}$ arrangements and in the $O_{\underline{3+4}}$ phylad (including $O_{\underline{3+4+8}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4}}$) by taking advantage of an expanded data set relative to that used in chapter 3.1 to test for departures from neutral evolution within the *Hsp70IR* locus that might relate to their adaptive character.

4.3 On the extant polymorphism in the *Hsp70IR* locus in several *D. subobscura* inversions.

4.3.1 The polymorphism at the *Hsp70IR* locus evolves close to a neutral model in the adaptive O_{ST} and $O_{\underline{3+4}}$ arrangements.

The sequencing of a ~10 kb region in the *Hsp70IR* locus of our complete *D. subobscura* data set consisting in 12 lines and four arrangements ($O_{\underline{3+4+8}}$ (1), $O_{\underline{3+4+16}}$ (2), O_{ST} (4) and $O_{\underline{3+4}}$ (5)) confirms that a similar genomic organization prevails in these four representative arrangements as previously described in Chapter 1 (**Figure 1 in section 3.3.1.4**). To explore and understand the nucleotide diversity patterns in the different gene arrangements, we divided the ~10 kb reconstructed region into seven parts according to their distinctive molecular function and characteristics, including the two intronless CDSs of the *hsp70A* and *hsp70B* genes, their 5' and 3' untranslated flanking regions (UTRs) and the intergenic, nonduplicated region that separates the two copies (IGR).

Several neutrality tests (including the Tajima's *D* (Tajima 1989), Fu and Li's *D* and *F* (Fu and Li 1993), and the normalized Fay and Wu's H-test (Zeng et al. (2006)) were performed on the extant single nucleotide polymorphism (SNP) of the seven regions to detect departures from neutrality in the different gene arrangements and in our complete *D. subobscura* data set. The results indicate that the nucleotide polymorphism in the seven regions, when analyzed separately and combined, evolves close to the expectations of neutral evolution in the O_{ST} and $O_{\underline{3+4}}$ arrangements, in the $O_{\underline{3+4}}$ phylad and in *D. subobscura* (**Table 1 in section 3.3.1.4 and Table S3 in Appendix 7.3.4**). Surprisingly, polymorphism levels vary sharply between neighboring regions and in a similar fashion in the adaptive O_{ST} and $O_{\underline{3+4}}$ arrangements as depicted in **Figure 2, section 3.3.1.4**, with substantially higher levels in the IGR relative to the selectively constrained CDSs and their proximal promoters. As expected, polymorphism levels in the nonsynonymous sites are only a tenth of those in the synonymous sites in the two conserved *hsp70* copies, and show a trend towards significantly negative Tajima's *D* values that indicate that most nonsynonymous changes are moderately deleterious, and therefore rarely increase their frequency in the population (**Table S3 in Appendix 7.3.4)**. Estimates of insertion/deletion (INDEL) polymorphism in five regions (including the 5' and 3' UTRs, and the IGR) indicate that this type of polymorphism is

most abundant and diverse in the IGR, which is agreement with SNP estimates **(Table 6 in section 3.3.1.4)**. The increased nucleotide polymorphism in the IGR might relate to the relatively abundance of repetitive, non-homologous DNA motifs in this region that complicate multiple sequence alignment. It is worth mentioning that some of these repetitive DNA sequences show moderate similarity to transposable elements (TEs) from diverse families (not shown, but **see section 3.3.1.5 for more details**), and that certain TEs are thought to have played an important role in the organization of *hsp70* gene clusters in some *Drosophila* by means of homologous recombination and retrotransposition mechanisms (Bettencourt and Feder 2001; Evgen'ev et al. 2004; reviewed in Garbuz and Evgen'ev (2017)).

Interestingly, the Hudson-Kreitman-Aguadé (HKA) test indicates that the two *hsp70* paralogous genes (including their CDS and the 5' and 3'UTRs) evolve at a similar rate in the different *D. subobscura* inversions, including O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and the $O_{\underline{3+4}}$ phylad (see **section 3.3.1.4**). Similarly, the McDonald and Kreitman (MKT) test suggests that there is no departure from neutrality in any of the two *hsp70* CDSs when we use their respective orthologs in *D. guanche* as outgroup sequences, which corroborates that the two *hsp70* genes evolve at similar rates in the aforementioned arrangements of *D. subobscura* (see **Table 4** in **section 3.3.1.4**). Surprisingly, the *hsp70A* CDS seems to have accumulated more changes at both synonymous and nonsynonymous sites between *D. guanche* and *D. subobscura* orthologs relative to the *hsp70B* CDS (see **section 3.3.1.4**), indicating that *hsp70A* has somehow evolved faster during the last 1.8-2.8 MY that separate the *D. guanche* and *D. subobscura*/*D. madeirensis* lineages (Ramos-Onsins et al. 1998). These differences are rather unexpected, and more *D. guanche* alleles are needed to confirm these outcomes.

Overall, the high conservation between the two *hsp70* duplicates and the similarities in the nucleotide diversity patterns indicate in our opinion that the *Hsp70IR* locus alone is unlikely to explain the adaptive value of the cold climate-associated O_{ST} and the warm climate-associated $O_{\underline{3+4}}$ arrangements, and that other candidate loci must be searched.

4.3.2 The reduced nucleotide diversity in the O_{ST} arrangement points to a low effective population size

Although the nucleotide diversity patterns are similar in the seven regions of the *Hsp70IR* locus between O_{ST} and $O_{\underline{3+4}}$, we noted that the nucleotide diversity is substantially lower in the former arrangement in most regions of the locus, including the CDSs, 5' UTRs and the IGR (**Table 1 and Figure 2 in section 3.3.1.4**). More precisely, estimates on the synonymous polymorphism in the *hsp70* genes and in the IGR indicate that nucleotide diversity in the O_{ST} arrangement represented here by 4 haplotypes is only ½ to ⅓ of that of the $O_{\underline{3+4}}$ arrangement represented by three haplotypes. This difference is reflected in the estimates of the coalescence times that we made using the synonymous polymorphism of the two *hsp70* genes that are about half in O_{ST} (~0.25 MYO) relative to $O_{\underline{3+4}}$ (~0.52 MYO) assuming the same number of generations per year (**Table 5 in section 3.3.1.4**). This indicates that the effective population size is significantly lower in O_{ST} than in $O_{\underline{3+4}}$. The higher diversity levels in $O_{\underline{3+4}}$ are unlikely to be caused by a more diverse origin of its three haplotypes, because these stem from wild males collected seasonally in the same population and year as the four O_{ST} haplotypes. Consequently, other explanations should be elaborated.

Some former studies that have provided estimates on the age of these inversions indicate that O_{ST} is either the same age or slightly younger than $O_{\underline{3+4}}$ (Rozas and Aguadé 1994; Navarro-Sabaté et al. 1999; Rozas et al. 1999; Pegueroles et al. 2013). Unfortunately, we cannot infer the age of our inversions using the polymorphism at the *Hsp70IR* locus because no departure from neutrality is observed, which indicates that these are no longer in the rapid population expansion phase that occurs after the origin of an inversion that leaves a footprint in the pattern of nucleotide diversity that can be used to calculate its age. Consequently, we have no evidence to prove whether the reduced levels in O_{ST} are due to a younger origin of this inversion relative to $O_{\underline{3+4}}$. Considering that the diversity in the O_{ST} arrangement is very similar or only slightly lower than that of $O_{\underline{3+4}}$ in most studies, we think that this is likely not the principal explanation.

For instance, the reduced nucleotide diversity levels in O_{ST} are unlikely to be caused by selection, because our data indicates that the polymorphism at the *Hsp70IR* locus evolves close to the neutral expectations in all of the examined arrangements including

O_{ST}. Considering these observations, the lower effective population size of O_{ST} as deduced from the polymorphism at the *Hsp70IR* locus is possibly caused by stochastic fluctuations in the inversion frequencies in the source population, that is, some inversions become more prevalent in the population during certain periods of time while others, like O_{ST}, experience a loss of diversity. It is likely that natural selection is behind these fluctuations in the inversion frequencies, as suggested by long-term cyclic seasonal changes (Rodríguez-Trelles et al. 1996). However, *Hsp70IR* is apparently not among the loci that are being selected.

An alternative explanation to the reduced nucleotide diversity levels in O_{ST} postulates that undetected recombination tracts from different chromosomal arrangements result in inflated diversity levels for $O_{\underline{3+4}}$ as noted in Rozas et al. (1999). In fact, we have been able to detect relatively long (~1 kb) gene conversion tracts that likely resulted from gene flow between different arrangements in two $O_{\underline{3+4}}$ lines of our *D. subobscura* data set, indicating that these events might be more common than assumed (**Table 2 in section 3.3.1.4**). The study of a wider set of *hsp70* haplotypes from unrelated populations might help to make better estimates of the nucleotide diversity levels and the effective population size of the four considered arrangements.

4.3.3 Gene conversion is restricted to the proximal promoter and CDS regions in the four *D. subobscura* arrangements

In chapter 3.1 we provided evidence that the two highly conserved *hsp70* paralogs (including their CDSs and 5'UTRs) are more similar at the nucleotide level within the same arrangement and species, indicating that they evolve in concert by means of ectopic gene conversion (GC) events that periodically erase the accumulated divergence between them. In the third chapter we corroborated these trends and confirmed that the two *hsp70* paralogs are not only more similar within the same arrangement, but also in some cases within the same line (**Tables S4.1-S4.4 in Appendix 7.3.5**).

Considering that the genomic region that is affected by gene conversion is restricted to the functional CDS and its proximal promoter regions, it is likely that the high identity between the *hsp70A* and *hsp70B* genes in *D. subobscura* is the result of both purifying selection and recurrent GC events. The fact that GC has been detected in the four considered *D. subobscura* arrangements and also in *D. madeirensis* indicates that there

has been a selective pressure during the last 0.6-1.0 MY (Ramos-Onsins et al. 1998) to keep GC mechanisms active. Interestingly, GC has been detected in the *hsp70* genes of several *Drosophila* and also in *Anopheles*, indicating that selection to kept GC mechanisms operating has existed over a much longer time frame (Benedict et al. 1993; Bettencourt and Feder 2001; Bettencourt and Feder 2002). Interestingly, GC shapes the evolution of the 5' and CDSs regions but not the short (~200 bp) 3'UTRs of the two *hsp70* genes that show a much lower degree of identity between paralogs (~75%) than between orthologs (Tables 7.1-7.2 **in section 3.3.1.4**). This is perhaps hardly surprising considering that efficient gene conversion requires high (>80%) sequence identity (Mansai et al. 2011). Taking into account that AU-rich sequences in the 3'UTR have shown to play an essential role in the regulation of *hsp70* transcript half-life (Petersen and Lindquist 1989; Balakrishnan and De Maio 2006; Bönisch et al. 2007), the high divergence between the *hsp70A* and *hsp70B* paralogous 3'UTRs indicates that at least two different *hsp70* transcript isoforms are produced that might differ in their half-life.

Although GC is postulated to occur in O_{ST}, in the $O_{\underline{3+4}}$ phylad and in *D. madeirensis*, the GC rate might not be the same in all of these lineages. Indeed, in **chapter 3.3** we found a significantly higher identity between *hsp70* paralogs in the O_{ST} arrangement than in the $O_{\underline{3+4}}$ phylad that could result from an increased GC rate in the former. Distinct GC rates between inversions are possible due to their genetic isolation effects. On this line, certain polymorphisms enhancing more efficient and/or frequent GC rates might have been selected for in one inversion and have been m aintained in it due to the reduced recombination rates. For instance, we postulated in **chapter 3.3** that the increased GC rate in the O_{ST} arrangement might make up for its higher sensitivity to heat stress relative to the inversions of the $O_{\underline{3+4}}$ phylad (including $O_{\underline{3+4+8}}$, $O_{\underline{3+4}}$ and $O_{\underline{3+4+16}}$). In this line, more efficient GC would eliminate faster the deleterious mutations that accumulate in the two *hsp70* CDSs, resulting in a more efficient heat shock response. Nonetheless, the dynamics of the heat shock response in O_{ST} do not seem to differ from those of $O_{\underline{3+4+8}}$ according to the results of Calabria et al. (2012). If selection to mount a more efficient GC response is indeed stronger in O_{ST}, we would expect to find a higher degree of conservation in the nonsynonymous polymorphism and/or in the *cis*-regulatory proximal promoter regions of this arrangement. However, an analysis of the proximal promoter sequences in the *hsp70A* and *hsp70B* genes of several O_{ST} and $O_{\underline{3+4}}$ lines indicates that *cis*-regulatory elements are equally conserved in the two

arrangements (Supplementary Figures S1 and S2 in **Appendices 7.2.3 and 7.2.4**). Likewise, the nucleotide diversity levels at the nonsynonymous sites of the two *hsp70* genes do not seem to differ much among the different gene arrangements (including O_{ST}, $O_{\underline{3+4}}$ and the $O_{\underline{3+4}}$ phylad) (**Table 1 in section 3.3.1.4**), and the predicted amino acid sequences of the two paralogous *hsp70* CDSs tend to be most similar within the same line due to the effects of GC (**Figures 4A-4B in section and Tables S4.1-S4.4 in Appendix 7.3.5)**. Furthermore, an examination of codon usage bias in the two *hsp70* genes indicates that these are more pronounced in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad than in O_{ST} (Figure S3 in **Appendix 7.8.9)**. This suggests that selection to enhance a more efficient transcription is slightly stronger in $O_{\underline{3+4}}$ than in O_{ST}, which might relate to the aforementioned differences in the effective population size between these two arrangements.

In agreement with the results in **chapter 3.1** that point to higher levels of compactness in O_{ST}, we found a significant difference in the number of nucleotides that separate the two *hsp70* CDSs between O_{ST} and the $O_{\underline{3+4}}$ phylad, being ~120 bps shorter on average in the former arrangement (Figure S2 in **Appendix 7.3.7**). Although the loci implicated in GC mechanisms are not known, it is possible that a more compact disposition of *hsp70* copies in O_{ST} might somehow enhance more frequent and/or efficient GC events in this arrangement. In this line, it is worth mentioning that higher levels of compactness within *hsp70* gene clusters are observed in the genomes of some thermotolerant *Drosophila* strains where they have been postulated to enhance a faster and more efficient transcriptional response (Evgen'ev et al. 2004; Garbuz et al. 2011). However, we could not corroborate a higher identity between *hsp70* duplicates in O_{ST} than in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad using our expanded data set (Tables S6.1-S6.2 in **Appendix 7.3.8**), which indicates that the association between higher compactness and increased GC rates remains elusive. It is worth mentioning that, although high compactness might have been selected for to enhance a faster and more efficient heat shock response in certain taxa, it could also result from high rates of DNA loss associated to inactive TE copies as previously reported in some *Drosophila* lineages, including *D. melanogaster* and *D. virilis* (Petrov et al. 1996; Petrov and Hartl 1998). In this line, it is worth remembering that several traces of TE elements seem to exist in the intergenic, nonduplicated region (IGR).

Unfortunately, a big limitation of our analyses is that the number of lines per arrangement (4 for O_{ST}, 3 for $O_{\underline{3+4}}$, 2 for $O_{\underline{3+4+16}}$ and 8 for the $O_{\underline{3+4}}$ phylad) is rather low. A higher number of lines might help to confirm if there are indeed differences in the degree of divergence between *hsp70* paralogs in the different gene arrangements, and whether some of them like O_{ST} show a more compact disposition of their *hsp70* genes that might somehow influence their heat shock response.

5. CONCLUSIONS

1. *In situ* hybridization experiments performed in third-instar larvae isogenic for the representative O-chromosome arrangements O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$ of *D. subobscura* revealed that *hsp70* homologs are found within a single locus (‘*Hsp70IR’*) in all of them. This locus is located on the 94A band in the distal segment (SI) of the O chromosome that is comprised within the inverted region between these four arrangements.
2. The sequencing of a ~10 kb genomic region at the *Hsp70IR* locus in several lines isogenic for the O_{ST} (2), $O_{\underline{3+4}}$ (1), $O_{\underline{3+4+16}}$ (1) and $O_{\underline{3+4+8}}$ (1) arrangements indicates that it consists in two ~2.7 kb paralogous *hsp70* genes in divergent orientation separated by a ~1-kb long intergenic, nonduplicated region. A similar genomic organization prevails in the closely related *D. madeirensis* and *D. guanche*.
3. The two *hsp70* paralogous genes, here conventionally labeled as *hsp70A* and *hsp70B,* encompass each a 1.9 kb coding sequence (CDS) and its respective 5’ *cis*-regulatory region. The two paralogs are flanked, respectively, by the *CG5608* and *Dmt* genes forming a *CG5608-Hsp70IR-Dmt* syntenic block that is conserved in the four considered *D. subobscura* arrangements and in its sister species *D. madeirensis*.
4. The two paralogous CDSs and their 5’ *cis*-regulatory regions show a high degree of identity at the sequence level (>97%) in the four studied *D. subobscura* arrangements. This similarity is even higher within the same arrangement and species, which indicates that the two *hsp70* copies evolve in concert by means of ectopic gene conversion (GC).
5. GC is detected among the *hsp70* duplicates of the four *D. subobscura* arrangements and in *D. madeirensis*, but it appears to be turned off in *D. guanche,* where the divergence between the two paralogs is substantially higher. A possibility is that the *hsp70* duplicates in *D. guanche* have escaped GC as an aftereffect of relaxed selection in a long-term isolated species with a low effective population size.
6. A higher identity at the sequence level and a more compact disposition between the two *hsp70* genes is observed in the two examined O_{ST} lines relative to the $O_{\underline{3+4}}$ phylad (including $O_{\underline{3+4}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4+8}}$). A possibility to explain this increased similarity is that GC rates are higher and/or more efficient in O_{ST} than in any of the other three arrangements.
7. The increased GC rate could have been selected for in the cold climate-associated O_{ST} to compensate for the increased sensitivity of its bearers to thermal stress

relative to the warm climate-associated $O_{\underline{3+4}}$. On this line, higher GC rates in O_{ST} would enhance a more efficient elimination of deleterious mutations akin to an increase in the effective population size, which might result in a more efficient heat shock response in this inversion.

8. The analysis of the proximal promoters in the two *hsp70* genes of the different lines and species show that they contain several conserved *cis*-regulatory elements, including four heat shock elements (HSEs 1-4) and three binding sites for the GAGA (GAF) factor. Interestingly, the HSE2 motif shows two extra pentanucleotide units in the promoter of the *hsp70B* gene in *D. guanche*, which might or might not have consequences on gene expression regulation.
9. A comparison of basal *hsp70* mRNA expression levels in several replicates of adult males and females isogenic for the $O_{\underline{3+4}}$ and O_{ST} arrangements indicate that these are similar between O_{ST} and $O_{\underline{3+4}}$ females and between O_{ST} males and females, yet they are significantly lower in $O_{\underline{3+4}}$ males than in females of the same arrangement. This indicates that *hsp70* gene expression might be influenced by sex in *D. subobscura*.
10. The basal Hsp70 protein levels are found at comparable levels between males and females and also between O_{ST} and $O_{\underline{3+4}}$ when the two sexes are analyzed together. However, there is an unexpected interaction effect between sex and arrangement, where protein levels tend to be increased in O_{ST} males relative to females, while the opposite trend is observed in $O_{\underline{3+4}}$. This pattern indicates that unidentified loci in the O chromosome might interact epistatically with sex-linked factors, which is in turn reinforced by the mRNA data analysis.
11. A lack of correlation between basal *hsp70* mRNA and protein levels is observed within the same line. It is also possible that *hsp70* mRNA levels are too low at normal temperatures to cause a significant increase in Hsp70 protein levels.
12. The similarities in gene expression regulation between O_{ST} and $O_{\underline{3+4}}$ are corroborated by the analysis of *cis*-regulatory elements in the proximal promoters of the *hsp70A* and *hsp70B* genes in the 12 lines that show the same set of conserved elements described in point 7, including four HSEs and three GAGA-motif binding factors.
13. The sequencing of a ~10 kb region in the *Hsp70IR* locus of an expanded data set of 12 *D. subobscura* lines isogenic for the $O_{\underline{3+4}}$ (5), O_{ST} (4), $O_{\underline{3+4+8}}$ (1) and $O_{\underline{3+4+16}}$ (2) confirms that a similar genomic organization prevails in these four arrangements. Polymorphism levels vary sharply between seven delimited regions within the locus

yet in a similar fashion between arrangements, being highest in the nonduplicated intergenic region (IGR).

14. No departure from neutrality is detected after analyzing the polymorphism in the seven aforementioned distinctive regions (including the two paralogous CDSs and their 5' and 3' flanking regions, and the IGR) in the two adaptive *D. subobscura* arrangements, namely the warm climate-associated $O_{\underline{3+4}}$ and the cold climate-associated O_{ST}.
15. The two *hsp70* genes seem to evolve at a comparable rate in the different *D. subobscura* arrangements (including O_{ST}, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4}}$) but not in *D. guanche*. Estimates of the divergence between *D. guanche* and *D. subobscura* orthologs indicate that the *hsp70A* gene has accumulated more changes at both nonsynonymous and synonymous sites since the split between the two species, while *hsp70B* is more conserved.
16. The fixation index (F_{ST}) of population genetic differentiation indicates that there is a high degree of structure between O_{ST} and the $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad. In contrast, we failed to detect a significant genetic differentiation between the more closely related $O_{\underline{3+4+16}}$ and $O_{\underline{3+4}}$ arrangements using such a small data set.
17. The genetic diversity in the synonymous polymorphism of the O_{ST} arrangement is less than ½ of that in $O_{\underline{3+4}}$, which indicates that the former has a lower effective population size. Considering that no departure from neutrality is observed, these differences are best explained by random fluctuations in the inversion frequencies in the population of origin.
18. Gene conversion affects the polymorphism of the selectively constrained paralogous CDSs and their 5' *cis*-regulatory regions that tend to be most similar within the same arrangement and sometimes within the same line. The effects of GC are even observed in the nonsynonymous polymorphism, where the rarest amino acid sometimes only occurs in the two paralogs of the same line. In contrast, GC mechanisms do not appear to be active in the much more divergent paralogous 3'UTRs.
19. Using an extended data set of 12 lines, we corroborate that the compactness between the two *hsp70* CDSs is significantly higher in O_{ST} than in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad. In contrast, the silent divergence between the two *hsp70* CDSs does not differ significantly between these arrangements, which indicates that GC mechanisms might be similar in all of them.

6. BIBLIOGRAPHY

Alamiri, Z. 2000. Conducta de oviposicion en cuatro especies de *Drosophila*. *Gayana* **64**:123-131.

Ananina, G., Peixoto, A. A., Bitner-Mathé, B. C., Souza, W. N., Silva, L. B. D., Valente, V. L., and Klaczko, L. B. 2004. Chromosomal inversion polymorphism in *Drosophila mediopunctata*: seasonal, altitudinal, and latitudinal variation. *Genet. Mol. Biol.* **27:**61-69.

Anderson, A. R., Hoffmann, A. A., Mckechnie, S. W., Umina, P. A., and Weeks, A. R. 2005. The latitudinal cline in the In(3R)Payne inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Mol. Ecol.* **14**:851-858.

Anderson, W. W., Arnold, J., Baldwin, D. G., Beckenbach, A. T., Brown, C. J., Bryant, S. H., and Jeffery, D. E. 1991. Four decades of inversion polymorphism in *Drosophila pseudoobscura*. *Proc. Natl. Acad. Sci. U.S.A.* **88**:10367-10371.

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

Arbona, M., Frutos, R.D., Tanguay, R.M. 1993. Transcriptional and translational study of the *Drosophila subobscura HSP83* gene in normal and heat-shock conditions. *Genome* **36**:694-700.

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

Ayala, D., Ullastres, A., and González, J. 2014. Adaptation through chromosomal inversions in *Anopheles*. *Front. Genet.* **5**:129.

Ayala, F. J., Serra, L., and Prevosti, A. 1989. A grand experiment in evolution: the *Drosophila subobscura* colonization of the Americas. *Genome* **31**:246-255.

Bahrndorff, S., Mariën, J., Loeschcke, V., and Ellers, J. 2009. Dynamics of heat-induced thermal stress resistance and hsp70 expression in the springtail, *Orchesella cincta*. *Funct. Ecol.* **23**:233-239.

Balakrishnan, K., and De Maio, A. 2006. Heat shock protein 70 binds its own messenger ribonucleic acid as part of a gene expression self-limiting mechanism. *Cell Stress Chaperon.* **11**:44-50.

Balanyá, J., Huey, R. B., Gilchrist, G. W., and Serra, L. 2009. The chromosomal polymorphism of *Drosophila subobscura*: a microevolutionary weapon to monitor global change. *Heredity* **103**:364-367.

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

Balanyá, J., Serra, L., Gilchrist, G. W., Huey, R. B., Pascual, M., Mestres, F., and Solé, E. 2003. Evolutionary pace of chromosomal polymorphism in colonizing populations of *Drosophila subobscura*: an evolutionary time series. *Evolution* **57**:1837-1845.

Balanyá, J., Solé, E., Oller, J. M., Sperlich, D., and Serra, L. 2004. Long-term changes in the chromosomal inversion polymorphism of Drosophila subobscura. II. European populations. *J. Zool. Syst. Evol. Res.* **42**:191-201.

Balding, D. J., Nichols, R. A., and Hunt, D. M. 1992. Detecting gene conversion: primate visual pigment genes. *Proc. Royal Soc. B.* **249**: 275-280.

Barrio, E., and Ayala, F. J. 1997. Evolution of the *Drosophila obscura* species group Inferred from the *Gpdh* and *Sod* genes. *Mol. Phylogenet. Evol.* **7**:79-93.

Belyayev, A. 2014. Bursts of transposable elements as an evolutionary driving force. *J. Evol. Biol.* **27**:2573-2584.

Benedict, M. Q., Cockburn, A. F., and Seawright, J. A. 1993. The *Hsp70* heat-shock gene family of the mosquito *Anopheles albimanus*. *Insect Mol. Biol.* **2**:93-102.

Betrán, E., Rozas, J., Navarro, A., and Barbadilla, A. 1997. The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. *Genetics* **146**:89-99.

Bettencourt, B. R., and Feder, M. E. 2001. Hsp70 duplication in the *Drosophila melanogaster* species group: how and when did two become five?. *Mol. Biol. Evol.* **18**:1272-1282.

Bettencourt, B. R., and Feder, M. E. 2002. Rapid concerted evolution via gene conversion at the Drosophila *Hsp70* genes. *J. Mol. Evol.* **54**:569-586.

Bettencourt, B. R., Hogan, C. C., Nimali, M., and Drohan, B. W. 2008. Inducible and constitutive heat shock gene expression responds to modification of *Hsp70* copy number in *Drosophila melanogaster* but does not compensate for loss of thermotolerance in *Hsp70* null flies. *BMC Biol.* **6**:5-15.

Bettencourt, B. R., Kim, I., Hoffmann, A. A., and Feder, M. E. 2002. Response to natural and laboratory selection at the *Drosophila Hsp70* genes. *Evolution* **56**:1796-1801.

Bettencourt, B.R., Feder, M.E., Cavicchi, S. 1999. Experimental evolution of *Hsp70* expression and thermotolerance in *Drosophila melanogaster*. *Evolution*. **53**:484-492.

Bombardier, E., Vigna, C., Iqbal, S., Tiidus, P.M., Tupling, A.R. 2009. Effects of ovarian sex hormones and downhill running on fiber-type-specific *HSP70* expression in rat soleus. *J. Appl. Ecol.* **106**:2009-2015.

Bönisch, C., Temme, C., Moritz, B., and Wahle, E. 2007. Degradation of hsp70 and other mRNAs in *Drosophila* via the 5′–3′pathway and its regulation by heat shock. *J. Biol. Chem.* **282**:21818-21828.

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

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

Brown, D. D., Wensink, P. C., and Jordan, E. 1972. A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J. Mol. Biol.* **63**:57-73.

Burla, H., and Götz, W. 1965. Veränderlichkeit des chromosomalen Polymorphismus bei *Drosophila subobscura*. *Genetica* **36**:83-104.

Burla, H., Jungen, H., and Bächli, G. 1986. Population structure of *Drosophila subobscura*: Non-random microdispersion of inversion polymorphism on a mountain slope. *Genetica* **70**:9-15.

Cáceres, M., Puig, M., and Ruiz, A. 2001. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res.* **11**:1353-1364.

Cáceres, M., Sullivan, R. T., Thomas, J. W., and National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program. 2007. A recurrent inversion on the eutherian X chromosome. *Proc. Natl. Acad. Sci. U.S.A.* **104**:18571-18576.

Calabria, G. 2012. Inversions cromosòmiques, clines i adaptació a *Drosophila subobscura*: aproximació mitjançant marcadors moleculars. PhD Thesis, Universitat de Barcelona.

Calabria, G., Dolgova, O., Rego, C., Castañeda, L. E., Rezende, E. L., Balanya, J., and Santos, M. 2012. *Hsp70* protein levels and thermotolerance in *Drosophila subobscura*: a reassessment of the thermal co-adaptation hypothesis. *J. Evolution. Biol.* **25**:691-700.

Carlini, D. B., and 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.

Casals, F., and Navarro, A. 2007. Chromosomal evolution: inversions: the chicken or the egg?. *Heredity* **99**:479-480.

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

Castañeda, L. E., Balanya, J., Rezende, E. L., and Santos, M. 2013. Vanishing chromosomal inversion clines in *Drosophila subobscura* from Chile: is behavioral thermoregulation to blame?. *Am. Nat.* **182**:249-259.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**:540-552.

Charlesworth, B. 2001. Genome analysis: more *Drosophila* Y chromosome genes. *Curr. Biol.* **11**:R182-R184.

Chen, B., Shilova, V. Y., Zatsepina, O. G., Evgen'ev, M. B., and Feder, M. E. 2008. Location of *P* element insertions in the proximal promoter region of *Hsp70A* is consequential for gene expression and correlated with fecundity in *Drosophila melanogaster*. *Cell Stress Chaperon.* **13**:11-17.

Chen, J. M., Cooper, D. N., Chuzhanova, N., Férec, C., and Patrinos, G. P. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat. Rev. Genet.* **8**:762.

Cheng C., White B. J., Kamdem C., Mockaitis K., Costantini C., Hahn M. W., Besansky, N.J. 2012. Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. *Genetics* **190**:1417–1432.

Clark, J. B., and Kidwell, M. G. 1997. A phylogenetic perspective on *P* transposable element evolution in *Drosophila. Proc. Natl. Acad. Sci. U.S.A.* **94**:11428-11433.

Colinet, H., Lee, S. F., and Hoffmann, A. 2010. Temporal expression of heat shock genes during cold stress and recovery from chill coma in adult *Drosophila melanogaster*. *FEBS J.* **277**:174-185.

Colinet, H., Overgaard, J., Com, E., and Sørensen, J. G. 2013. Proteomic profiling of thermal acclimation in Drosophila melanogaster. *Insect Biochem. Mol. Biol.* **43**:352-365.

Collin, J.E. 1936. Note. *Drosophila subobscura* n. sp. *J. Genet*. **33**:60.

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., and de Frutos, R. 1998. Ultrastructure of regions containing homologous loci in polytene chromosomes of *Drosophila melanogaster* and *Drosophila subobscura*. *Chromosoma* **107**:113-126.

Dahlgaard J, Loeschcke V, Michalak P, Justesen J. 1998. Induced thermotolerance and associated expression of the heat-shock protein Hsp70 in adult *Drosophila melanogaster*. *Funct. Ecol.* **12**:786–793.

Datta, A., Hendrix, M., Lipsitch, M., and Jinks-Robertson, S. 1997. Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc. Natl. Acad. Sci USA* **94**: 9757-9762.

Daveran-Mingot, M. L., Campo, N., Ritzenthaler, P., and Le Bourgeois, P. 1998. A Natural Large Chromosomal Inversion inLactococcus lactis Is Mediated by Homologous Recombination between Two Insertion Sequences. *J. Bacteriol.* **180**:4834-4842.

De Frutos R, Prevosti A. 1984. Temporal changes of chromosomal polymorphism in natural populations of *Drosophila subobscura*. *Genetica* **63**:181–187.

De Jong, L., Moreau, X., Jean, S., Scher, O., and Thiéry, A. 2006. Expression of the heat shock protein Hsp70 in chloride target cells of mayfly larvae from motorway retention pond: A biomarker of osmotic shock. *Sci. Total Environ.* **366**:164-173.

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

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.

Delprat, A., Negre, B., Puig, M., and Ruiz, A. 2009. The transposon *Galileo* generates natural chromosomal inversions in Drosophila by ectopic recombination. *PloS One* **4**:e7883.

Deng, W., Maust, B. S., Nickle, D. C., Learn, G. H., Liu, Y., Heath, L., Kosakovsky Pond, S.L., and 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. 1948. Genetics of natural populations. XVI. Altitudinal and seasonal changes produced by natural selection in certain populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. *Genetics* **33**:158-176.

Dobzhansky, T. 1950. Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura*. *Genetics* **35**:288-302.

Dobzhansky, T. 1962. Rigid vs. flexible chromosomal polymorphisms in *Drosophila*. *Am. Nat.* **96**:321-328.

Dobzhansky, T. 1970. Genetics of the evolutionary process. New York, NY: Columbia Univ. Press.

Dobzhansky, T., and Ayala, F. J. 1973. Temporal frequency changes of enzyme and chromosomal polymorphisms in natural populations of *Drosophila. Proc. Natl. Acad. Sci. U.S.A.* **70**:680-683.

Dobzhansky, T., and Epling, C. 1948. The suppression of crossing over in inversion heterozygotes of Drosophila pseudoobscura. *Proc. Natl. Acad. Sci. U.S.A.* **34**:137-141.

Dolgova, O., Rego, C., Calabria, G., Balanyà, J., Pascual, M., Rezende, E.L., Santos, M. 2010. Genetic constraints for thermal coadaptation in *Drosophila subobscura*. *BMC Evol. Biol*. **10**:363.

Dyson-Hudson, V. R. D. 1956. The daily activity rhythm of *Drosophila* subobscura and *D. obscura. Ecology* **37**:562-567.

Etges, W. J. 1984. Genetic structure and change in natural populations of *Drosophila robusta*: systematic inversion and inversion association frequency shifts in the Great Smoky Mountains. *Evolution* **38**:675-688.

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.

Evgen'ev, M. B., Zatsepina, O. G., Garbuz, D., Lerman, D. N., Velikodvorskaya, V., Zelentsova, E., and Feder, M. E. 2004. Evolution and arrangement of the *Hsp70* gene cluster in two closely related species of the virilis group of *Drosophila*. *Chromosoma* **113**:223-232.

Ezawa, K., Oota, S., and Saitou, N. 2006. Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Mol. Evol. Biol.* **23**:927-940.

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

Fay, J. C., and Wu, C. I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**:1405-1413.

Feder, J. H., Rossi, J. M., Solomon, J., Solomon, N., and Lindquist, S. 1992. The consequences of expressing *Hsp70* in *Drosophila* cells at normal temperatures. *Genes Dev.* **6**:1402-1413.

Feder, M. E., and Hofmann, G. E. 1999. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu. Rev. Physiol.* **61**:243-282.

Feder, M. E., and Krebs, R. A. 1998. Natural and genetic engineering of the heat-shock protein *Hsp70* in *Drosophila melanogaster*: consequences for thermotolerance. *Am. Zool.* **38**:503-517.

Feder, M. E., Cartaño, N. V., Milos, L., Krebs, R. A., and Lindquist, S. L. 1996. Effect of engineering *Hsp70* copy number on *Hsp70* expression and tolerance of ecologically relevant heat shock in larvae and pupae of *Drosophila melanogaster*. *J. Exp. Biol.* **199**:1837-1844.

Feder, M. E., Krebs, R. A. 1998. Natural and genetic engineering of the heat-shock protein *Hsp70* in *Drosophila melanogaster*: consequences for thermotolerance. *Am Zool.* **38**:503-517.

Felger, I., and Sperlich, D. 1989. Cytological localization and organization of dispersed middle repetitive DNA sequences of *Drosophila subobscura*. *Chromosoma* **98**:342-350.

Fernández-Palacios, J. M., de Nascimento, L., Otto, R., Delgado, J. D., García-del-Rey, E., Arévalo, J. R., and Whittaker, R. J. 2011. A reconstruction of Palaeo-

Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *J. Biogeogr*. **38**:226-246.

Feuk, L., MacDonald, J. R., Tang, T., Carson, A. R., Li, M., Rao, G., and Scherer, S. W. 2005. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS Genet.* **1**:e56.

Folk, D. G., Zwollo, P., Rand, D. M., and Gilchrist, G. W. 2006. Selection on knockdown performance in *Drosophila melanogaster* impacts thermotolerance and heat-shock response differently in females and males. *J. Exp. Biol.* **209**:3964-3973.

Fontdevila, A., Zapata, C., Alvarez, G., Sanchez, L., Méndez, J., and Enriquez, I. 1983. Genetic coadaptation in the chromosomal polymorphism of *Drosophila subobscura*. I. Seasonal changes of gametic disequilibrium in a natural population. *Genetics* **105**:935-955.

Fortune, P. M., Roulin, A., and Panaud, O. 2008. Horizontal transfer of transposable elements in plants. *Commun and Integr. Biol.* **1**:74-77.

Fu, Y. X., and Li, W. H. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**:693-709.

Galtier, N. 2003. Gene conversion drives GC content evolution in mammalian histones. *Trends Genet.* **19**:65-68.

Garbuz, D. G., and 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. *Russ. J. Genet.* **53**:21-38.

Garbuz, D. G., Molodtsov, V. B., Velikodvorskaia, V. V., Evgen'ev, M. B., and Zatsepina, O. G. 2002. Evolution of the response to heat shock in genus *Drosophila*. *Russ. J. Genet.* **38**:925-936.

Garbuz, D. G., Yushenova, I. A., Zatsepina, O. G., Przhiboro, A. A., Bettencourt, B. R., and Evgen'ev, M. B. 2011. Organization and evolution of *Hsp70* clusters strikingly differ in two species of Stratiomyidae (Diptera) inhabiting thermally contrasting environments. *BMC Evol. Biol.* **11**:74.

Garbuz, D., Evgen'ev, M. B., Feder, M. E., and Zatsepina, O. G. 2003. Evolution of thermotolerance and the heat-shock response: evidence from inter/intraspecific comparison and interspecific hybridization in the virilis species group of *Drosophila*. I. Thermal phenotype. *J. Exp. Biol.* **206**:2399-2408.

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.

García Guerreiro, M. G. 2012. What makes transposable elements move in the Drosophila genome?. *Heredity* **108**:461.

García Guerreiro, M.P.G. 2015. Changes of Osvaldo expression patterns in germline of male hybrids between the species *Drosophila buzzatii* and *Drosophila koepferae*. *Mol. Genet. Genomics.* **290**:1471-1483.

Garcia, G. C. 2012. Inversions cromosòmiques, clines i adaptació a *Drosophila subobscura*: aproximació mitjançant marcadors moleculars. PhD Thesis, Universitat de Barcelona.

Gonçalves, J. W., Valiati, V. H., Delprat, A., Valente, V. L., and Ruiz, A. 2014. Structural and sequence diversity of the transposon Galileo in the *Drosophila willistoni* genome. *BMC Genom.* **15**:792.

Gong, W. J., and Golic, K. G. 2004. Genomic deletions of the *Drosophila melanogaster* Hsp70 genes. *Genetics* **168**: 1467-1476.

Gong, W. J., and Golic, K. G. 2006. Loss of Hsp70 in *Drosophila* is pleiotropic, with effects on thermotolerance, recovery from heat shock and neurodegeneration. *Genetics* **172**:275-286.

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.

Govin, J., Caron, C., Escoffier, E., Ferro, M., Kuhn, L., Rousseaux, S., and Khochbin, S. 2006. Post-meiotic shifts in HSPA2/HSP70. 2 chaperone activity during mouse spermatogenesis. *J. Biol. Chem.* **281**:37888–37892.

Grant, P. R., Grant, B. R., Huey, R. B., Johnson, M. T. J., Knoll, A. H., and Schmitt, J. 2017. Evolution caused by extreme events. *Proc. R. Soc. Lond. [Biol]* **372**:20160146.

Graves, J. A. M. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* **124**:901-914.

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. *Syst. Biol.* **59**:307-321.

Haldane, J. B. S. 1940. *Keeping cool: and other essays*. Chatto and Windus. London.

Haldane, J. B. S. 1957. The conditions for coadaptation in polymorphism for inversions. *J. Genet.* **55**:218.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Series.* **41**:95-98.

Hashikawa, N., Mizukami, Y., Imazu, H., and Sakurai, H. 2006. Mutated yeast heat shock transcription factor activates transcription independently of hyperphosphorylation. *J. Biol. Chem.* **281**:3936-3942.

Hasson, E., Rodriguez, C., Fanara, J. J., Naveira, H., Reig, O. A., and Fontdevila, A. 1995. The evolutionary history of *Drosophila buzzatti*. XXVI. Macrogeographic patterns of inversion polymorphism in New World populations. *J. Evol. Biol.* **8**:369-384.

Herrig, D. K., Modrick, A. J., Brud, E., and 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., and Montooth, K. L. 2013. Inducing extra copies of the Hsp70 gene in *Drosophila melanogaster* increases energetic demand. *BMC Evol. Biol.* **13**:68.

Hoffmann, A. A., Sgrò, C. M., and Weeks, A. R. 2004. Chromosomal inversion polymorphisms and adaptation. *Trends Ecol. Evol.* **19**:482-488.

Hoffmann, A. A., and Rieseberg, L. H. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation?. *Annu. Rev. Ecol. Evol. Syst.* **39**:21-42.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**:1-44.

Hudson, R. R., Boos, D. D., and Kaplan, N. L. 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**:138-151.

Hudson, R. R., Kreitman, M., Aguadé, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**:153-159.

Huey, R. B., and Pascual, M. 2009. Partial thermoregulatory compensation by a rapidly evolving invasive species along a latitudinal cline. *Ecology* **90**:1715-1720.

Hughes, L. 2000. Biological consequences of global warming: is the signal already apparent? *Trends Ecol. Evol.* **15**:56-61.

Hunt, C., and Morimoto, R. I. 1985. Conserved features of eukaryotic hsp70 genes revealed by comparison with the nucleotide sequence of human hsp70. *Proc. Natl. Acad. Sci. U.S.A.* **82**:6455-6459.

Hurles, M. E., Willey, D., Matthews, L., and Hussain, S. S. 2004. Origins of chromosomal rearrangement hotspots in the human genome: evidence from the AZFa deletion hotspots. *Genome Biol.* **5**:R55.

Huson, D. H., & Bryant, I. D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **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.

Ireland, H. E., Harding, S. J., Bonwick, G. A., Jones, M., Smith, C. J., and Williams, J. H. 2004. Evaluation of heat shock protein 70 as a biomarker of environmental stress in *Fucus serratus* and *Lemna minor*. *Biomarkers* **9**:139-155.

Iriarte, P. F., Norry, F. M., and Hasson, E. R. 2003. Chromosomal inversions effect body size and shape in different breeding resources in *Drosophila buzzatii*. *Heredity* **91**:51-59.

Jedlicka, P., Mortin, M. A., and Wu, C. 1997. Multiple functions of *Drosophila* heat shock transcription factor in vivo. *EMBO J.* **16**:2452-2462.

Jensen, L. T., Cockerell, F. E., Kristensen, T. N., Rako, L., Loeschcke, V., McKechnie, S. W., and Hoffmann, A. A. 2010. Adult heat tolerance variation in *Drosophila melanogaster* is not related to Hsp70 expression. *J. Exp. Zool. A Ecol. Genet. Physiol.* **313**:35-44.

Jinks-Robertson, S., Michelitch, M., and Ramcharan, S. H. A. R. I. 1993. Substrate length requirements for efficient mitotic recombination in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **13**:3937-3950.

Jukes, T. H., and Cantor, C. R. 1969. Evolution of protein molecules. Mammalian protein metabolism. *New York Academic Press* (pg. 21-132).Katoh, K., and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**:772-780.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**:462-467.

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

Kapun, M., Fabian, D. K., Goudet, J., Flatt T. 2016. Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Mol. Biol. Evol.* **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.

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

Katoh, K., and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**:772-780.

Keightley, P. D., Ness, R. W., Halligan, D. L., and Haddrill, P. R. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics* **196**:313-320.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**:241-254.

Kenig, B., Jelić, M., Kurbalija, Z., Stamenković-Radak, M., and Anđelković, M. 2010. Inversion polymorphism in populations of *Drosophila subobscura* from urban and non-urban environments. *Arch. Biol. Sci.* **62**:565-574.

Kenig, B., Novičić, Z. K., Patenković, A., Stamenković-Radak, M., and Anđelković, M. 2015. Adaptive role of inversion polymorphism of *Drosophila subobscura* in lead stressed environment. *PloS One* **10**:e0131270.

Kennington, W. J., Hoffmann, A. A., and Partridge, L. 2007. Mapping regions within cosmopolitan inversion In (3R) Payne associated with natural variation in body size in *Drosophila melanogaster*. *Genetics* **177**:549-556.

Kent, W. J. 2002. BLAT – The BLAST-like alignment tool. *Genome Res.* **12**:656-664.

Khadem, M., and 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., and Segarra, C. 2012. Multilocus analysis of nucleotide variation in *Drosophila madeirensis*, an endemic species of the Laurisilva forest in Madeira. *J. Evol. Biol.* **25**:726-739.

Khadem, M., Rozas, J., Segarra, C., and Aguadé, M. 2001. DNA variation at the *rp49* gene region in *Drosophila madeirensis* and *D. subobscura* from Madeira: inferences about the origin of an insular endemic species. *J. Evol. Biol.* **14**:379-387.

Khadem, M., Rozas, J., Segarra, C., Brehm, A., and Aguadé, M. 1998. Tracing the colonization of Madeira and the Canary Islands by *Drosophila subobscura* through the study of the rp49 gene region. *J. Evol. Biol.* **11**:439-452.

Kidwell, M. G., and Lisch, D. R. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**:1-24.

Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press.

Kirkpatrick, M. 2010. How and why chromosome inversions evolve. *PLoS Biol.* **8**:e1000501.

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

Kirkpatrick, M., and Kern, A. 2012. Where's the money? Inversions, genes, and the hunt for genomic targets of selection. *Genetics* **190**:1153-1155.

Kohany, O., Gentles, A. J., Hankus, L., and Jurka, J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinform.* **7**:474.

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

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

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. W. 2006. *Mol. Biol. Evol.* **23**:1891–1901.

Koske, T., Smith, J.M. 1954. Genetics and cytology of *Drosophila subobscura*. X. The fifth linkage group. *J. Genet.* **52**:521-541.

Krebs, R. A. 1999. A comparison of Hsp70 expression and thermotolerance in adults and larvae of three *Drosophila* species. *Cell Stress Chaperon.* **4:** 243.

Krebs, R. A., and Feder, M. E. 1997a. Natural variation in the expression of the heat-shock protein HSP70 in a population of Drosophila melanogaster and its correlation with tolerance of ecologically relevant thermal stress. *Evolution* **51**:173-179.

Krebs, R. A., and Feder, M. E. 1997b. Deleterious consequences of Hsp70 overexpression in *Drosphila melanogaster* larvae. *Cell Stress Chaperon.* **2**:60-71.

Krebs, R. A., Roberts, S. P., Bettencourt, B. R., and Feder, M. E. 2001. Changes in thermotolerance and Hsp70 expression with domestication in *Drosophila melanogaster*. *J. Evol. Biol.* **14**:75-82.

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. 1993. *Drosophila subobscura*: Biology, Genetics and Inversion polymorphism. Verlag Dr. Kovac, Hamburg.

Krimbas, C. B., and 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.

Kristensen, T. N., Dahlgaard, J., and Loeschcke, V. 2002. Inbreeding affects Hsp70 expression in two species of *Drosophila* even at benign temperatures. *Evol. Ecol. Res.* **4**:1209-1216.

Kruskal, W. H., and Wallis, W. A. 1952. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**: 583-621.

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. *Mol. Biol. Evol.* **4**:1470-1480.

Kumar, S., Stecher, G., & Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **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.

Laayouni, H., García-Franco, F., Chávez-Sandoval, B. E., Trotta, V., Beltran, S., Corominas, M., and Santos, M. 2007. Thermal evolution of gene expression profiles in *Drosophila subobscura*. *BMC Evol. Biol.* **7**:42.

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

Lakhotia, S. C., and Prasanth, K. V. 2002. Tissue-and development-specific induction and turnover of hsp70 transcripts from loci 87A and 87C after heat shock and during recovery in *Drosophila melanogaster*. *J. Exp. Biol.* **205**:345-358.

Larruga, J. M., Cabrera, V. M., González, A. M., and 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., and Feder, M. E. 2003. Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. *Mol. Biol. Evol.* **20**:135-144.

Lerman, D. N., and Feder, M. E. 2004. Naturally occurring transposable elements disrupt hsp70 promoter function in *Drosophila melanogaster*. *Mol. Evol. Biol.* **22**:776-783.

Letunic, I., Bork, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* Nucleic acids research, 44.W1:W242-W245.

Levitan, M. 1992. Chromosomal variation in *Drosophila robusta* Sturtevant. *Drosophila inversion polymorphism.* Boca Raton, FL: CRC Press (pp. 221-338).

Levitan, M., and Etges, W. J. 2005. Climate change and recent genetic flux in populations of *Drosophila robusta*. *BMC Evol. Biol.* **5**:4.

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

Librado, P., & Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451-1452.

Lichten, M., and Haber, J. E. 1989. Position effects in ectopic and allelic mitotic recombination in Saccharomyces cerevisiae. *Genetics* **123***:*261-268.

Lindquist S, Craig, E.A. 1988. The heat-shock proteins. *Annu. Rev. Genet.* **22**, 631-677.

Lindquist, S. 1986. The heat-shock response. *Annu. Rev. Biochem.* **55**:1151-1191.

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.

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

Loukas, M., Vergini, V., and Krimbas, C. B. 1981. The genetics of *Drosophila subobscura* populations XVIII. Multiple insemination and sperm displacement in *Drosophila subobscura*. *Genetica* **57**:29-37.

Lyttle, T. W., and Haymer, D. S. 1992. The role of the transposable element hobo in the origin of endemic inversions in wild populations of *Drosophila melanogaster*. *Genetica* **86**:113-126.

Mano, S., and Innan, H. 2008. The evolutionary rate of duplicated genes under concerted evolution. *Genetics* **180**:493-505.

Mansai, S. P., Kado, T., and Innan, H. 2011. The rate and tract length of gene conversion between duplicated genes. *Genes* **2**:313-331.

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

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**:1-5.

Martínez-Sebastián, M. J., Latorre, A., and De Frutos, R. 1984. Chromosomal polymorphism of *Drosophila subobscura* populations sampled at different altitudes. *Genét. Ibér.* **36**:1-11.

Mathiopoulos, K. D., della Torre, A., Santolamazza, F., Predazzi, V., Petrarca, V., and Coluzzi, M. 1999. Are chromosomal inversions induced by transposable elements? A paradigm from the malaria mosquito *Anopheles gambiae*. *Parassitologia* **41**:119-123.

Maynard Smith J. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34**:126-129.

McDonald, J. H., and Kreitman, M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**:652-65.

McFarquhar, A. M., and Robertson, F. W. 1963. The lack of evidence for co-adaptation in crosses between geographical races of Drosophila subobscura Coll. *Genet. Res.* **4**:104-131.

McGarry, T. J., and Lindquist, S. 1985. The preferential translation of Drosophila *hsp70* mRNA requires sequences in the untranslated leader. *Cell* **42**:903-911.

Menozzi, P., Krimbas, C.B. 1992. The inversion polymorphism of *D. subobscura* revisited: synthetic maps of gene arrangement frequencies and their interpretation. *J. Evol. Biol.* **5**:625-641.

Mestres, F., Balañà, J., Segarra, C., Prevosti, A., and Serra, L. 1992. Colonization of America by *Drosophila subobscura:* Analysis of the O5 inversions from Europe and America and their implications for the colonizing process. *Evolution* **46**:1564-1568.

Mézard, C., Pompon, D., and Nicolas, A. 1992. Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. Cell 70:659-670.

Miller, W. J., Nagel, A., Bachmann, J., & Bachmann, L. 2000. Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. *Mol. Biol.Evol.* **17**:1597-1609.

Milne, K.J., Noble, E.G. 2008. Response of the myocardium to exercise: sex-specific regulation of *HSP70*. *Med.Sci. Sports. Exerc.* **40**:655-663.

Minh, B. Q., Nguyen, M. A. T., and von Haeseler, A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol.Evol.* **30**:1188-1195.

Mohn, N., and Spiess, E. B. 1963. Cold resistance of karyotypes in *Drosophila persimilis* from Timberline of California. *Evolution* **17**:548-563.

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

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., Martínez-Sebastián, M. J., and Frutos, R. D. 1992. Genetic analysis of heat shock response in three *Drosophila* species of the *obscura* group. *Genome* **35**:870-880.

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.

Morimoto, R. I. 1993. Cells in stress: transcriptional activation of heat shock genes. *Science* **259**:1409-1409.

Munté, A., Rozas, J., Aguadé, M., and Segarra, C. 2005. Chromosomal Inversion Polymorphism Leads to Extensive Genetic Structure. *Genetics* **169**:1573-1581.

Nadeau, D., Corneau, S., Plante, I., Morrow, G., and Tanguay, R. M. 2001. Evaluation for Hsp70 as a biomarker of effect of pollutants on the earthworm Lumbricus terrestris. *Cell Stress Chaperon.* **6**:153-163.

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

Navarro, A., Barbadilla, A., and Ruiz, A. 2000. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. *Genetics* **155**:685-698.

Navarro-Sabaté, A., Aguadé, M., and 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**:871-889.

Neal, S. J., Karunanithi, S., Best, A., So, A. K. C., Tanguay, R. M., Atwood, H. L., and Westwood, J. T. 2006. Thermoprotection of synaptic transmission in a *Drosophila* heat shock factor mutant is accompanied by increased expression of Hsp83 and DnaJ-1. *Physiol. Genomics* **25**:493-501.

Nei, M., and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418-426.

Nei, M., and Rooney, A. P. 2005. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**:121-152.

Nikesitch, N., and Ling, S. C. 2015. Molecular mechanisms in multiple myeloma drug resistance. *J. Clin. Pathol.* **69**:97-101.

Nikolaidis, N., and Nei, M. 2004. Concerted and nonconcerted evolution of the Hsp70 gene superfamily in two sibling species of nematodes. *Mol. Evol. Biol.* **21**:498-505.

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

Obbard, D. J., Maclennan, J., Kim, K. W., Rambaut, A., O'Grady, P. M., and Jiggins, F. M. 2012. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol. Evol. Biol.* **29**:3459-3473.

Omichinski, J. G., Pedone, P. V., Felsenfeld, G., Gronenborn, A. M., and Clore, G. M. 1997. The solution structure of a specific GAGA factor–DNA complex reveals a modular binding mode. *Nat. Struct. Mol. Biol.* **4**:122-132.

Orengo, D. J., and Prevosti, A. 1994. Preadult competition between *Drosophila subobscura* and *Drosophila pseudoobscura. J. Zool. Syst. Evol. Res.* **32**:44-50.

Orengo, D. J., and Prevosti, A. 1996. Temporal changes in chromosomal polymorphism of *Drosophila subobscura* related to climatic changes. *Evolution* **50**:1346-1350.

Orengo, D. J., Puerma, E., and Aguadé, M. 2016. Monitoring chromosomal polymorphism in *Drosophila subobscura* over 40 years. *Entomol. Sci.* **19**:215-221.

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

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

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

Parsell, D. A., and Lindquist, S. 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annu. Rev. Genet.* **27**:437-496.

Pascual, M., Aquadro, C. F., Soto, V., and Serra, L. 2001. Microsatellite variation in colonizing and palearctic populations of *Drosophila subobscura*. *Mol. Evol. Biol.* **18:** 731-740.

Pascual, M., Chapuis, M. P., Mestres, F., Balanya, J., Huey, R. B., Gilchrist, G. W., and Estoup, A. 2007. Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Mol. Ecol.* **16:** 3069-3083.

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

Pegueroles, C., Araúz, P. A., Pascual, M., and Mestres, F. 2010b. A recombination survey using microsatellites: the O chromosome of *Drosophila subobscura*. *Genetica* **138**:795-804.

Pegueroles, C., Ferrés-Coy, A., Martí-Solano, M., Aquadro, C. F., Pascual, M., and Mestres, F. 2016. Inversions and adaptation to the plant toxin ouabain shape DNA sequence variation within and between chromosomal inversions of *Drosophila subobscura. Sci. Rep.-UK*. **6**:23754.

Pegueroles, C., Ordóñez, V., Mestres, F., and Pascual, M. 2010a. Recombination and selection in the maintenance of the adaptive value of inversions. *J. Evol. Biol.* **23**:2709-2717.

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

Pentzos-Daponte, A. 1964. Qualitative und quantitative Untersuchungen über den chromosomalen Polymorphismus natürlicher Population von *Drosophila*

subobscura in der Umgebung von Thessaloniki/Griechenland. *Z. Vererbungsl.* **95**:129–144.

Pentzos-Daponte, A., and Sperlich, D. 1965. Vitalitätsprüfungen an intra-und interlokalen Heterozygoten vonDrosophila Subobscura. *Z. Vererbungsl.* **97**:46-54.

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

Perisic, O., Xiao, H., and 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.

Petersen, R.B., Lindquist, S. 1989. Regulation of synthesis by messenger RNA degradation. *Cell Regul.* **1**:135-149.

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.

Petrov, D. A., Lozovskaya, E. R., and Hartl, D. L. 1996. High intrinsic rate of DNA loss in Drosophila. *Nature* **384**: 346-349.

Petrov, D. A., and Hartl, D. L. 1998. High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups. *Mol. Evol. Biol.* **15**:293-302.

Pietzenuk, B., Markus, C., Gaubert, H., Bagwan, N., Merotto, A., Bucher, E., and Pecinka, A. 2016. Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. *Genome Biol.* **17**:209.

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

Pinsker, W., and Böhm, I. 1989. Allozyme divergence between different gene arrangements in Drosophila subobscura. *J. Evol. Biol.* **2**:353-366.

Posada, D., and Crandall, K. A. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. U S A* **98**:13757-13762.

Powell, J. R. 1997. *Progress and prospects in evolutionary biology: the Drosophila model*. Oxford University Press.

Powell, J.R. 1992. Inversion polymorphisms in *Drosophila pseudoobscura and Drosophila persimilis*. In: Krimbas CV, Powell JR, editors. *Drosophila inversion polymorphism*. Boca Raton .FL.: CRC Press; 1992. (pp. 73–126).

Prevosti, A. 1964. Chromosomal polymorphism in *Drosophila subobscura* populations from Barcelona .Spain.. *Genet. Res.* **5**:27-38.

Prevosti, A. 1967. Inversion heterozygosity and selection for wing length in *Drosophila subobscura*. *Genet. Res.* **10**:81-93.

Prevosti, A. 1971. Chromosomal polymorphism in *Drosophila subobscura* Coll. populations from the Canary Islands. *Genét. Ibér.* **23**:69–84.

Prevosti, A. 1972. Chromosomal polymorphism in *Drosophila subobscura* populations from the Madeira Island.*Genét. Ibér.* **24**:11–21.

Prevosti, A., Ribo, G., Serra, L., Aguade, M., Balaña, J., Monclus, M., and Mestres, F. 1988. Colonization of America by Drosophila subobscura: experiment in natural populations that supports the adaptive role of chromosomal-inversion polymorphism. *Proc. Natl. Acad. Sci. USA* **85**:5597-5600.

Prevosti, A., Serra, L., Ribo, G., Aguade, M., Sagarra, E., Monclus, M., and Garcia, M. P. 1985. The colonization of *Drosophila subobscura* in Chile. II. Clines in the chromosomal arrangements. *Evolution* **39**:838-844.

Prevosti, A., Serra, L., Segarra, C., Aguade, M., Ribo, G., and Monclus, M. 1990. Clines of chromosomal arrangements of *Drosophila subobscura* in South America evolve closer to Old World patterns. *Evolution* **44**:218-221.

Puerma, E., Orengo, D. J., and Aguadé, M. 2016. The origin of chromosomal inversions as a source of segmental duplications in the *Sophophora* subgenus of *Drosophila*. *Sci. Rep.* **6**:30715.

Puerma, E., Orengo, D. J., Salguero, D., Papaceit, M., Segarra, C., and Aguadé, M. 2014. Characterization of the breakpoints of a polymorphic inversion complex detects strict and broad breakpoint reuse at the molecular level. *Mol. Evol. Biol.* **31**:2331-2341.

Puig-Giribets, M., García Guerreiro, M. P., Santos, M., Ayala, F. J., Tarrío, R., and Rodríguez-Trelles, F. 2018. Chromosomal inversions promote genomic islands of concerted evolution of *Hsp70* genes in the *Drosophila subobscura* species subgroup. *Mol. Ecol.*, accepted. doi: org/10.1111/mec.14511.

Quintana, A., and Prevosti, A. 1991. Genetic and environmental factors in the resistance of *Drosophila subobscura* adults to high temperature shock. *Genetica* **84**:165-170.

Ramos-Onsins, S., Segarra, C., Rozas, J., and Aguadé, M. 1998. Molecular and chromosomal phylogeny in the obscura group of *Drosophila* inferred from sequences of the *rp49* gene region. *Mol. Phylogenet. Evol.* **9**:33-41.

Reddy, G.S., Krishnamurthy, N.B. 1974. Altitudinal gradients in frequencies of three common inversions in *Drosophila ananassae*. *Dros. Inf. Serv.* **51**:136–137.

Regner, L. P., Pereira, M. S. O., Alonso, C. E. V., Abdelhay, E., and Valente, V. L. S. 1996. Genomic distribution of *P* elements in *Drosophila willistoni* and a search for their relationship with chromosomal inversions. *J. Hered.* **87**:191-198.

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

Rego, C., Balanya, J., Fragata, I., Matos, M., Rezende, E. L., and 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.

Rezende, E. L., Balanyà, J., Rodríguez-Trelles, F., Rego, C., Fragata, I., Matos, M., and Santos, M. 2010. Climate change and chromosomal inversions in *Drosophila subobscura*. *Clim. Res.* **43**:103-114.

Rezende, E. L., Tejedo, M., and Santos, M. 2011. Estimating the adaptive potential of critical thermal limits: methodological problems and evolutionary implications. *Funct. Ecol.* **25**:111-121.

Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**:351-358.

Roberts, S. P., and Feder, M. E. 2000. Changing fitness consequences of hsp70 copy number in transgenic *Drosophila* larvae undergoing natural thermal stress. *Funct. Ecol.* **14**:353-357.

Rodríguez, C., Fanara, J.J., Hasson, E. 1999. Inversion polymorphism, longevity, and body size in a natural population of *Drosophila buzzatii*. *Evolution* **53**:612-620.

Rodríguez-Trelles, F. 2003. Seasonal cycles of allozyme-by-chromosomal-inversion gametic disequilibrium in *Drosophila subobscura*. *Evolution* **57**:839-848.

Rodríguez-Trelles, F., and Rodríguez, M. A. 2007. Comment on 'Global genetic change tracks global climate warming in *Drosophila subobscura*'. *Science* **315**:1497.

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

Rodríguez-Trelles, F., and Rodríguez, M. A. 1998. Rapid micro-evolution and loss of chromosomal diversity in *Drosophila* in response to climate warming. *Evol. Ecol.* **12**:829-838.

Rodríguez-Trelles, F., Tarrío, R., and Santos, M. 2013. Genome-wide evolutionary response to a heat wave in Drosophila. *Biol. Lett.* **9**:20130228.

Roeder, G. S. 1983. Unequal crossing-over between yeast transposable elements. *Mol. Gen. Genet.* **190**:117-121.

Romero-Soriano, V., and Guerreiro, M. P. G. 2016. Expression of the retrotransposon Helena reveals a complex pattern of TE Deregulation in Drosophila Hybrids. *PloS One* **11**:e0147903

Rozas, J., and Aguadé, M. 1990. Evidence of extensive genetic exchange in the rp49 region among polymorphic chromosome inversions in *Drosophila subobscura*. *Genetics* **126**:417-426.

Rozas, J., and Aguade, M. 1994. Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **91**:11517-11521.

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

Rozas, J., Segarra, C., Zapata, C., Alvarez, G., and Aguadé, M. 1995. Nucleotide polymorphism at the rp49 region of *Drosophila subobscura*: lack of geographic subdivision within chromosomal arrangements in Europe. *J. evol. biol.* **8**:355-367.

Rozen, S., Skaletsky, H., Marszalek, J. D., Minx, P. J., Cordum, H. S., Waterston, R. H., and Page, D. C. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**:873-876.

Russnak, R. H., and Candido, E. P. 1985. Locus encoding a family of small heat shock genes in *Caenorhabditis elegans:* two genes duplicated to form a 3.8-kilobase inverted repeat. *Mol. Cell. Biol.* **5**:1268-1278.

Russo, C. A., Takezaki, N., and Nei, M. 1995. Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* **12**:391-404.

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.

Santos, M. 2009. Recombination load in a chromosomal inversion polymorphism of Drosophila subobscura. *Genetics* **181**:803-809.

Santos, M., Céspedes, W., Balanya, J., Trotta, V., Calboli, F. C., Fontdevila, A., and Serra, L. 2005. Temperature-related genetic changes in laboratory populations of *Drosophila subobscura*: evidence against simple climatic-based explanations for latitudinal clines. *Am. Nat.* **165**:258-273.

Santos, M., Iriarte, P. F., Céspedes, W., Balanyà, J., Fontdevila, A., and Serra, L. 2004. Swift laboratory thermal evolution of wing shape .but not size. in *Drosophila subobscura* and its relationship with chromosomal inversion polymorphism. *J. Evol. Biol.* **17**:841-855.

Savković, V., Stamenković-Radak, M., and Andjelković, M. 2004. Diurnal variability of gene arrangement frequencies in *Drosophila subobscura* populations from two habitats. *J. Zool. Syst. Evol. Res.* **42**:208-214.

Sawyer, S. 1989. Statistical test for detecting gene conversion. *Mol. Biol. Evol.* **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.

Schmidt, H. A., Strimmer, K., Vingron, M., and von Haeseler, A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**:502-504.

Schmittgen, T.D., Livak, K. J. 2008. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3**:1101-1108.

Schwartz, A., Chan, D. C., Brown, L. G., Alagappan, R., Pettay, D., Disteche, C., and Page, D. C. 1998. Reconstructing hominid Y evolution: X-homologous block, created by X–Y transposition, was disrupted by Yp inversion through LINE—LINE recombination. *Hum. Mol. Genet.* **7**:1-11.

Segarra, C., Ribó, G., and Aguadé, M. 1996. Differentiation of Muller's chromosomal elements D and E in the obscura group of Drosophila. *Genetics* **144**:139-146.

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 Res.* **43**:W7-W14.

Serra i Camó, Ll. 2013. Antoni Prevosti i Maria Monclús: dos naturalistes vocacionals. *Semblances biogràfiques* **58**: 73-99.

Sharp, P. M., Stenico, M., Peden, J. F., and Lloyd, A. T. 1993. Codon usage: mutational bias, translational selection, or both?. *Biochem. Soc. T.* **21**:835–841.

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

Slotkin, R. K., and Martienssen, R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**:272-285.

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

Smith, J. M. 1956. Fertility, mating behaviour and sexual selection in *Drosophila subobscura*. *J. Genet.* **54**: 261-279.

Solé, E., Balanyà, J., Sperlich, D., and Serra, L. 2002. Long-term changes in the chromosomal inversion polymorphism of *Drosophila subobscura*. I. Mediterranean populations from southwestern Europe. *Evolution* **56**: 830-835.

Son, W. Y., Han, C. T., Hwang, S. H., Lee, J. H., Kim, S., and Kim, Y. C. 2000. Repression of hspA2 messenger RNA in human testes with abnormal spermatogenesis. *Fertil. Steril.* **73**:1138-1144.

Sørensen, J. G., and Loeschcke, V. 2001. Larval crowding in *Drosophila melanogaster* induces Hsp70 expression, and leads to increased adult longevity and adult thermal stress resistance. *J. Insect Physiol.* **47**:1301-1307.

Sørensen, J. G., and Loeschcke, V. 2002. Decreased heat-shock resistance and down-regulation of Hsp70 expression with increasing age in adult *Drosophila melanogaster*. *Funct. Ecol.* **16**:379-384.

Sørensen, J. G., Dahlgaard, J., and Loeschcke, V. 2001. Genetic variation in thermal tolerance among natural populations of *Drosophila buzzatii*: down regulation of Hsp70 expression and variation in heat stress resistance traits. *Funct. Ecol.* **15**:289-296.

Sørensen, J. G., Loeschcke, V., and Kristensen, T. N. 2013. Cellular damage as induced by high temperature is dependent on rate of temperature change–investigating consequences of ramping rates on molecular and organismal phenotypes in *Drosophila melanogaster*. *J. Exp. Biol*. **216**:809-814.

Sørensen, J.G., Michalak, P., Justesen, J., Loeschcke, V. 1999. Expression of the heat-shock protein *HSP70* in *Drosophila buzzatii* lines selected for thermal resistance. *Hereditas* **131**:155-164.

Sperlich, D., and Feuerbach, H. 1966. Ist der chromosomale strukturpolymorphismus von *Drosophila subobscura* stabil oder flexibel?. *Mol. Gen. Genet.* **98**:16-24.

Sperlich, D., and Pfriem, P. 1986. Chromosomal polymorphism in natural and experimental populations. In Ashburner, M., Carson, H.L., and Thompson, jr, J.N. (Eds.). *The Genetics and Biology of Drosophila.* Academic Press N.Y. (pp. 257-309).

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

Stalker, H. D. 1980. Chromosome studies in wild populations of *Drosophila melanogaster*. II. Relationship of inversion frequencies to latitude, season, wing-loading and flight activity. *Genetics* **95**:211-223.

Stanke, M., and Morgenstern, B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**:W465-W467.

Strand, D. J., and McDonald, J. F. 1985. *Copia* is transcriptionally responsive to environmental stress. *Nucleic Acids Res.* **13**:4401-4410.

Sugino, R. P., and Innan, H. 2006. Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. *Trends Genet.* **22**:642–644.

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

Sung, P., and Klein, H. 2006. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature reviews. Mol. Cell Biol.* **7**:739-750.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585-595.

Tamura, K., Subramanian, S., and Kumar, S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol. Biol. Evol.* **21**:36-44.

Tarrío, R., Rodríguez-Trelles F., and Ayala, F. J. 2001. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the drosophilidae. *Mol. Biol. Evol.* **8**:1464-1473.

Tian, S., Haney, R. A., and Feder, M. E. 2010. Phylogeny disambiguates the evolution of heat-shock cis-regulatory elements in Drosophila. *PLoS One* **5**: e10669.

Tower, J. 2011. Heat shock proteins and *Drosophila* aging. *Exp. Gerontol.* **46**:355-362.

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 Res.* **44**:W232-W235.

Trombetta, B., and Cruciani, F. 2017. Y chromosome palindromes and gene conversion. *Hum. Genet.* **136**:605-619.

Tsukiyama, T., Becker, P. B., and 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. Identification of novel chromatin structures flanking the heat-shock locus that may define the boundaries of higher order domains. *J. Mol. Biol.* **185**:341– 358.

Umina, P. A., Weeks, A. R., Kearney, M. R., McKechnie, S. W., and 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., and Leunissen, J. A. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**:W71-W74.

Velikodvorskaia, V. V., Lyozin, G. T., Feder, M. E., Evgen'ev, M. B. 2005. Unusual arrangement of the hsp68 locus in the virilis species group of *Drosophila* implicates evolutio nary loss of an hsp68 gene. *Genome* **48**:234-240.

Walser, J. C., Chen, B., and Feder, M. E. 2006. Heat-shock promoters: targets for evolution by *P* transposable elements in *Drosophila*. *PLoS Genet.* **2**:e165.

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

Wasserman, M. 1968. Recombination-induced chromosomal heterosis. *Genetics* **58**:125-139.

White, B. J., Hahn, M. W., Pombi, M., Cassone, B. J., Lobo, N. F., Simard, F., and Besansky, N. J. 2007. Localization of candidate regions maintaining a common polymorphic inversion 2La in *Anopheles gambiae*. *PLoS Genet.* **3**:e217.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., and Paux, E. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**:973-982.

Wilcoxon, F. 1992. Individual comparisons by ranking methods. In Breakthroughs in Statistics (pp. 196-202). Springer, New York, NY.

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

Williams, K. D., Helin, A. B., Posluszny, J., Roberts, S. P., and Feder, M. E. 2003. Effect of heat shock, pretreatment and hsp70 copy number on wing development in *Drosophila melanogaster*. *Mol. Ecol.* **12**:1165-1177.

Woolfit, M., and Bromham, L. 2005. Population size and molecular evolution on islands. *Proc. R. Soc. Lond. [Biol]* **272**:2277-2282.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**:1586–1591.

Zatsepina, O. G., Velikodvorskaia, V. V., Molodtsov, V. B., Garbuz, D., Lerman, D. N., Bettencourt, B. R., and Evgenev, M. B. 2001. A *Drosophila melanogaster* strain from sub-equatorial Africa has exceptional thermotolerance but decreased Hsp70 expression. *J. Exp. Biol.* **204**:1869-1881.

Zeng, K., Fu, Y. X., Shi, S., and Wu, C. I. 2006. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**:1431-1439.

Zivanovic, G., and Mestres, F. 2011. Changes in chromosomal polymorphism and global warming: The case of *Drosophila subobscura* from Apatin .Serbia.. *Genet. Mol. Biol.* **34**:489-495.

Zivanovic, G., Arenas, C., and Mestres, F. 2014. Inversion polymorphism in two Serbian natural populations of *Drosophila subobscura*: analysis of long-term changes. *Russ. J. Genet.* **50**:557-562.

Zivanovic, G., Arenas, C., and Mestres, F. 2016. Individual inversions or their combinations: which is the main selective target in a natural population of *Drosophila subobscura*?. *J. Evol. Biol.* **29**:657-664.

7. APPENDICES

7.1 Supplementary data of "Chromosomal inversions promote genomic islands of concerted evolution of *Hsp70* genes in the *Drosophila subobscura* species subgroup" (Chapter 3.1)

7.1.1 Figure S1: Cytological location of the *Hsp70IR* locus in the O_{ST}, O_{3+4}, O_{3+4+8} and O_{3+4+16} arrangements.

Figure S1. Results of the *in situ* hybridization of a *D. subobscura* 1.9 kb long *Hsp70 cds* probe to isogenic lines for the **A**. O_{ST}, **B**. O_{3+4}, **C**. O_{3+4+8} and **D**. O_{3+4+16} gene arrangements. Hybridization signals and their cytological location are indicated. C and T denote centromere and telomere of the O chromosome.

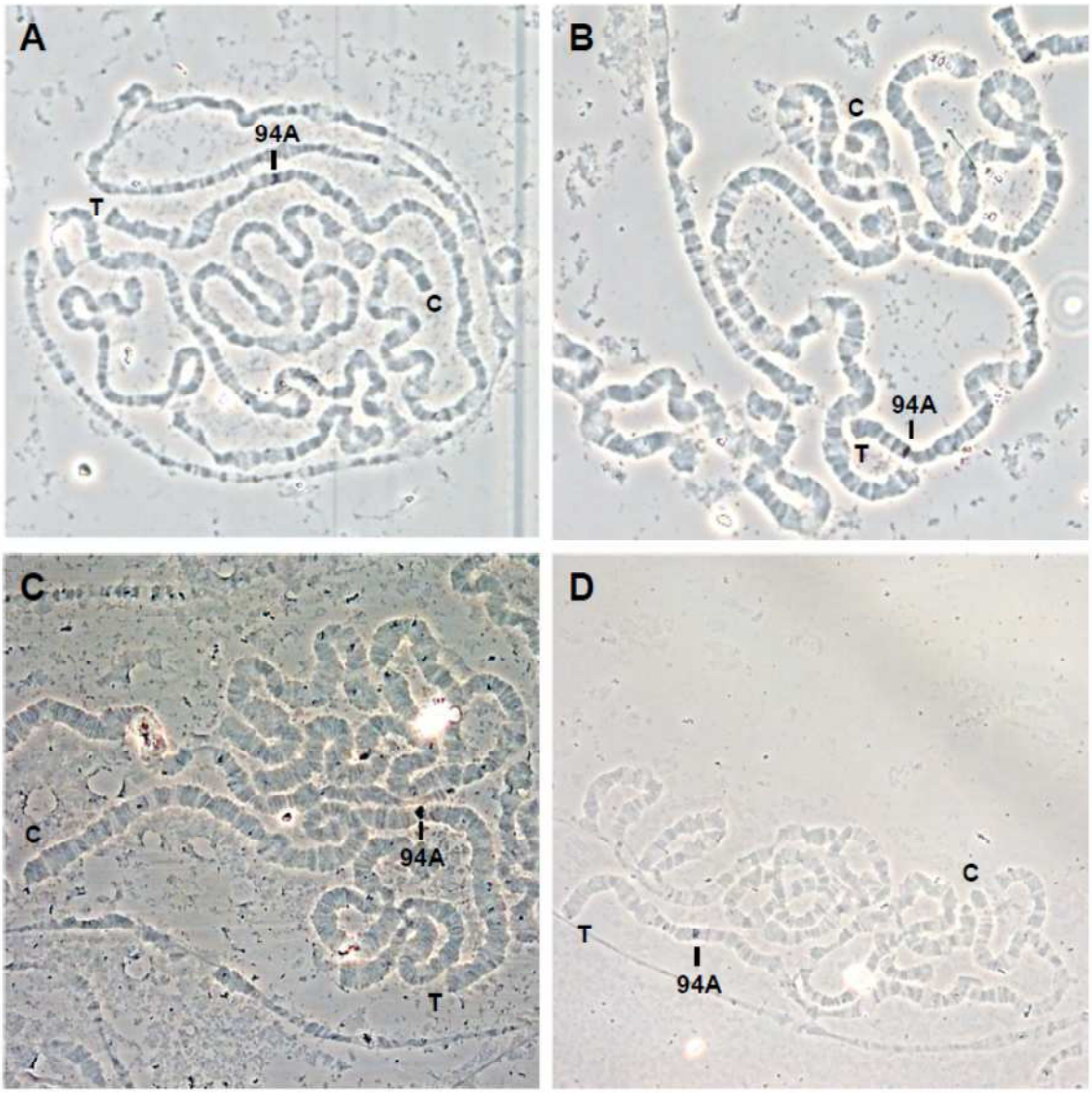


Figure S1

7.1.2 Table S1: Partitioning scheme data

Table S1. Optimal partitioning schemes and corresponding best-fitting substitution models selected using ModelFinder based on the Bayesian Information Criterion (BIC) for the evolution of the *Hsp70IR* region in the *Drosophila subobscura* species cluster. Shown are the partitioning schemes, the specific partition names and characters included the partitions, and the models.

Data set	Partitioning scheme	Partition	Character set	Best model
DNA	Unpartitioned		1-7042	TN93+I
	A priori	1: Hsp70B_3'	1-211	F81+I
		2: Hsp70B_3rd	212-2140\3	K3Pu+G4
		3: Hsp70B_2nd	213-2140\3	F81
		4: Hsp70B_1st	214-2140\3	F81
		5: Hsp70B_5'	2141-2641	F81
		6: Hsp70B_Hsp70A_IGC	2642-3843	HKY
		7: Hsp70A_5'	3844-4379	F81
		8: Hsp70A_1st	4380-6308\3	F81
		9: Hsp70A_2nd	4381-6308\3	F81
		10: Hsp70A_3rd	4382-6308\3	TPM+G4
		11: Hsp70A_3'	6309-6748	TPM2+I
		12: Dmt_3rd	6749-7042\3	F81
		13: Dmt_2nd	6750-7042\3	HKY
		14: Dmt_1st	6751-7042\3	JC
	Best merging	1: Hsp70B_3', Hsp70B_5', Hsp70B_Hsp70A_IGC, Hsp70A_5', Hsp70A_3', Dmt_2nd	1-211; 2141-2641; 2642-3843; 3844-4379; 6309-6748	HKY
		2: Hsp70B_1st, Hsp70A_1st, Dmt_1st	214-2140\3; 4380-6308\3; 6751-7042\3	F81
		3: Hsp70B_2nd, Hsp70A_2nd	213-2140\3; 4381-6308\3	F81
		4: Hsp70B_3rd, Hsp70A_3rd, Dmt_3rd	212-2140\3; 4382-6308\3; 6749-7042\3	K3Pu+G4
Mixed DNA + PROT	A priori	1: Hsp70B_3'	1-211	F81
		2: Hsp70B_PROT	NT2AA , 212-2140	JTTDCMut
		3: Hsp70B_5'	2141-2641	F81+I
		4: Hsp70B_Hsp70A_IGC	2642-3843	HKY
		5: Hsp70A_5'	3844-4379	F81
		6: Hsp70A_PROT	NT2AA , 4380-6308	JTT
		7: Hsp70A_3'	6309-6748	TPM2+I
		8: Dmt_PROT	NT2AA , 6749-7042	HIVb
	Best merging	1: Hsp70B_3', Hsp70B_5', Hsp70B_Hsp70A_IGC, Hsp70A_5', Hsp70A_3'	1-211; 2141-2641; 2642-3843; 3844-4379; 6309-6748	HKY + I
		2: Hsp70B_PROT, Hsp70A_PROT, Dmt_PROT	NT2AA , 212-2140; 4380-6308; 6749-7042	JTT
Mixed DNA + CODONS	A priori	1: Hsp70B_3'	1-211	F81+I
		2: Hsp70B_CODON	CODONS , 212-2140	MGK+F3X4+I
		3: Hsp70B_5'	2141-2641	F81
		4: Hsp70B_Hsp70A_IGC	2642-3843	HKY
		5: Hsp70A_5'	3844-4379	F81
		6: Hsp70A_CODON	CODONS , 4380-6308	MGK+F3X4+G4

		7: Hsp70A_3’	6309-6748	TPM2+I
		8: Dmt_CODON	CODONS , 6749-7042	SCHN05
	Best merging	1: Hsp70B_3’, Hsp70B_5’, Hsp70B_Hsp70A_IGC, Hsp70A_5’, Hsp70A_3’	1-211; 2141-2641; 2642-3843; 3844-4379; 6309-6748	HKY
		2: Hsp70B_CODON, Hsp70A_CODON, Dmt_CODON	CODONS , 212-2140; 4380-6308; 6749-7042	GY+F+I

7.1.3 Table S2 Parameters of partitioning schemes

Table S2. Optimal model selection for the evolution of the *Hsp70* ortholog and paralog genes in the *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 S2.

Data set	Partitioning scheme	P	lnL	BIC	*k*
NUC	Unpartitioned	1	-4,913.38	10,068.79	31
	A priori	4	-4,643.59	10,332.55	118
	Best merging	2	-4,754.45	10,049.34	61

7.1.4 Table S3 Best-fitting substitution models for different partitioning schemes

Table S3. Optimal partitioning schemes and corresponding best-fitting substitution models selected using ModelFinder based on the Bayesian Information Criterion (BIC) for the evolution of the *Hsp70* ortholog and paralog genes in the *Drosophila subobscura* species cluster. Shown are the partitioning schemes, the specific partition names and characters included the partitions, and the models.

Partitioning scheme	Partition	Character set	Best model
Unpartitioned		1-2459	TN93+I
A priori	1: Hsp70_5’	1-530	F81+I
	2: Hsp70_1st	531-2459\3	HKY+I
	3: Hsp70_2nd	532-2459\3	F81
	4: Hsp70_3rd	533-2459\3	K3Pu+G4
Best merging	1: Hsp70_5’	1-530	HKY+I
	2: Hsp70_1st, Hsp70_2nd, Hsp70_3rd	531-2459\3; 532-2459\3; 533-2459\3	K3Pu+G4

7.2 Supplementary data of "Quantification of *hsp70* mRNA and Hsp70 protein levels under non heat-shock conditions in the warm climate-associated O_{3+4} and the cold climate-associated O_{ST} arrangements of *Drosophila subobscura*" (Chapter 3.2)

7.2.1 Table S1: *Hsp70* mRNA relative expression levels in the complete data set

Table S1: List of samples used to quantify *hsp70* mRNA relative expression levels under basal conditions in the adaptive O_{ST} and O_{3+4} arrangements of *D. subobscura*. Two or three biological replicates have been made for six isogenic lines of the two arrangements and in the two sexes separately making up 70 samples. *hsp70* expression has been calculated using the comparative CT method described in Schmittgen and Livak (2008) and the ribosomal protein 49 (*rp49*) expression levels as reference.

LINE	REPLICATE	SEX	ARRANGEMENT	C_T increase (ΔC_T) (HSP70 C_T -RP49 C_T)	HSP70 relative expression ($2^{-\Delta CT}$)
OST (1)	1	♀	OST	8.17	0.0034744490
OST (1)	2	♀	OST	6.55	0.0106795900
OST (1)	3	♀	OST	7.84	0.0043674290
OST (2)	1	♀	OST	7.33	0.0062194380
OST (2)	2	♀	OST	6.84	0.0087348580
OST (3)	1	♀	OST	5.62	0.0203475660
OST (3)	2	♀	OST	4.88	0.0339840120
OST (3)	3	♀	OST	4.84	0.0349394330
OST (4)	1	♀	OST	7.36	0.0060914440
OST (4)	2	♀	OST	5.65	0.0199288190
OST (4)	3	♀	OST	8.61	0.0025611370
OST (5)	1	♀	OST	6.73	0.0094269060
OST (5)	2	♀	OST	5.02	0.0308411420
OST (5)	3	♀	OST	7.89	0.0042186580
OST (6)	1	♀	OST	4.69	0.0387677280
OST (6)	2	♀	OST	1.30	0.4064078010
OST (6)	3	♀	OST	5.41	0.0235357890
O3+4 (1)	1	♀	O3+4	5.12	0.0287758030
O3+4 (1)	2	♀	O3+4	9.15	0.0017614750
O3+4 (1)	3	♀	O3+4	7.65	0.0049822050
O3+4 (2)	1	♀	O3+4	8.01	0.0038819570
O3+4 (2)	2	♀	O3+4	7.82	0.0044283960
O3+4 (2)	3	♀	O3+4	7.54	0.0053769360
O3+4 (3)	1	♀	O3+4	5.56	0.0212116400

O3+4 (3)	2	♀	O3+4	5.94	0.0162998220
O3+4 (3)	3	♀	O3+4	4.96	0.0321508350
O3+4 (4)	1	♀	O3+4	5.83	0.0175912280
O3+4 (4)	2	♀	O3+4	5.42	0.0233732150
O3+4 (4)	3	♀	O3+4	4.39	0.0477286720
O3+4 (5)	1	♀	O3+4	5.61	0.0204890940
O3+4 (5)	2	♀	O3+4	5.41	0.0235357890
O3+4 (5)	3	♀	O3+4	9.03	0.0019142570
O3+4 (6)	1	♀	O3+4	5.88	0.0169920060
O3+4 (6)	2	♀	O3+4	5.43	0.0232117640
O3+4 (6)	3	♀	O3+4	6.48	0.0112105430
OST (1)	1	♂	OST	9.11	0.0018097423
OST (1)	2	♂	OST	6.19	0.0136969644
OST (1)	3	♂	OST	5.61	0.0204748969
OST (2)	1	♂	OST	5.13	0.0285572328
OST (2)	2	♂	OST	6.00	0.0156250000
OST (3)	1	♂	OST	8.11	0.0036194846
OST (3)	2	♂	OST	1.17	0.4444213406
OST (3)	3	♂	OST	7.72	0.0047429488
OST (4)	1	♂	OST	7.95	0.0040440036
OST (4)	2	♂	OST	7.01	0.0077585351
OST (4)	3	♂	OST	8.22	0.0033537712
OST (5)	1	♂	OST	5.22	0.0268301699
OST (5)	2	♂	OST	8.75	0.0023226701
OST (5)	3	♂	OST	5.74	0.0187106048
OST (6)	1	♂	OST	8.19	0.0034242411
OST (6)	2	♂	OST	6.80	0.0089742059
OST (6)	3	♂	OST	-0.19	1.1407637159
O3+4 (1)	1	♂	O3+4	9.27	0.0016197647
O3+4 (1)	2	♂	O3+4	11.91	0.0002598560
O3+4 (1)	3	♂	O3+4	7.46	0.0056795801
O3+4 (2)	1	♂	O3+4	7.10	0.0072893202
O3+4 (2)	2	♂	O3+4	7.09	0.0073400215
O3+4 (2)	3	♂	O3+4	7.64	0.0050133824
O3+4 (3)	1	♂	O3+4	4.77	0.0366510922
O3+4 (3)	2	♂	O3+4	7.06	0.0074942509
O3+4 (3)	3	♂	O3+4	7.96	0.0040160696
O3+4 (4)	1	♂	O3+4	8.48	0.0028006938
O3+4 (4)	2	♂	O3+4	8.97	0.0019941643
O3+4 (4)	3	♂	O3+4	7.71	0.0047759386
O3+4 (5)	1	♂	O3+4	7.06	0.0074942509
O3+4 (5)	2	♂	O3+4	5.72	0.0189717951
O3+4 (5)	3	♂	O3+4	9.73	0.0011775467

O3+4 (6)	1	♂	O3+4	8.40	0.0029603839
O3+4 (6)	2	♂	O3+4	9.55	0.0013340237
O3+4 (6)	3	♂	O3+4	8.83	0.0021973799

7.2.2 Table S2: Hsp70 protein levels in the complete data set

Supplementary Table S2: List of samples used to quantify basal Hsp70 protein by measuring absorbance at 490 nm in several lines of the adaptive O_{ST} and O_{3+4} arrangements of *D. subobscura*. Three biological replicates have been made for six isogenic lines of the two arrangements and in the two sexes separately making up 72 samples.

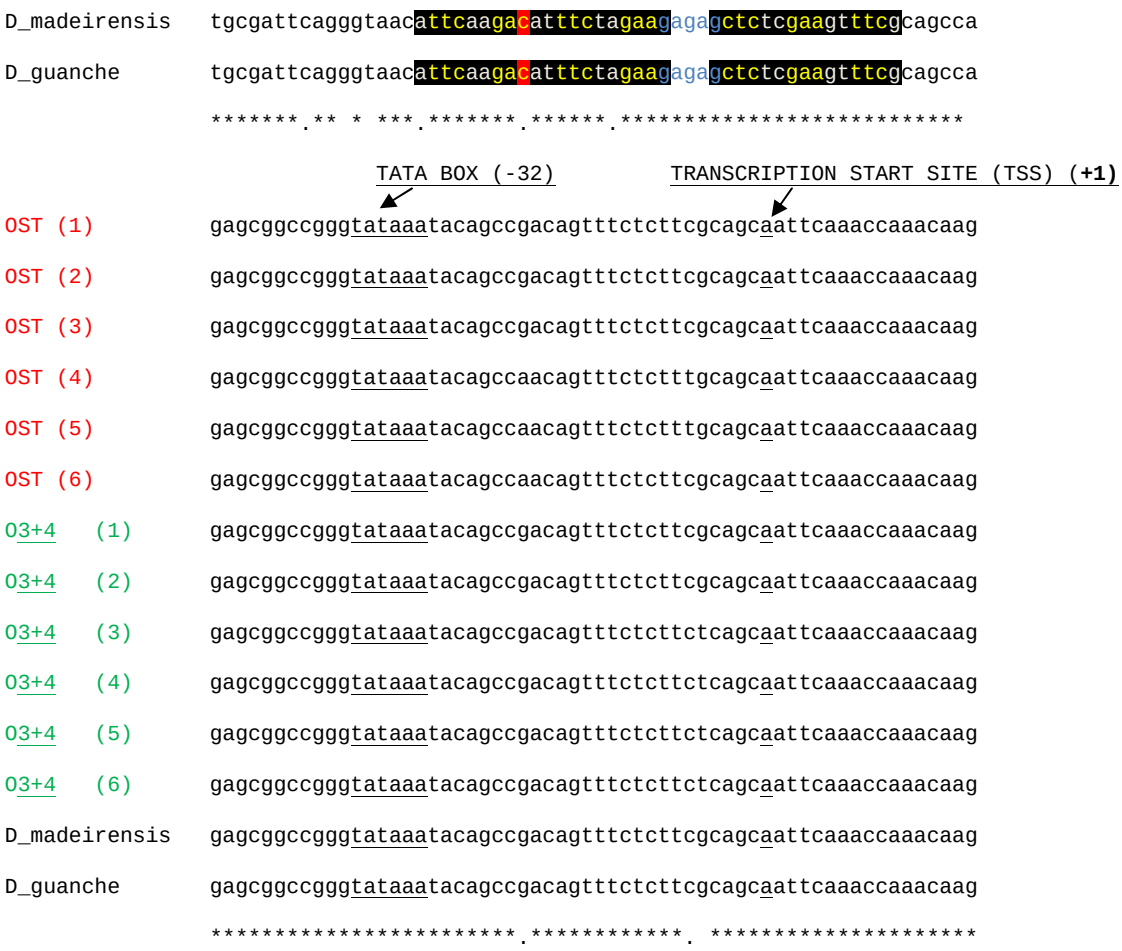
LINE	REPLICATE	SEX	ARRANGEMENT	Hsp70 protein levels (ABS 490 nm)
OST (1)	1	♀	OST	0.05167
OST (1)	2	♀	OST	0.03741
OST (1)	3	♀	OST	0.05230
OST (2)	1	♀	OST	0.05827
OST (2)	2	♀	OST	0.05230
OST (2)	3	♀	OST	0.05531
OST (3)	1	♀	OST	0.06338
OST (3)	2	♀	OST	0.04404
OST (3)	3	♀	OST	0.05639
OST (4)	1	♀	OST	0.04167
OST (4)	2	♀	OST	0.04139
OST (4)	3	♀	OST	0.04056
OST (5)	1	♀	OST	0.05062
OST (5)	2	♀	OST	0.05502
OST (5)	3	♀	OST	0.03829
OST (6)	1	♀	OST	0.04167
OST (6)	2	♀	OST	0.04404
OST (6)	3	♀	OST	0.05042
O3+4 (1)	1	♀	O3+4	0.05467
O3+4 (1)	2	♀	O3+4	0.04687
O3+4 (1)	3	♀	O3+4	0.05417
O3+4 (2)	1	♀	O3+4	0.04934
O3+4 (2)	2	♀	O3+4	0.04868
O3+4 (2)	3	♀	O3+4	0.05693
O3+4 (3)	1	♀	O3+4	0.03567
O O3+4 (3)	2	♀	O3+4	0.06129
O3+4 (3)	3	♀	O3+4	0.03943
O3+4 (4)	1	♀	O3+4	0.06593
O3+4 (4)	2	♀	O3+4	0.05465
O3+4 (4)	3	♀	O3+4	0.06236

O3+4 (5)	1	♀	O3+4	0.04679
O3+4 (5)	2	♀	O3+4	0.08516
O3+4 (5)	3	♀	O3+4	0.04396
O3+4 (6)	1	♀	O3+4	0.06867
O3+4 (6)	2	♀	O3+4	0.04777
O3+4 (6)	3	♀	O3+4	0.06617
OST (1)	1	♂	OST	0.04667
OST (1)	2	♂	OST	0.05863
OST (1)	3	♂	OST	0.06408
OST (2)	1	♂	OST	0.06593
OST (2)	2	♂	OST	0.05140
OST (2)	3	♂	OST	0.08367
OST (3)	1	♂	OST	0.04807
OST (3)	2	♂	OST	0.05465
OST (3)	3	♂	OST	0.04444
OST (4)	1	♂	OST	0.05367
OST (4)	2	♂	OST	0.05200
OST (4)	3	♂	OST	0.05871
OST (5)	1	♂	OST	0.04807
OST (5)	2	♂	OST	0.04958
OST (5)	3	♂	OST	0.05304
OST (6)	1	♂	OST	0.05967
OST (6)	2	♂	OST	0.05731
OST (6)	3	♂	OST	0.04227
O3+4 (1)	1	♂	O3+4	0.06467
O3+4 (1)	2	♂	O3+4	0.04687
O3+4 (1)	3	♂	O3+4	0.05304
O3+4 (2)	1	♂	O3+4	0.04169
O3+4 (2)	2	♂	O3+4	0.05230
O3+4 (2)	3	♂	O3+4	0.04499
O3+4 (3)	1	♂	O3+4	0.04767
O3+4 3)	2	♂	O3+4	0.04139
O3+4 (3)	3	♂	O3+4	0.04802
O3+4 (4)	1	♂	O3+4	0.04041
O3+4 (4)	2	♂	O3+4	0.04537
O3+4 (4)	3	♂	O3+4	0.03901
O3+4 (5)	1	♂	O3+4	0.04551
O3+4 (5)	2	♂	O3+4	0.05465
O3+4 (5)	3	♂	O3+4	0.05417
O3+4 (6)	1	♂	O3+4	0.05767
O3+4 (6)	2	♂	O3+4	0.05683
O3+4 (6)	3	♂	O3+4	0.05965

7.2.3 Figure S1: Identification of *cis*-regulatory elements in the proximal promoters of the *hsp70A* gene

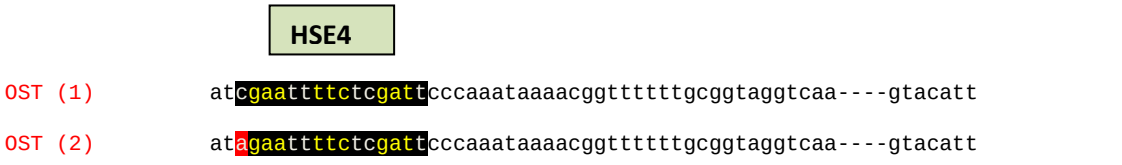
Supplementary Figure S1: Identification of several conserved CREs in the 5' *cis*-regulatory region of the *hsp70A* gene in 12 *D. subobscura* lines isogenic for the O_{ST} (6) and the O_{3+4} (6) adaptive gene arrangements and in the closely related *D. madeirensis* and *D. guanche*. The sequences of some of the lines shown in the aligment below have been described in Puig-Giribets et al. (2018) and are available in Genbank under the following accession numbers: OST (2): MG780233, OST (3): MG780234, *D. madeirensis:* MG780238, *D. guanche*: MG780239. Green boxes refer to four conserved HSEs described in Tian et al. (2010) in the *hsp70* proximal promoters of *Drosophila*. The three central nucleotides of each nGAAn/nTTCn unit have been colored in yellow. Nucleotides in red correspond to polymorphic sites. Red boxes represent conserved GAGA sites (G-: CTCTC; G+: GAGAG) whose sequences are colored in blue. The TATA box and the transcription start site (TSS) sequences have been underlined.

```
O3+4   (3)      gcacagagaaaaatcgagaaatttcgttcacaattcacctctcttacaaacacaagccc
O3+4   (4)      gcacagagaaaaatcgagaaatttcgttcacaattcacctctcttacaaacacaagccc
O3+4   (5)      gcacagagaaaagtcgagaaatttcgtccaaaatttacctctcttacaaacacaagccc
O3+4   (6)      gcacagagaaaagtcgagaaatttcgtcaacaa-tcacctctctt-ccaacacaagctg
D_madeirensis   gcagagagaaaaatcgagaaatttcgttcacaattcacctctcttacaaacacaaaccc
D_guanche       gcacagagaaaagtcgagaaatttcgttcacaattcacctctcttacgaacacaaaccc
                *** ********.**************. * ** *.********** * **.****.*.

                      G-

OST (1)         cttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgca
OST (2)         cttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgca
OST (3)         cttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgca
OST (4)         cttgcgttctctctgctgccagctgccagctgccaagacttgttgtctcgctctgacgca
OST (5)         cttgcgttctctctgctgccagctgccagctgccaagacttgttgtctcgctctgacgca
OST (6)         cttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgca
O3+4   (1)      cttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgca
O3+4   (2)      cttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgca
O3+4   (3)      attgtgctctctct-------gctgccagctgccaagacttgtcgtctcgctctgacgca
O3+4   (4)      attgtgctctctct-------gctgccagctgccaagacttgtcgtctcgctctgacgca
O3+4   (5)      attgtgctctcttt-------gctgccagctgccaagacttgtcgtctcgctctgacgca
O3+4   (6)      catgcgttctctct-------gctgccagctgccaagacttgtcgtctcgctctgacgca
D_madeirensis   attgtgctctatct-------gctgccagctgccaacacttgttgtctcgctctgacgca
D_guanche       attgtgctctatct-------gctgccagctgccaacacttgttgtctcgctctgacgca
                  **.*.*** *.*       *************** ******.****************

                                    HSE2          G+        HSE1
OST (1)         tgcgatttagtgtaacattcaagatatttctagaagagagctctcgaagtttcgcagcca
OST (2)         tgcgatttagtgaaacattcaagatatttctagaagagagctctcgaagtttcgcagcca
OST (3)         tgcgattcagggtaacgttcaagatatttctagaagagagctctcgaagtttcgcagcca
OST (4)         tgcgattcagggtaacgttcaagatatttctagaagagagctctcgaagtttcgcagcca
OST (5)         tgcgattcagggtaacgttcaagatatttctggaagagagctctcgaagtttcgcagcca
OST (6)         tgcgattcagggtaacgttcaagatatttctagaagagagctctcgaagtttcgcagcca
O3+4   (1)      tgcgatttagtgtaacattcaagatatttctagaagagagctctcgaagtttcgcagcca
O3+4   (2)      tgcgatttagtgtaacattcaagatatttctagaagagagctctcgaagtttcgcagcca
O3+4   (3)      tgcgattcagggtaacattcaagatatttctagaagagagctctcgaagtttcgcagcct
O3+4   (4)      tgcgattcagggtaacattcaagatatttctagaagagagctctcgaagtttcgcagcct
O3+4   (5)      tgcgattcagggtaacattcaagatatttctagaagagagctctcgaagtttcgcagcct
O3+4   (6)      tgcgattcagggtaacattcaagatatttctagaagagagctctcgaagtttcgcagcct
```



7.2.4 Figure S2: Identification of cis-regulatory elements in the proximal promoters of the *hsp70B* gene

Supplementary Figure S2: Identification of several conserved CREs in the 5' *cis*-regulatory region of the *hsp70B* gene in 12 *D. subobscura* lines isogenic for the O_{ST} (6) and the O_{3+4} (6) adaptive gene arrangements and in the closely related *D. madeirensis* and *D. guanche*. The sequences of some of the lines shown in the aligment below have been described in Puig-Giribets et al. (2018) and are available in Genbank under the following accession numbers: OST (2): MG780233, OST (3): MG780234, *D. madeirensis:* MG780238, *D. guanche*: MG780239. Green boxes refer to four conserved HSEs described in Tian et al. (2010) in the *hsp70* proximal promoters of *Drosophila*. The three central nucleotides of each nGAAn/nTTCn unit have been colored in yellow. Nucleotides in red correspond to polymorphic sites. Notice that the *D. guanche* allele has two extra pentanucleotide units in the HSE2 (see Puig-Giribets et al. (2018) for further information). Red boxes represent conserved GAGA sites (G-: CTCTC; G+: GAGAG) whose sequences are colored in blue. The TATA box and the transcription start site (TSS) sequences have been underlined.



```
OST (3)          atcgaattttctcgattcccaaataaaac-gttttttgcggtaggtcaa----gtacatt
OST (4)          atcgaattttctcgattcccaaataaaac-gttttttgcggtaggtcaa----gtacatt
OST (5)          atcgaattttctcgattcccaaataaaac-gttttttgcggtaggtcaa----gtacatt
OST (6)          atcgaattttctcgattcccaaataaaacgttttttgcggtaggtcaa----gtacatt
O3+4   (1)       atcgaattttctcgattcccaaataaaac-gttttttgcggtaggtcaa----gtacatt
O3+4   (2)       atcgaattttctcgattcccaaataaaacggttttttgcggtaggtcaa----gtacatt
O3+4   (3)       atcgaattttctcgattcccaaataaaacgggttttttgaggtaggtcaa----gtacatt
O3+4   (4)       atcgaattttctcgattcccaaataaaacggttttttgaggtaggtcaa----gtacatt
O3+4   (5)       atcgaattttctcgattcccaaataaaacggttttttgaggtaggtcaa----gtacatt
O3+4   (6)       atcgaattttctcgattcccaaataaaacgggttttttgaggtaggtcaa----gtacatt
D_madeirensis    atcgaattttctagattcccaaataaaacggttttttgcggtaggtcaa----gtacatt
D_guanche        atcgaaatttctggattcccaaataaaacggtttgttgtggtaggtcaagtatgtacatt
                 ** *** ***** ****************   ** *** **********    *******

                          HSE3                        G-
OST (1)          ggcacagagaaaagtcgagaaatttcgtcaaaaaatcaccctctctt-ccaacacaagct
OST (2)          ggcacagagaaaagtcgagaaatttcgtcaacaaatcaccctctctt-ccaacacaagct
OST (3)          ggcacagagaaaagtcgagaaatttcgtcaacaaatcaccctctctt-ccaacacaagct
OST (4)          ggcacagagaaaagtcgagaaatttcgtcaacaaatcaccctctctt-ccaacacaagct
OST (5)          ggcacagagaaaagtcgagaaatttcgtcaacaaatcaccctctctt-ccaacacaagct
OST (6)          ggcacagagaaaagtccagaaatttcgtcaacaaatcaccctctctt-ccaacacaagct
O3+4   (1)       ggcacagagaaaagtcgagaaatttcctcaacaaatcaccctctctt-ccaacacaagct
O3+4   (2)       ggcacagagaaaagtcgagaaatttcgtcaacaaatcaccctctctt-ccaatacaagct
O3+4   (3)       ggcacagagaaaagtcgagaaatttcgtccacaatttacactctcttacaaacacaagcc
O3+4   (4)       ggcacagagaaaagtcgagaaatttcgtccacaattcaccctctcttacaaacacaagcc
O3+4   (5)       ggcacagagaaaagtcgagaaatttcgtccacaatttaccctctcttacaaacacaagcc
O3+4   (6)       ggcacagagaaaagtcgagaaatttcgtcaacaa-tcaccctctctt-ccaacacaagct
D_madeirensis    ggcacagagaaaaatcgagaaatttcgttcacaattcaccctctcttacaaacacaaacc
D_guanche        ggcagagagaaaagtcgagaaatttcgtccacaattcaccctctctt-ccaacacaaacc
                 **** ********.** ********* *. * ** *.** ******* * **.****.*.

                       G-
OST (1)          gcttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgc
OST (2)          gcttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgc
OST (3)          gcttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgc
OST (4)          gcttgcgttctctctgctgccagctgccagctgccaagacttattgtctcgctctgacgc
OST (5)          gcttgcgttctctctgctgccagctgccagctgccaagacttgttgtctcgctctgacgc
OST (6)          gcttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgc
```

```
O3+4   (1)      gcatgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgc
O3+4   (2)      gcttgcgttctctct-------gctgccagctgccaagacttgttgtctcgctctgacgc
O3+4   (3)      cattgtgctctctct-------gctgccagctgccaagacttgtcgtctcgctctgacgc
O3+4   (4)      cattgtgctctcttt-------gctgccagctgccaagacttgtcgtctcgctctgacgc
O3+4   (5)      cattgtgctctcttt-------gctgccagctgccaagacttgtcgtctcgctctgacgc
O3+4   (6)      gcatgcgttctctct-------gctgccagctgccaagacttgtcgtctcgctctgacgc
D_madeirensis   cattgtgctctatct-------gctgccagctgccaacacttgttgtctcgctctgacgc
D_guanche       cattgtgctctctct-------gctgccagctgccaagacttgttgtctcgctctgacgc
                   **.*.*** *.*       *************** ****.*.***************
```

HSE2 G+ HSE1

```
OST (1)         atgcgatttagtgtaacattcaaga----------tatttctagaagagaggctctcgaag
OST (2)         atgcgatttagtgtaacattcaaga----------tatttctagaagagaggctctcgaag
OST (3)         atgcgattcagggtaacgttcaaga----------tatttctagaagagaggctctcgaag
OST (4)         atgcgatttagtgtaacattcaaga----------tatttctagaagagaggctctcgaag
OST (5)         atgcgattcagggtaacgttcaaga----------tatttctagaagagaggctctcgaag
OST (6)         atgcgattcagggtaacgttcaaga----------tatttctagaagagaggctctcgaag
O3+4   (1)      atgcgatttagtgtaacattcaaga----------tatttctagaagagaggctctcgaag
O3+4   (2)      atgcgatttagtgtaacattcaaga----------tatttctagaagagaggctctcgaag
O3+4   (3)      atgcgattcagggtaacattcaaga----------tatttctagaagagaggctctcgaag
O3+4   (4)      atgcgattcagggtaacattcaaga----------tatttctagaagagaggctctcgaag
O3+4   (5)      atgcgattcagggtaacattcaaga----------tatttctagaagagaggctctcgaag
O3+4   (6)      atgcgattcagggtaacattcaaga----------tatttctagaagagaggctctcgaag
D_madeirensis   atgcgattcagggtaacattcaaga----------tatttctagaagagaggctctcgaag
D_guanche       atgcgattcagggtaacattcaagaacattcattctatttctagaagagaggctctcgaag
                ********.** *****.*******          *************************
```

TATA BOX (-32) TRANSCRIPTION START SITE (TSS) (+1)

```
OST (1)         tttcgcagccagagcggccgggtataaatacagccgacagtttctcttcgcagcaattca
OST (2)         tttcgcagccagagcggccgggtataaatacagccgacagtttctcttcgcagcaattca
OST (3)         tttcgcagccagagcggccgggtataaatacagccgacagtttctcttcgcagcaattca
OST (4)         tttcgcagccagagcggccgggtataaatacagccgacagtttctcttcgcagcaattca
OST (5)         tttcgcagccagagcggccgggtataaatacagccaacagtttctcttcgcagcaattca
OST (6)         tttcgcagccagagcggccgggtataaatacagccaacagtttctcttcgcagcaattca
O3+4   (1)      tttcgcagccagagcggccgggtataaatacagccgacagtttctcttcgcagcaattca
O3+4   (2)      tttcgcagccagagcggccgggtataaatacagccgacagtttctcttcgcagcaattca
O3+4   (3)      tttcgcagcctgagcggccgggtataaatacagccgacagtttctcttctcagcaattca
O3+4   (4)      tttcgcagcctgagcggccgggtataaatacagccgacagtttctcttctcagcaattca
O3+4   (5)      tttcgcagcctgagcggccgggtataaatacagccgacagtttctcttctcagcaattca
```

```
O3+4   (6)      tttcgcagcctgagcggccgggtataaatacagccgacagtttctcttctcagcaattca
D_madeirensis   tttcgcagccagagcggccgggtataaatacagccgacagtttctcttcgcagcaattca
D_guanche       tttcgcagccagagcggccgggtataaatacagccgacagtttctcttcgcagcatttca
                ********** ************************.************* ***** ****
```

7.3 Supplementary data of "A complex interplay between stability, convergence and divergence drives the evolution of the *Hsp70IR* locus in *Drosophila subobscura"* (Chapter 3.3)

7.3.1 Table S1: Origin of the lines and species used in this work

Supplementary table S1: Shown are the geographical location (coordinates) and time of capture of wild *D. madeirensis* females and wild *D. subobscura* males used to establish the 12 isogenic lines. The *D. guanche* strain is derived from material provided by the San Diego Stock Center (Stock ID: 14011-0095.01) and is therefore not shown in the table. The *D. madeirensis* individuals were provided by Dr. Pilar García-Guerreiro.

Line/species name	Geographical location (coordinates)	Season and year
OST_1	"Berbikiz" (43°11′20.31″ N, 3°5′23.74″W)	autumn 2012
OST_2	"Berbikiz" (43°11′20.31″ N, 3°5′23.74″W)	late summer 2012
OST_3	"Berbikiz" (43°11′20.31″ N, 3°5′23.74″W)	autumn 2012
OST_4	"Berbikiz" (43°11′20.31″ N, 3°5′23.74″W)	spring 2012
O3+4_1	"Berbikiz" (43°11′20.31″ N, 3°5′23.74″W)	spring 2012
O3+4_2	"Berbikiz" (43°11′20.31″ N, 3°5′23.74″W)	autumn 2012
O3+4_3	"Berbikiz" (43°11′20.31″ N, 3°5′23.74″W)	early summer 2012
O3+4_4	"Berbikiz" (43°11′20.31″ N, 3°5′23.74″W)	early summer 2012
O3+4_5	"Berbikiz" (43°11′20.31″ N, 3°5′23.74″W)	autumn 2012
O3+4+8	"Vélez de Benaudalla" (36°50'23.66"N, 3°30'59.32"W)	spring 2014
O3+4+16_1	"Jérez del Marquesado" (37°11'6.74"N, 3°10'30.37"W)	late summer 2014
O3+4+16_2	"Juviles" (43°38'19.39"N, 116°14'28.86"W)	early summer 2014
D. madeirensis	"Ribeiro Frio" (32°43'00" N, 16°52'00" W)	autumn 2011

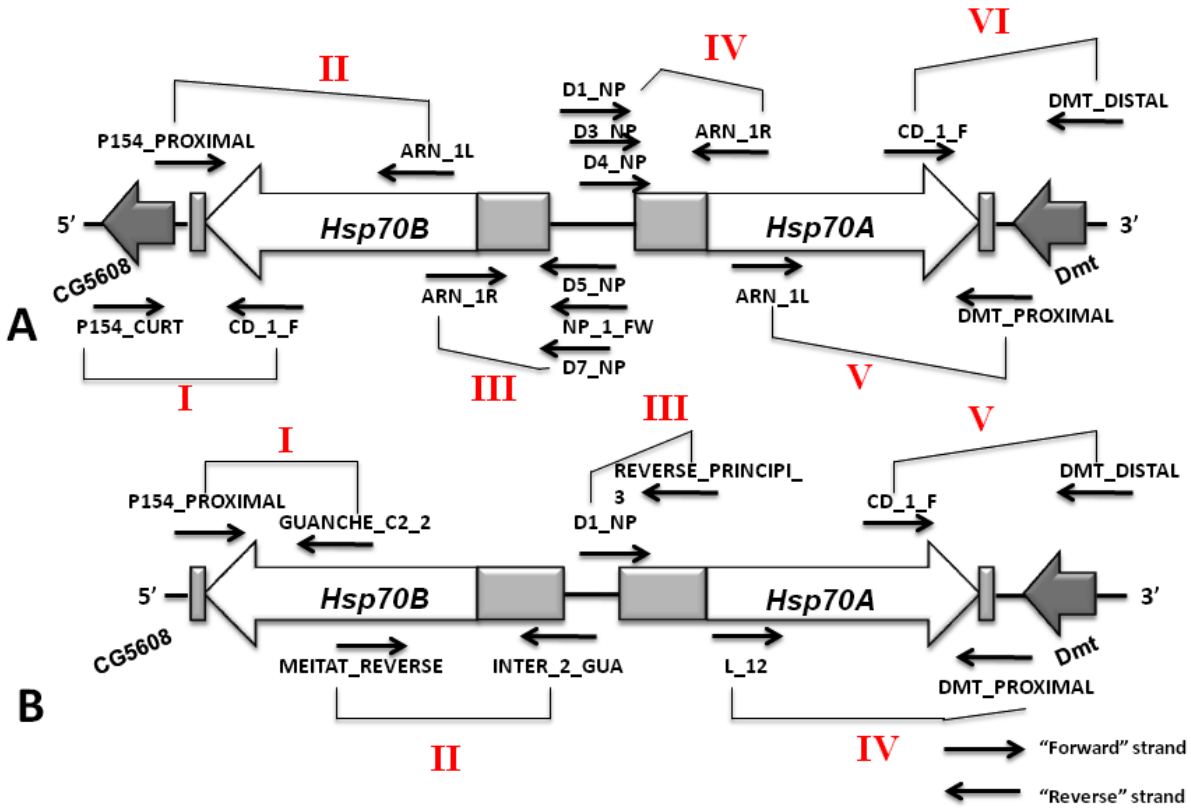
7.3.2 Table S2: Primers used to PCR-amplify the *Hsp70IR* locus in the different lines and species

Supplementary table S2: List of primers used to PCR-amplify a 9-10 kb genomic region at the *Hsp70IR* locus in the four O-chromosome arrangements of *D. subobscura* and in the closely related species *D. madeirensis* and *D. guanche.*

Primer name	Primer sequence (5’ ---> 3’)
P154_CURT	5' CCCGAGTTGTTCTTGTGGTT 3
P154_PROXIMAL	5' ACCCGCAAAATTGAACCCAA 3
ARN_1R	5' TGTGATGCTTTGGCCCAGAT 3
ARN_1L	5' CACAGTCTTTGACGCCAAGC 3
CD_1_F	5' ACCATACAGAACGACAAGGGT 3
MEITAT_REVERSE	5' CTTTGGTCGCCACTGAGGAT 3
REVERSE_PRINCIPI_3	5' TTTCCGTGCTGGTAGACACC 3
D1_NP	5' CTTCGGTGGGTTGGATTGTAGATTC 3
D3_NP	5' CTACTTTCTCTCCTGTGTATGTCTGG 3
D4_NP	5’ CTCTCCTGTGTACATATGTCTTGCG 3’
D5_NP	5' GATCCTCAAGGAATGCGAAA 3
D7_NP	5’ CTGTAGATCTTCAGGCATAAGCTG 3’
NP_1_FW	5’ GTATTGTGGCTTCTTAACGAGGTTCC 3’
L_12	5' TGGTATTGACTTGGGCACCAC 3
INTER_2_GUA	5' GGGAGACGGCATAGCATTTA 3
GUANCHE_C2_2	5' ATGCCAAGATGGACAAGAGC 3
DMT_PROXIMAL	5' AGTCGGAATTGTGAAGCCTT 3
DMT_DISTAL	5' GCTGCACGAGAAACTGCA 3

7.3.3 Figure S1: Location of primer combinations used to PCR-amplify different regions across *Hsp70IR* locus

Supplementary figure S1. Location of primer combinations used to PCR-amplify several 1.5-2.0-kb long overlapping segments (indicated in roman numerals) to reconstruct a 9-10 kb region at the *Hsp70IR* locus in the 14 lines used in this work. **A:** Primer combinations used for all *D. subobscura* and *D. madeirensis* lines. **B**: Primer combinations used to reconstruct a ~9-kb long region in *D. guanche*. PCR amplifications could not be performed downstream the *hsp70B* copy in the latter species with the available primers. Arrows indicate transcription direction relative to a 5' --> 3' oriented strand. Although most primers worked in all *D. subobscura* lines, a specific primer had to be used to amplify the targeted region in the following lines: *D3_NP* (O3+4_3, O3+4_2, O3+4+8); *D4_NP* (OST_3); *D7_NP* (O3+4+16_2); *NP_1_FW* (OST_1, OST_3).



7.3.4 Table S3: Neutrality tests and estimates of nucleotide diversity across the *Hsp70IR* locus

Supplementary table S3. No departure from neutrality is observed in any of the seven delimited regions at the *Hsp70IR* locus. The estimates of nucleotide diversity and several neutrality tests have been performed using a 5,978 bp cured alignment in the O_{3+4} phylad (including all O_{3+4+16}, O_{3+4} and O_{3+4+8} lines), O_{3+4} phylad[§] (analyses after excluding the recombinant O3+4_1 line) and in the complete *D. subobscura* data set. *S*, number of segregating sites. η (eta), total number of mutations. ∏, nucleotide diversity; ∏(a), nonsynonymous nucleotide diversity; ∏(s), silent nucleotide diversity. **3'B**: 3'UTR of the *hsp70B* copy, **CDS_B**: coding region of the *hsp70B* copy. **5'B**: 5'UTR of the *hsp70B* copy. **IGR:** intergenic nonduplicated region. **5'A**: 5'UTR of the *hsp70A* copy, **CDS_A**: coding region of the *hsp70A c*opy. **3'A**: 3'UTR of the *hsp70A* copy. ***Hsp70IR***: includes the genomic region comprising the two *hsp70* copies and the IGR. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; (two-tailed). JC: values after Jukes-Cantor distance correction (Jukes and Cantor 1969). Na = not applicable due to insufficient samples. SD: standard deviation of ∏. D. sub*: D. subobscura.*

Genomic region	**3'B**			**CDS_B**			**5'B**			**IGR**	**IGR**	**IGR**
Arrangement/species	**O3+4 phylad**	**O3+4 phylad[§]**	***D. sub***	**O3+4 phylad**	**O3+4 phylad[§]**	***D. sub***	**O3+4 phylad**	**O3+4 phylad[§]**	***D. sub***	**O3+4 phylad**	**O3+4 phylad[§]**	***D. sub***
Sample size	8	7	12	8	7	12	8	7	12	8	7	12
Nonsynonymous sites	0	0	0	1467.04	1467.05	1466.76	0	0	0	0	0	0
Synonymous and non-coding sites	201	201	201	458.96	458.95	459.24	225	225	225	1300	1300	1300
S(η)	13(13)	12 (12)	14 (14)	39 (39)	39 (39)	58 (58)	10 (10)	5(5)	11 (11)	140 (145)	103 (105)	159 (164)
∏	0.0232	0.0241	0.0176	0.0068	0.0069	0.0089	0.0131	0.0084	0.0179	0.0384	0.0295	0.0450
±SD	±0.0027	±0.0032	±0.0030	±0.0008	±0.0009	±0.0007	±0.0039	±0.0022	±0.0020	±0.0075	±0.0049	±0.0033
∏(a) (JC)	na	na	na	0.0009	0.0011	0.0015	na	na	na	na	na	na
∏(s) (JC)	0.02328	0.02416	0.0179	0.0260	0.0262	0.0336	0.0131	0.0084	0.0182	0.0384	0.0295	0.0468
Tajima's *D (NonSyn)*	na	na	na	-1.1753	-1.0237	-1.7134	na	na	na	na	na	na

Tajima's *D (Syn)*	-0.3385	-0.0453	-1.0035	-0.5533	-0.8535	-0.2373	-1.1455	-0.3302	0.4429	-0.5784	-0.6103	0.3660
Fu and Li's *D*	-0.6100	-0.3538	-1.1670	-0.5522	-0.8456	-0.6322	-1.2247	-0.3703	-0.0096	-0.4995	-0.7117	-0.0498
Fu and Li's *F*	-0.6068	-0.3126	-1.2787	-0.6443	-0.9484	-0.6671	-1.3366	-0.3928	0.1221	-0.5787	-0.7636	0.0684
Fu and Li's *D* + outgroup	-0.3484	-0.0373	-0.8734	-0.9436	-1.3361	-1.0297	-0.5006	-0.6963	-0.1600	-0.6484	-1.2647	-0.2076
Fu and Li's *F* + outgroup	-0.4177	-0.0474	-1.1016	-1.0592	-1.4751	-1.0717	-0.7993	-0.7255	0.0117	-0.7751	-1.3476	-0.0261
Normalized Fay and Wu's FW-H_n	-1.0700	-0.8825	-1.9023	0.1448	0.2936	0.4795	-1.3267	0.8669	0.3470	0.1657	0.5071	0.4405
Genomic region	**3'A**			**CDS_A**			**5'A**			***Hsp70IR***	***Hsp70IR***	***Hsp70IR***
Arrangement/species	**O3+4 phylad**	**O3+4 phylad[§]**	***D. sub***	**O3+4 phylad**	**O3+4 phylad[§]**	***D. sub***	**O3+4 phylad**	**O3+4 phylad[§]**	***D. sub***	**O3+4 phylad**	**O3+4 phylad[§]**	***D. sub***
Sample size	8	7	12	8	7	12	8	7	12	8	7	12
Nonsynonymous sites	0	0	0	1466.96	1466.95	1466.76	0	0	0	2934.00	2934.00	3044.47
Synonymous and non-coding sites	175	175	175	459.04	459.05	459.24	219	219	219	3044.00	3044.00	2933.53
S(η)	9 (9)	9 (9)	18 (18)	37 (37)	35 (35)	53 (53)	9 (9)	8 (8)	13 (13)	257 (262)	211 (213)	328 (333)
∏	0.0189	0.0212	0.0343	0.0060	0.0062	0.0085	0.0156	0.0126	0.0202	0.0149	0.0129	0.0185
±SD	±0.0039	±0.0039	±0.0045	±0.0009	±0.0011	±0.0008	±0.0033	±0.0035	±0.0019	±0.0019	±0.0015	±0.0011
∏(a) (JC)	na	na	na	0.0006	0.0007	0.0011	na	na	na	0.0008	0.0009	0.0013
∏(s) (JC)	0.0189	0.0212	0.0354	0.0234	0.0243	0.0330	0.0156	0.0126	0.0205	0.0247	0.0252	0.0333
Tajima's *D (NonSyn)*	na	na	na	-1.5347	-1.4341	-1.7468	na	na	na	-1.4712	-1.3188	-1.9108*
Tajima's *D (Syn)*	-0.2111	0.0590	0.0399	-0.9008	-0.7676	-0.1366	-0.0599	-0.0599	0.1175	-0.7369	-0.8250	-0.1909
Fu and Li's *D*	0.3621	0.4600	-0.3538	-1.0693	-0.9298	-0.7588	0.7006	-0.8709	-0.4120	-0.5746	-0.7150	-0.3618
Fu and Li's *F*	0.2532	0.4053	-0.2858	-1.1763	-1.0153	-0.7272	0.6106	-0.9374	-0.3122	-0.6607	-0.7798	-0.2951
Fu and Li's *D* + outgroup	0.0446	0.1349	-0.7466	-1.4921	-1.4344	-1.2871	1.0892	-0.6450	-0.1600	-0.7257	-1.1936	-0.6187
Fu and Li's *F* + outgroup	-0.0502	0.1100	-0.6823	-1.6772	-1.5671	-1.2157	1.0223	-0.7825	0.0352	-0.8664	-1.2907	-0.5099
Normalized Fay and Wu's FW-H_n	-0.1158	0.0763	0.5745	-0.3409	0.1324	0.5500	-0.4246	-0.0281	0.4004	-0.0906	0.3089	0.3695

7.3.5 Tables S4.1-S4.4: Divergence and number of nucleotide differences between the two paralogous *hsp70* CDSs per line

Supplementary table S4.1. Estimated pairwise synonymous (Ks, below diagonal) and nonsynonymous (Ka, above diagonal in bold) substitutions per site (Nei and Gojobori 1986, equations 1-3) between the coding regions of *hsp70* orthologous and paralogous genes in the O_{ST} arrangement here represented by four lines.

		Copy **A**				Copy **B**			
		OST 1	OST 2	OST 3	OST 4	OST 1	OST 2	OST 3	OST 4
Copy **A**	OST 1		**0**	**0**	**0**	**0.0007**	**0**	**0.0007**	**0**
	OST 2	0.0153		**0**	**0**	**0.0007**	**0**	**0.0007**	**0**
	OST 3	0.0153	0.0043		**0**	**0.0007**	**0**	**0.0007**	**0**
	OST 4	0.0175	0.0065	0.0109		**0.0007**	**0**	**0.0007**	**0**
Copy **B**	OST 1	0.0043	0.0109	0.0109	0.0131		**0.0007**	**0.0014**	**0.0007**
	OST 2	0.0131	0.0109	0.0109	0.0175	0.0087		**0.0007**	**0**
	OST 3	0.0109	0.0087	0.0043	0.0153	0.0065	0.0065		**0.0007**
	OST 4	0.0153	0.0087	0.0131	0.0109	0.0109	0.0109	0.0087	

Supplementary table S4.2. Pairwise number of synonymous (below diagonal) and nonsynonymous (above diagonal in bold) nucleotide differences between the coding regions of *hsp70* orthologous and paralogous genes in the O_{ST} arrangement here represented by four lines.

		Copy **A**				Copy **B**			
		OST 1	OST 2	OST 3	OST 4	OST 1	OST 2	OST 3	OST 4
Copy **A**	OST 1		**0**	**0**	**0**	**1**	**0**	**1**	**0**
	OST 2	7		**0**	**0**	**1**	**0**	**1**	**0**
	OST 3	7	2		**0**	**1**	**0**	**1**	**0**
	OST 4	8	3	5		**1**	**0**	**1**	**0**
Copy **B**	OST 1	2	5	5	6		**1**	**2**	**1**
	OST 2	6	5	5	8	4		**1**	**0**
	OST 3	5	4	2	7	3	3		**1**
	OST 4	7	4	6	5	5	5	4	

Supplementary table S4.3. Estimated pairwise synonymous (Ks, below diagonal) and nonsynonymous (Ka, above diagonal in bold) distances (Nei and Gojobori 1986, equations 1-3) between the coding regions of *hsp70* orthologous and paralogous genes in the $O_{\underline{3+4+8}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4}}$ arrangements here represented by one, two and five lines, respectively.

		Copy **A**								Copy **B**							
		O3+4 1	O3+4 2	O3+4 3	O3+4 4	O3+4 5	O3+4+8	O3+4+16 1	O3+4+16 2	O3+4 1	O3+4 2	O3+4 3	O3+4 4	O3+4 5	O3+4+8	O3+4+16 1	O3+4+16 2
C	O3+4 1		**0**	**0**	**0.0014**	**0**	**0.0014**	**0**	**0**	**0**	**0**	**0**	**0.0014**	**0.0014**	**0.0014**	**0**	**0**
o	O3+4 2	0.0131		**0**	**0.0014**	**0**	**0.0014**	**0**	**0**	**0**	**0**	**0**	**0.0014**	**0.0014**	**0.0014**	**0**	**0**
p	O3+4 3	0.0131	0.0175		**0.0014**	**0**	**0.0014**	**0**	**0**	**0**	**0**	**0**	**0.0014**	**0.0014**	**0.0014**	**0**	**0**
y	O3+4 4	0.0309	0.0175	0.0354		**0.0014**	**0.0027**	**0.0014**	**0.0014**	**0.0014**	**0.0014**	**0.0014**	**0**	**0.0027**	**0.0014**	**0.0014**	**0.0014**
	O3+4 5	0.0242	0.0109	0.0287	0.0242		**0.0014**	**0**	**0**	**0**	**0**	**0**	**0.0014**	**0.0014**	**0.0014**	**0**	**0**
	O3+4+8	0.0265	0.0265	0.0265	0.0466	0.0387		**0.0014**	**0.0014**	**0.0014**	**0.0014**	**0.0014**	**0.0027**	**0.0027**	**0.0014**	**0.0014**	**0.0014**
A	O3+4+16 1	0.0131	0.0087	0.0175	0.0264	0.0197	0.0220		**0**	**0**	**0**	**0**	**0.0014**	**0.0014**	**0.0014**	**0**	**0**
	O3+4+16 2	0.0242	0.0153	0.0287	0.0153	0.0219	0.0378	0.0242		**0**	**0**	**0**	**0.0014**	**0.0014**	**0.0014**	**0**	**0**
C	O3+4 1	0.0087	0.0175	0.0043	0.0354	0.0287	0.0310	0.0175	0.0287		**0**	**0**	**0.0014**	**0.0014**	**0.0014**	**0**	**0**
o	O3+4 2	0.0197	0.0109	0.0242	0.0242	0.0175	0.0242	0.0153	0.0175	0.0242		**0**	**0.0014**	**0.0014**	**0.0014**	**0**	**0**
p	O3+4 3	0.0220	0.0087	0.0264	0.0264	0.0197	0.0310	0.0131	0.0242	0.0264	0.0153		**0.0014**	**0.0014**	**0.0014**	**0**	**0**
y	O3+4 4	0.0309	0.0175	0.0354	0	0.0242	0.0446	0.0264	0.0153	0.0354	0.0242	0.0264		**0.0027**	**0.0014**	**0.0014**	**0.0014**
	O3+4 5	0.0242	0.0197	0.0332	0.0332	0.0131	0.0470	0.0287	0.0309	0.0287	0.0265	0.0198	0.0332		**0.0027**	**0.0014**	**0.0014**
	O3+4+8	0.0175	0.0220	0.0175	0.0400	0.0287	0.0176	0.0175	0.0332	0.022	0.0242	0.0265	0.0400	0.0378		**0.0014**	**0.0014**
B	O3+4+16 1	0.0242	0.0109	0.0153	0.0242	0.0175	0.0378	0.0153	0.0219	0.0153	0.0175	0.0197	0.0242	0.0264	0.0332		**0**
	O3+4+16 2	0.0264	0.0175	0.0264	0.0175	0.0242	0.0401	0.0264	0.0022	0.0264	0.0197	0.0264	0.0175	0.0332	0.0355	0.0197	

Supplementary table S4.4. Pairwise number of synonymous (below the diagonal) and nonsynonymous (above diagonal in **bold**) nucleotide differences between the coding regions of *hsp70* orthologous and paralogous genes in the $O_{\underline{3+4+8}}$, $O_{\underline{3+4+16}}$ and $O_{\underline{3+4}}$ arrangements here represented by one, two and five lines, respectively.

		Copy **A**								Copy **B**							
		O3+4 1	O3+4 2	O3+4 3	O3+4 4	O3+4 5	O3+4+8	O3+4+16 1	O3+4+16 2	O3+4 1	O3+4 2	O3+4 3	O3+4 4	O3+4 5	O3+4+8	O3+4+16 1	O3+4+16 2
C	O3+4 1		**0**	**0**	**2**	**0**	**2**	**0**	**0**	**0**	**0**	**0**	**2**	**2**	**2**	**0**	**0**
o	O3+4 2	6		**0**	**2**	**0**	**2**	**0**	**0**	**0**	**0**	**0**	**2**	**2**	**2**	**0**	**0**
p	O3+4 3	6	8		**2**	**0**	**2**	**0**	**0**	**0**	**0**	**0**	**2**	**2**	**2**	**0**	**0**
y	O3+4 4	14	8	16		**2**	**4**	**2**	**2**	**2**	**2**	**2**	**0**	**4**	**2**	**2**	**2**
	O3+4 5	11	5	13	11		**2**	**0**	**0**	**0**	**0**	**0**	**2**	**2**	**2**	**0**	**0**
	O3+4+8	12	12	12	20	17		**2**	**2**	**2**	**2**	**2**	**4**	**4**	**2**	**2**	**2**
A	O3+4+16 1	6	4	8	12	9	10		**0**	**0**	**0**	**0**	**2**	**2**	**2**	**0**	**0**
	O3+4+16 2	11	7	13	7	10	17	11		**0**	**0**	**0**	**2**	**2**	**2**	**0**	**0**
C	O3+4 1	4	8	2	16	13	14	8	13		**0**	**0**	**2**	**2**	**2**	**0**	**0**
o	O3+4 2	9	5	11	11	8	11	7	8	11		**0**	**2**	**2**	**2**	**0**	**0**
p	O3+4 3	10	4	12	12	9	14	6	11	12	7		**2**	**2**	**2**	**0**	**0**
y	O3+4 4	14	8	16	0	11	20	12	7	16	11	12		**4**	**2**	**2**	**2**
	O3+4 5	11	9	15	15	6	21	13	14	13	12	9	15		**4**	**2**	**2**
	O3+4+8	8	10	8	18	13	8	8	15	10	11	12	18	17		**2**	**2**
B	O3+4+16 1	11	5	7	11	8	17	7	10	7	8	9	11	12	15		**0**
	O3+4+16 2	12	8	12	8	11	18	12	1	12	9	12	8	15	16	9	

7.3.6 Table S5: Silent nucleotide divergence, distance and codon usage bias in the *hsp70* paralogs of the different lines

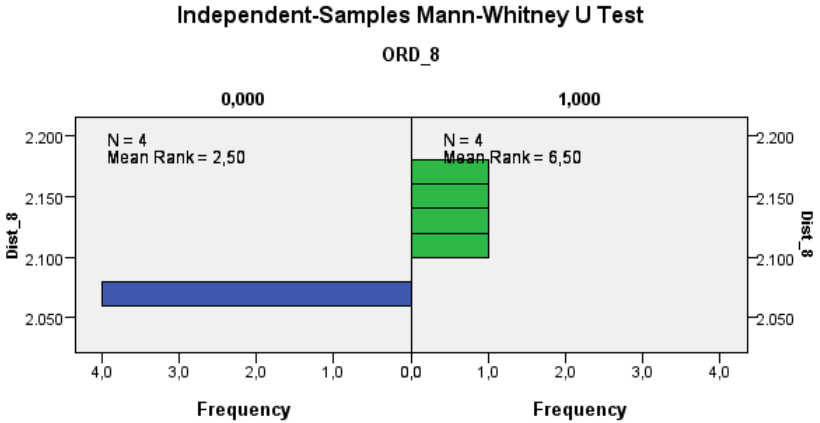
Supplementary table S5. Several polymorphisms have been quantified in our data set of 14 lines to detect differences in selection strength at the *Hsp70IR* locus that might exist between O_{ST} and the O_{3+4} phylad according to Giribets et al. (2018). These include the silent divergence between the two complete paralogous *hsp70* CDSs (K_{SIL}, including all synonymous sites plus the stop codon), the physical distance (in nts) separating the two head-to-head oriented *hsp70* CDSs (counting from the first nucleotide of the start codon) and the Effective Number of Codons index (ENC) implemented in DnaSP. A value of 20 in this index indicates the most extreme bias (only one codon per amino acid is used) while a value of 61 indicates no bias at all (all codons for a given amino acid are used equally). The ENC values shown in Table S6.1 for the different lines have been calculated combining the polymorphism levels of the two *hsp70* duplicates, although similar results were obtained with separate comparisons for *hsp70A* and *hsp70B* (data not shown). JC: values after Jukes-Cantor correction (Jukes and Cantor 1969). Box in grey: The physical distance separating the two copies in the O3+4_1 line has not been considered for statistical analyses because its intergenic region appears to be more akin to O_{ST} than to O_{3+4} (see Results section and Table 2 for more details).

Line	Arrangement/species	K_{SIL} *hsp70* A|B (JC)	Silent sites	Nuc. differences *hsp70* A|B	Physical distance *hsp70* CDS A|B (nts)	Codons analyzed	ENC index (*hsp70*A- *hsp70*B)
OST_1	O_{ST} (*D. subobscura*)	0.0043	463	2	2071	462	42.192
OST_2	O_{ST} (*D. subobscura*)	0.0109	462.67	5	2068	462	42.242
OST_3	O_{ST} (*D. subobscura*)	0.0043	462	2	2073	462	42.124
OST_4	O_{ST} (*D. subobscura*)	0.0109	463.17	5	2063	462	42.395
	Median ($\tilde{x}$) or mean ($\bar{x}$) Standard deviation (SD)	$\tilde{x}$: (O_{ST}): 0.0076 SD: 0.0038			$\bar{x}$: (O_{ST}): 2068.75; SD: 4.34.		$\bar{x}$: (O_{ST}): 42.238; SD: 0.1151.
O3+4_1	O_{3+4} (*D. subobscura*)	0.0087	462	4	2062*	462	41.923
O3+4_2	O_{3+4} (*D. subobscura*)	0.0109	462.67	5	2129	462	41.638
O3+4_3	O_{3+4} (*D. subobscura*)	0.0264	462	12	2145	462	41.834
O3+4_4	O_{3+4} (*D. subobscura*)	0	462.67	0	2103	462	41.798
O3+4_5	O_{3+4} (*D. subobscura*)	0.0131	461.83	6	2161	462	42.094
	Median ($\tilde{x}$) or mean ($\bar{x}$) Standard deviation (SD)	$\tilde{x}$: (O_{3+4}): 0.0109; SD: 0.0095			$\bar{x}$: (O_{3+4}): 2134.50; SD: 24.73.		$\bar{x}$: (O_{3+4}); 41.857; SD: 0.1677.
O3+4+8	O_{3+4+8} (*D. subobscura*)	0.0176	461	8	2207	462	42.159

O3+4+16_1	$O_{\underline{3+4+16}}$ (*D. subobscura*)	0.0153	462.17	7	2174	462	41.803
O3+4+16_2	$O_{\underline{3+4+16}}$ (*D. subobscura*)	0.0022	462.33	1	2423	462	41.837
	Median ($\tilde{x}$) or mean ($\bar{x}$) Standard deviation (SD)	$\tilde{x}$ ($O_{\underline{3+4}}$ phylad): 0.0120; SD: 0.0084.			$\bar{x}$: ($O_{\underline{3+4}}$ phylad): 2191.71; SD: 107.21.	$\bar{x}$: ($O_{\underline{3+4}}$ phylad): 41.886; SD: 0.1692.	
Dmad	*D. madeirensis*	0.0175	461.25	8	2125	462	41.838
Dgua	*D. guanche*	0.0790	460.17	34.5	2442	462	43.355

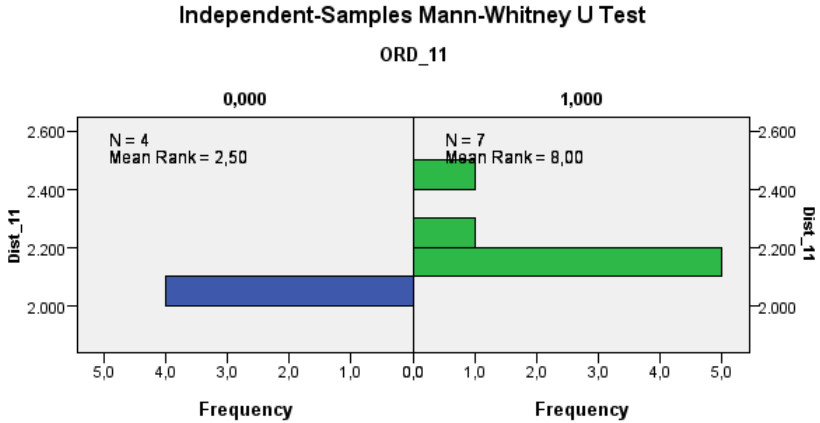
7.3.7 Figure S2: Compactness between the two *hsp70* genes in different gene arrangements

Supplementary figure S2: Higher levels of compactness are observed between the *hsp70A* and *hsp70B CDSs* in O_{ST} than in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad. The non-parametric test Mann-Whitney-U has been used to test for significant differences in the physical distance (in nucleotides) separating the two *hsp70* CDSs (starting to count from the first position of the start codon, ATG) between four O_{ST} and five $O_{\underline{3+4}}$ lines (**A**) and between four O_{ST} lines and seven lines of the three arrangements of the $O_{\underline{3+4}}$ phylad ($O_{\underline{3+4+16}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4}}$) (**B**). Note: the recombinant O3+4_1 line has not been considered for this analysis. The differences are statistically significant at $P < 0.05$, indicating that the distance separating the two CDSs is shorter in O_{ST} than in the other arrangements. The values of O_{ST} lines are shown in blue bars, while those of $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad lines are shown in green bars.



Total N	8
Mann-Whitney U	16,000
Wilcoxon W	26,000
Test Statistic	16,000
Standard Error	3,464
Standardized Test Statistic	2,309
Asymptotic Sig. (2-sided test)	,021
Exact Sig. (2-sided test)	,029

A



Total N	11
Mann-Whitney U	28,000
Wilcoxon W	56,000
Test Statistic	28,000
Standard Error	5,292
Standardized Test Statistic	2,646
Asymptotic Sig. (2-sided test)	,008
Exact Sig. (2-sided test)	,006

B

7.3.8 Tables S6.1-S6.2: Silent nucleotide divergence between the two paralogous *hsp70* CDSs in different arrangements.

Supplementary tables S6.1-S6.2: The number of silent nucleotide substitutions between the two paralogous *hsp70* CDSs is similar in the different gene arrangements. The parametric t-Test for independent samples has been used to detect significant differences among gene arrangements in the silent nucleotide divergence levels between the two *hsp70* CDSs (K_{SIL}: includes all synonymous sites plus the stop codon). The test has been performed between four O_{ST} and five $O_{\underline{3+4}}$ lines (**S6.1**) and between four O_{ST} lines and eight lines of the three arrangements of the $O_{\underline{3+4}}$ phylad ($O_{\underline{3+4+16}}$, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4}}$) (**S6.2**). The differences are not statistically significant in any of the two comparisons ($P > 0.05$), indicating that the divergence between the two *hsp70* CDSs is comparable in the different gene arrangements. Note: the recombinant O3+4_1 line has been considered for this analysis because there is no compelling evidence of GC tracts of a different arrangement in its CDSs.

Independent Samples Test

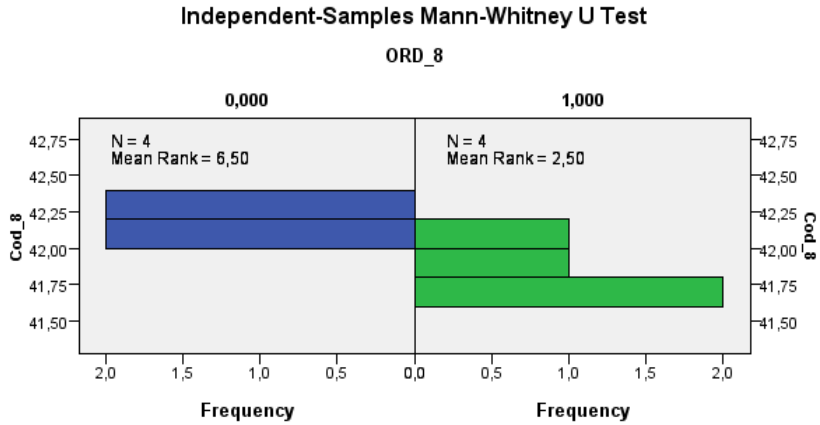
S6.1		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Ksil_9	Equal variances assumed	,883	,379	-,824	7	,437	-,0042200000	,0051218105	-,0163311573	,0078911573
	Equal variances not assumed			-,903	5,463	,405	-,0042200000	,0046749759	-,0159373753	,0074973753

Independent Samples Test

S6.2		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Ksil_12	Equal variances assumed	1,337	,274	-,923	10	,378	-,004175000	,004525407	-,014258236	,005908236
	Equal variances not assumed			-1,176	9,993	,267	-,004175000	,003550239	-,012086161	,003736161

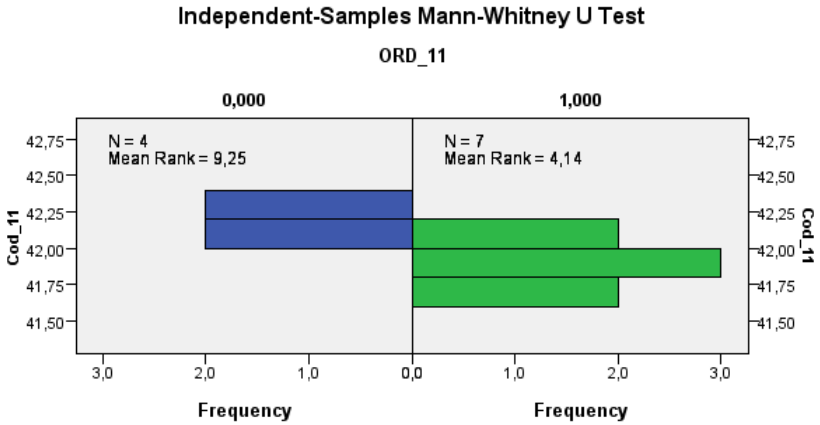
7.3.9 Figure S3: Codon usage bias in different arrangements.

Supplementary figure S3: Codon usage bias are stronger in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad than in O_{ST}. The non-parametric test Mann-Whitney-U has been used to test for significant differences between codon usage bias in the different arrangements using the Effective Number of Codons (ENC) index. The ENC index has been calculated combining the polymorphism levels of the two *hsp70* CDSs in the different lines. A value of 20 in this index indicates the most extreme bias (only one codon per amino acid is used) while a value of 61 indicates no bias at all (all codons for a given amino acid are used equally). The test has been performed between four O_{ST} lines and four $O_{\underline{3+4}}$ lines (**A**) and between four O_{ST} lines and seven $O_{\underline{3+4}}$ phylad lines (**B**). The differences in both comparisons are statistically significant at $P < 0.05$, indicating that bias are stronger in $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad than in O_{ST}. The values of O_{ST} lines are shown in blue bars, while those of $O_{\underline{3+4}}/O_{\underline{3+4}}$ phylad lines are shown in green bars.



Total N	8
Mann-Whitney U	,000
Wilcoxon W	10,000
Test Statistic	,000
Standard Error	3,464
Standardized Test Statistic	-2,309
Asymptotic Sig. (2-sided test)	,021
Exact Sig. (2-sided test)	,029

A



Total N	11
Mann-Whitney U	1,000
Wilcoxon W	29,000
Test Statistic	1,000
Standard Error	5,292
Standardized Test Statistic	-2,457
Asymptotic Sig. (2-sided test)	,014
Exact Sig. (2-sided test)	,012

B

7.3.10 Figures S4.1-S4.3: Predicted amino acid sequences of the two *hsp70* paralogs in the different lines and species

Supplementary figures S4.1-S4.3. The effects of gene conversion are also detectable in the nonsynonymous variability at the *Hsp70IR* locus. Alignment of the predicted amino acid sequences from the complete *hsp70A* (S5.1), *hsp70B* (5.2) and the two paralogous copies (S5.3) in the 14 rconstructed lines. Only the positions that contain amino acid polymorphisms are shown. Figures within brackets indicate the number of invariable positions separating two variable columns. Notice that for some polymorphisms, the less frequent amino acid occurs only in the two *hsp70* duplicates of the same line/species possibly as a result of gene conversion, including O3+4_4 ($O_{\underline{3+4}}$), O3+4+8 ($O_{\underline{3+4+8}}$) and in *D. guanche*.

S4.1. Annotated amino acid (aa) polymorphisms from the predicted *hsp70A* coding region (642 aa, 14 sequences).

```
D_guanche_A      [13]R[51]D[41]H[105]L[80]I[12]S[5]P[31]NM[134]S[12]E[29]D[38]A[12]L[39]A[3]S[6]W[2]G[2]S[8]
D_madeirensis_A  [13]C[51]N[41]Q[105]P[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]P[39]S[3]G[6]G[2]G[2]A[8]
OST_1_A          [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
OST_2_A          [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
OST_3_A          [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
OST_4_A          [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
O3+4_1_A         [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
O3+4_2_A         [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
O3+4_3_A         [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
O3+4_4_A         [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]G[38]T[12]L[39]S[3]G[6]G[2]G[2]T[8]
O3+4_5_A         [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
O3+4+8_A         [13]C[51]D[41]Q[105]L[80]M[12]F[5]S[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]S[2]A[8]
O3+4+16_1_A      [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
O3+4+16_2_A      [13]C[51]D[41]Q[105]L[80]M[12]F[5]P[31]SL[134]G[12]K[29]D[38]T[12]L[39]S[3]G[6]G[2]G[2]A[8]
```

S4.2. Annotated amino acid (aa) polymorphisms from the predicted *hsp70B* coding region (642 aa, 14 sequences).

```
D_guanche_B      [63]A[18]D[24]H[108]E[7]H[21]F[47]I[45]I[4]NM[48]S[13]K[16]S[13]C[83]D[9]A[11]A[68]P[14]G[2]T[8]
D_madeirensis_B  [63]V[18]D[24]Q[108]V[7]H[21]Y[47]M[45]T[4]SL[48]P[13]M[16]P[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
OST_1_B          [63]V[18]G[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
OST_2_B          [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
OST_3_B          [63]V[18]D[24]Q[108]V[7]N[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
OST_4_B          [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
O3+4_1_B         [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
O3+4_2_B         [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
```

```
O3+4_3_B        [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
O3+4_4_B        [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]G[9]S[11]V[68]P[14]G[2]T[8]
O3+4_5_B        [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]S[14]G[2]A[8]
O3+4+8_B        [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]S[2]T[8]
O3+4+16_1_B     [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
O3+4+16_2_B     [63]V[18]D[24]Q[108]V[7]H[21]F[47]M[45]I[4]SL[48]P[13]K[16]S[13]Y[83]D[9]S[11]V[68]P[14]G[2]A[8]
```

S4.3. Annotated amino acid (aa) substitutions from the predicted h*sp70A* and *hsp70B* coding regions (642 aa, 28 sequences).

```
D_guanche_A     [13]R[49]V[1]D[16]D[24]H[105]L[2]V[7]H[21]F[47]I[12]S[5]P[26]I[4]NM[48]P[13]K[16]S[13]Y[40]S
D_guanche_B     [13]C[49]A[1]D[16]D[24]H[105]L[2]E[7]H[21]F[47]I[12]F[5]P[26]I[4]NM[48]S[13]K[16]S[13]C[40]G
D_madeirensis_A [13]C[49]V[1]N[16]D[24]Q[105]P[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
D_madeirensis_B [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]Y[47]M[12]F[5]P[26]T[4]SL[48]P[13]M[16]P[13]Y[40]G
OST_1_A         [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
OST_1_B         [13]C[49]V[1]D[16]G[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
OST_2_A         [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
OST_2_B         [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
OST_3_A         [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
OST_3_B         [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]N[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
OST_4_A         [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
OST_4_B         [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_1_A        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_1_B        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_2_A        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_2_B        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_3_A        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_3_B        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_4_A        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_4_B        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_5_A        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4_5_B        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4+8_A        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]S[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4+8_B        [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4+16_1_A     [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4+16_1_B     [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4+16_2_A     [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
O3+4+16_2_B     [13]C[49]V[1]D[16]D[24]Q[105]L[2]V[7]H[21]F[47]M[12]F[5]P[26]I[4]SL[48]P[13]K[16]S[13]Y[40]G
```

```
D_guanche_A  (cont.)        [12]E[29]D[9]S[11]V[16]A[12]L[38]PA[3]S[6]W[2]G[2]S[8]
D_guanche_B  (cont.)        [12]K[29]D[9]A[11]A[16]T[12]L[38]PS[3]G[6]G[2]G[2]T[8]
D_madeirensis_A  (cont.)    [12]K[29]D[9]S[11]V[16]T[12]P[38]PS[3]G[6]G[2]G[2]A[8]
D_madeirensis_B  (cont.)    [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
OST_1_A  (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
OST_1_B  (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
OST_2_A  (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
OST_2_B  (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
OST_3_A  (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
OST_3_B  (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
OST_4_A  (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
OST_4_B  (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4_1_A (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4_1_B (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4_2_A (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4_2_B (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4_3_A (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4_3_B (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4_4_A (cont.)            [12]K[29]G[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]T[8]
O3+4_4_B (cont.)            [12]K[29]G[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]T[8]
O3+4_5_A (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4_5_B (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]SS[3]G[6]G[2]G[2]A[8]
O3+4+8_A (cont.)            [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]S[2]A[8]
O3+4+8_B  (cont.)           [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]S[2]T[8]
O3+4+16_1_A  (cont.)        [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4+16_1_B   (cont.)       [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4+16_2_A   (cont.)       [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
O3+4+16_2_B   (cont.)       [12]K[29]D[9]S[11]V[16]T[12]L[38]PS[3]G[6]G[2]G[2]A[8]
```