Doctoral Thesis

UAB Universitat Autònoma de Barcelona

Facultat de Ciències

Departament de Genètica y Microbiologia

GENETIC AND PHENOTYPIC DIFFERENTIATION IN THREE CHROMOSOMAL ARRANGEMENTS OF *DROSOPHILA SUBOBSCURA*

Olga Dolgova

Bellaterra, 2013

A thesis submitted by Olga Dolgova graduated in Biology in conformity with the requirements for the degree of Philosophy Doctor under supervision of Dr. Marta Pascual, Dr. Joan Balanyà and Dr. Mauro Santos.

This Thesis was accomplished thanks to the fellowship for "Formación de Personal Investigador (FPI) del Ministerio de Ciencia y Tecnología", granted for the period from October 2007 to October 2011.

Doctor and Full Professor of the Department of Genetics and Microbiology, the Faculty of Sciences of the Autonomous University of Barcelona **Mauro Santos**, Doctor and Associate Professor **Marta Pascual** and Doctor **Joan Balanyà** from the Department of Genetics, the Faculty of Biology of the University of Barcelona,

Certify that: **Olga Dolgova** has conducted under their direction the research work in the Department of Genetics and Microbiology, the Faculty of Sciences of the Autonomous University of Barcelona which has led to the development of the Doctoral Thesis entitled: Genetic and phenotypic differentiation in three chromosomal arrangements of *Drosophila subobscura.*

Dra. Marta Pascual Berniola
Professora titular
Facultat de Biologia
Universitat de Barcelona

Dr. Joan Balanyà i Maymó
Professor lector
Facultat de Biologia
Universitat de Barcelona

Dr. Mauro Santos Maroño
Profesor catedrático
Facultat de Ciències
Universitat Autònoma de Barcelona

- A Antonio, por tu apoyo y amor incondicional,

- И моей маме, сестре Наталье и тёте Гале за безграничные любовь и поддержу посвящается.

ACKNOWLEDGEMENTS

En primer lugar, me gustaría expresar mi sincera gratitud a mis directores de tesis, el Dr. Mauro Santos, la Dra. Marta Pascual y el Dr. Joan Balanyà, por proporcionar el contexto científico e intelectual que han permitido realizar esta investigación doctoral. Su paciencia, orientación y apoyo a lo largo de todo este proceso formativo son realmente apreciados. Además, me gustaría expresar mi más sincero agradecimiento al investigador principal inicial del proyecto, el Dr. Antonio Fontdevila, por aceptarme como una estudiante de Postgrado en su laboratorio.

Quisiera agradecer a todo el Grupo de Biología Evolutiva, tanto a sus miembros actuales como a los anteriores, Carla Rego, Cristina Aragón, Luz Angela Betancourt, Tomás Morán, Doris Vela, Francisco Rodríguez-Trelles, Rosa Tarrío y Pilar Garcia. Muchas gracias por su apoyo, su interés, consejos y frecuentes ayudas en esta tesis, por su comprensión y ánimo en los momentos más difíciles, y por su buen sentido de humor. Con la misma sinceridad agradezco a los miembros del Grupo de Genètica Evolutiva de la Universitat de Barcelona, con quien hemos compartido tantos momentos del trabajo y diversión: Gemma Calabria, Cinta Pegueroles, Ferran Palero, Francesc Mestres y Lluis Serra.

Especialment vui agrair a la Montse Peiró la seva assistència tècnica, ensenyament de diferents metodologies i paciència inesgotable envers mi. També voldria agrair a la Montse Sales i la Raquel per proporcionar l'aliment de les mosques i assistència tècnica. Moltes gràcies a la Julia, la Elena, la Maite i la Conchi pel seu ajut logístic i administratiu.

I also thank Dr. Ronald Bonnet for giving me the opportunity to stay in his lab and learn new techniques of molecular and analytical work. Thanks for your guidance and support, for being always ready to help with anything, for so much fun and hard-working time, for your infinite enthusiasm and very familiar and informal spirit, for the expedition to Mississippi and Appalachian mountains, which filled my short stay with lots of impressions. I´d like to thank as well the rest of the guys from the Department of Life Science of University of Tulsa: Anna, Mike and Sam. Thanks to Jessica and Michael for hosting me in their house and joyful time we spent together.

Agradeço também aos meus colegas em Lisboa: Inês Fragata, Pedro Simões, Josiane Santos e ao resto das pessoas por muito dias agradáveis juntos, tornando a estadia em Portugal muito mais fácil e alegre. Eu realmente gostei de trabalhar em equipa, onde o trabalho duro e diversão eram abundantes. Um agradecimento especial à Dra. Margarida Matos por me dar a oportunidade de desenvolver uma atividade científica neste grupo.

Es imprescindible para mí agradecer la gente que trabaja en los grupos de los doctores Antonio Barbadilla, Alfredo Ruiz y Mario Cacères, con quien he compartido mucho tiempo de la última fase de trabajo de esta tesis, por los muchos momentos bonitos que hemos pasado hablando y discutiendo sobre diversas temas científicos y no científicos en los seminarios de genómica, los "Journal Club" y "Political Club", por las comidas del viernes, calçotadas anuales, y en general el ambiente acogedor, estimulante y creativo que han desarrollado.

I would like to thank all my friends who live beside me and that ones from other countries for their encouragements and love, despite of the distance.

Finally, I dedicate this work to my husband, Antonio, for your love and valuable advices, and the rest of my family: my mother, sister and aunt who have supported me with love, understanding and encouragement during my academic studies. I dedicate it also to my daughter Alisa who fills my life with more sense and happiness.

Por último, agradezco al Ministerio de Educación y Ciencia por la concesión de una beca FPI (BES-2007-17438). Esta investigación fue apoyada por subvenciones CGL2006-13423-CGL2006 y C01-13423-C02-02 del Ministerio de Ciencia y Tecnología (España), CGL2009-12912-C03-01, CGL2010-15395 y BFU2009-07564 de el Ministerio de Ciencia e Innovación, y 2009SGR 636 de la Generalitat de Catalunya para el Grupo de Biología Evolutiva.

CONTENTS

SUMMARY/RESUMEN **14**

PART 1 | INTRODUCTION **19**

1.1. Chromosomal polymorphism **21**

1.1.1. Chromosomal polymorphisms for inversions 21

1.1.2. Rigid and flexible chromosomal polymorphisms 24

1.1.3. Latitudinal, altitudinal and seasonal types of polymorphism 25

1.1.4. Inversions as suppressors of recombination 30

1.1.5. The genetic basis of fitness variation in the chromosomal arrangements 35

1.1.6. Hypotheses on the maintenance of the inversions in natural populations 37

1.2. Adaptation to environment **41**

1.2.1. Phenotypic variation 42

1.2.2. Thermal traits 45

1.3. *Drosophila subobscura* as a model species for the study of thermal adaptation **51**

1.3.1. *Drosophila subobscura* 51

1.3.2. Adaptive value of chromosomal polymorphism for inversions in *D. subobscura* 55

1.3.3. The chromosomal arrangements as the genetic markers of global climate change 59

1.3.4. Thermal adaptation in the laboratory 61

1.3.5. Studies of inversions in *D. subobscura* at the molecular level 62

1.3.6. Chromosomal arrangements selected for the present study 65

PART 2 | OBJECTIVES **71**

PART 3 | MATERIAL AND METHODS **75**

3.1. Material collection **77**

3.2. Establishment of the homokaryotypic lines **78**

3.3. Characterization of the *O* chromosome gene arrangements **80**

3.4. Candidate gene localization **81**

3.4.1. DNA isolation 81

3.4.2. DNA amplifications 81

3.4.3. Polythene chromosome preparation and *in situ* hybridization 82

3.5. Sequencing of candidate genes **84**

3.5.1. Sample settings 84

3.5.2. DNA sequencing 87

3.5.3. Cloning 88

3.5.4. Data analysis 88

3.6. Measurement of thermal preference and thermal resistance **92**

3.6.1. Experimental settings and procedures 92

3.6.2. Thermal preference behaviour in a laboratory gradient genomics 94

3.6.3. Heat resistance 95

3.6.4. Statistical methods 96

PART 4 | RESULTS AND DISCUSSION **101**

4.1. Chromosomal polymorphism of *Drosophila subobscura* in Europe **103**

4.1.1 Distribution of gene arrangement frequencies 103

4.1.2. Temporal changes in chromosomal polymorphisms in European populations 107

4.2. Physical localization of candidate genes **114**

4.3. Genetic diversity and divergence of the gene regions **119**

4.3.1. Genetic differentiation between populations within chromosomal arrangements 120

4.3.2. Nucleotide diversity 123

4.3.3. Genetic differentiation between chromosomal arrangements 125

4.3.4. Linkage disequilibrium 137

4.3.5. Gene flux 141

4.3.6. Tests for neutrality and footprints of selection 155

4.3.7. Patterns of selection 161

4.3.8. Age of chromosomal arrangements 170

4.4. Thermal coadaptation **174**

4.4.1. Association between thermal preference and knock out temperature 174

4.4.2. Consanguinity and developmental effects 175

4.4.3. Gene arrangement effects in the outbred lines 177

PART 5 | CONCLUSIONS ________ **187**

BIBLIOGRAPHY **191**

SUPPLEMENTARY MATERIAL **217**

APPENDIX I **254**

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.

APPENDIX II **271**

CALABRIA, G., DOLGOVA, O., REGO, C., CASTAÑEDA, L.E., REZENDE, E.L., BALANYÀ, J., PASCUAL, M., SØRENSEN, J.G., LOESCHCKE, V., SANTOS, M. 2012. Hsp70 protein levels and thermotolerance in *Drosophila subobscura*: a reassessment of the thermal co-adaptation hypothesis. *J Evol Biol* **25**:691-700.

SUMMARY

Background: Global warming is affecting many wild species in different ways. One of the species demonstrating thermal adaptation on the population genetic level is *Drosophila subobscura*. Latitudinal clines in the frequency of many chromosomal inversions of this species were well documented in the original Palearctic populations, and the discovery of parallel clinal patterns a few years after the colonization of South and North Americas provided compelling evidence that the clines evolved by natural selection**.** However, the selective process maintaining inversions in populations is not yet clear. Traditionally three selective hypotheses have been advanced to explain the maintenance of the chromosomal polymorphism, according to the level of operation of natural selection: chromosome, coadapted genes ("supergenes") and individual genes.

Objectives:

To distinguish between different hypotheses the following aspects were studied in *D. subobscura:*

1. The distribution of chromosomal arrangements along the thermal gradient;
2. The nucleotide variation in six genes inside the three most frequent chromosomal inversions;
3. The genetic basis of thermal preference and heat shock tolerance in isochromosomal lines.

Results and conclusions: The frequencies of the most abundant chromosomal arrangements in general correlated with temperature gradient, forming latitudinal clines. The arrangement O_{ST} positively correlated with latitude and its frequency increased from the south to the north. At the same time the frequency of $O_{\underline{3+4}+7}$ shows a negative correlation with latitude and reaches its maximum frequency in the south of Europe disappearing in the north. The $O_{\underline{3+4}}$ arrangement has a negative correlation with the latitude. Therefore, the arrangement O_{ST} is supposed to be cold-adapted while the other arrangements are considered to be warm-adapted.

The nucleotide variation of the most frequent chromosome arrangements was analyzed in two distant Spanish populations situated along a latitudinal gradient. No within-inversion genetic differences were detected among populations, which suggest that the gene content along the gradient is rather constant for the various chromosomal arrangements and genetic flow is high.

Although gene flux between different inversions was detected, significant genetic differentiation among inversions for all genes was found. Genetic differentiation between arrangements was also detected by linkage disequilibrium analysis, showing significant associations between informative sites when comparing arrangement pairs, which could be explained by low recombination rate between inversions and probable epistasis between some genes. The footprints of selection nearly in all genes, either in coding or noncoding parts, were detected using several neutrality tests. The Local Adaptation hypothesis is the one that fits better to our data and would explain the maintenance of the coadapted gene complexes within inversions in *D. subobscura.*

Our results corroborate that arrangements on chromosome *O* affect adult thermal preference in a laboratory temperature gradient, with cold-climate O_{ST} carriers displaying a lower thermal preference than their warm-climate $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ counterparts. However, these chromosome arrangements did not have any effect on adult heat tolerance and, hence, we putatively discard a genetic covariance between both traits arising from linkage disequilibrium between genes affecting thermal preference and genes of heat shock resistance. Therefore, thermal preference and heat tolerance in the isochromosomal lines of *D. subobscura* appear to be genetically independent, which might potentially prevent a coherent response of behavior and physiology (i.e., coadaptation) to thermal selection. If this pattern were general to all chromosomes, then any correlation between thermal preference and heat resistance across latitudinal gradients would likely reflect a pattern of correlated selection rather than genetic correlation.

RESUMEN

Introducción

El calentamiento global afecta de modo distinto a diferentes especies. Una de las especies en las que se ha documentado una respuesta genética a la adaptación térmica es *Drosophila subobscura*. Las clinas latitudinales en la frecuencia de muchas inversiones cromosómicas descritas en esta especie en las poblaciones originales del Paleártico y el descubrimiento de patrones clinales paralelos pocos años después de la colonización de América del Sur y del Norte han proporcionado una de las pruebas más convincentes de que las clinas de inversiones son el producto de la selección natural. Sin embargo, se desconoce el tipo de selección responsable del mantenimiento del polimorfismo cromosómico de inversiones asociado a las clinas. Tradicionalmente se han propuesto tres hipótesis selectivas para dar cuenta del polimorfismo cromosómico, las cuales abarcan distintas unidades de selección: el cromosoma, los genes coadaptados ("supergenes") y los genes individuales.

Objetivos

Para llegar a entender los mecanismos de selección en la especie *D. subobscura,* en este trabajo se ha estudiado:

1. La distribución de las ordenaciones cromosómicas a lo largo de un gradiente térmico;
2. La variación nucleotídica en seis genes incluidos en las tres ordenaciones cromosómica más frecuentes;
3. La base genética de la preferencia térmica y de la tolerancia al calor en líneas isocromosómicas.

Resultados y Conclusiones

Las frecuencias de las ordenaciones cromosómicas en general están correlacionadas con el gradiente de temperatura, formando clinas latitudinales. La ordenación O_{ST} se correlaciona positivamente con la latitud y su frecuencia aumenta conforme se avanza desde el sur hacia el norte. Inversamente, la frecuencia de $O_{\underline{3+4}+7}$ muestra una correlación negativa con la latitud: alcanza su máxima frecuencia en el sur de Europa y

desaparece en el norte. La ordenación O_{3+4} también exhibe una correlación negativa con la latitud. Estas correlaciones indican que la ordenación O_{ST} está adaptada al frío, mientras que las otras ordenaciones pueden considerarse adaptadas a temperaturas más elevadas.

La variación nucleotídica de las ordenaciones más frecuentes se analizó en dos poblaciones españolas distantes latitudinalmente. Aunque las frecuencias de las inversiones difieren entre ambas poblaciones, no se han detectado sin embargo diferencias nucleotídicas dentro de cada inversión entre las poblaciones. Se ha detectado flujo genético entre las diferentes inversiones, pero éste no es suficiente para evitar la existencia de diferenciación genética significativa entre las inversiones para todos los genes analizados. La diferenciación genética entre las ordenaciones también se detectó mediante el análisis de desequilibrio de ligamiento. Aparte de la baja tasa de recombinación entre las inversiones, la epistasis en eficacia entre algunos genes podría también contribuir a la diferenciación observada. Mediante la aplicación de diversas pruebas de neutralidad, se ha podido detectar la huella de la selección prácticamente en todos los genes analizados, ya sea en regiones codificadoras o no codificadoras. La hipótesis de la adaptación local es la que se ajusta mejor a nuestros datos, o sea, las inversiones mantienen complejos de genes coadaptados en inversiones de *D. subobscura*.

Nuestros resultados corroboran que las ordenaciones del cromosoma *O* afectan la preferencia térmica en adultos en un gradiente termal producido en el laboratorio, y que moscas que llevan la ordenación O_{ST} adaptada al frío muestran una preferencia térmica hacia temperaturas más bajas que aquellas que tienen las ordenaciones O_{3+4} y O_{3+4+8} adaptadas al calor. Sin embargo, estas ordenaciones cromosómicas no tienen ningún efecto sobre la tolerancia al calor en adultos y, por lo tanto, podemos suponer que no hay covarianza genética entre ambos rasgos. La preferencia térmica y la tolerancia al calor en las líneas isocromosómicas de *D. subobscura* parecen pues ser genéticamente independientes, lo que podría impedir una respuesta coherente del comportamiento y la fisiología (es decir, la coadaptación) a la selección térmica.

Part 1

Introduction

The White Rabbit put on his spectacles.

"Where shall I begin, please your Majesty?" he asked.

"Begin at the beginning", the King said gravely, "and go on till you come to the end: then stop".

Lewis Carroll,

Alice's Adventures in Wonderland

1

INTRODUCTION

1.1. CHROMOSOMAL POLYMORPHISM

1.1.1 CHROMOSOMAL POLYMORPHISMS FOR INVERSIONS

The eminent Russian geneticist Theodosius Dobzhansky, in *Genetics and the Origin of Species* (1937, p. 13), wrote: *Mutations and chromosomal changes arise in every sufficiently studied organism with a certain finite frequency... Once produced, mutations* (chromosomal changes) *are injected in the genetic composition of the population, where their further fate is determined by the dynamic regularities of the physiology of populations... The influences of selection, migration and geographical isolation then mold the genetic structure of populations into new shapes, in conformity with the secular environment and the ecology, especially the breeding habitats, of the species... on which the impact of the environment produces historical changes in the living population*. Such basic statements have the great merit of distinguishing between empirical patterns observed in nature, where one must necessarily proceed by inference, and the underlying mechanisms responsible for such patterns. Whenever feasible, a route to unravel patterns from mechanisms is to employ the classic experimental approach: studying the phenomena by isolating factors one by one.

The theory of Natural Selection, proposed by Charles Darwin in the XIX century, is fundamental and one of the most studied topics in biology. However, due to their complexity, the exact mechanisms involved in evolution and speciation are still unclear. Chromosomal inversions are clearly subjected to natural selection, since their evolutionary significance has been studied using many groups of animals and plants (reviewed by HOFFMAN and RIESENBERG 2008). One of the factors conditioning natural selection is the presence of variability for a trait. We can speak of *polymorphism* when two or more alleles or two or more inherited phenotypes coexist in considerable frequencies in the same population. Although a polymorphism can be controlled by alleles of a single locus (as in the case of the human ABO blood groups), some more complex forms of polymorphism are controlled by supergenes, formed by the alleles of many genetic loci

closely linked on the same chromosome. A particular case is chromosomal polymorphism, where individuals can be different with respect to number or structure of chromosomes. Of particular relevance for *Drosophila* species is the chromosomal polymorphisms for inversions. Succinctly, an inversion is a change in chromosomal structure in which a fragment of the chromosome involving several genes rotates 180°, resulting in a region where the positions of these genes are inverted relative to the original sequence. When the inversion includes the centromere it is called *pericentric*; otherwise it is called *paracentric* (Figure 1.1). In natural populations of *Drosophila* only *paracentric* inversions have evolutionary significance and it is widely accepted that the origin of these inversions is unique (POWELL 1997).

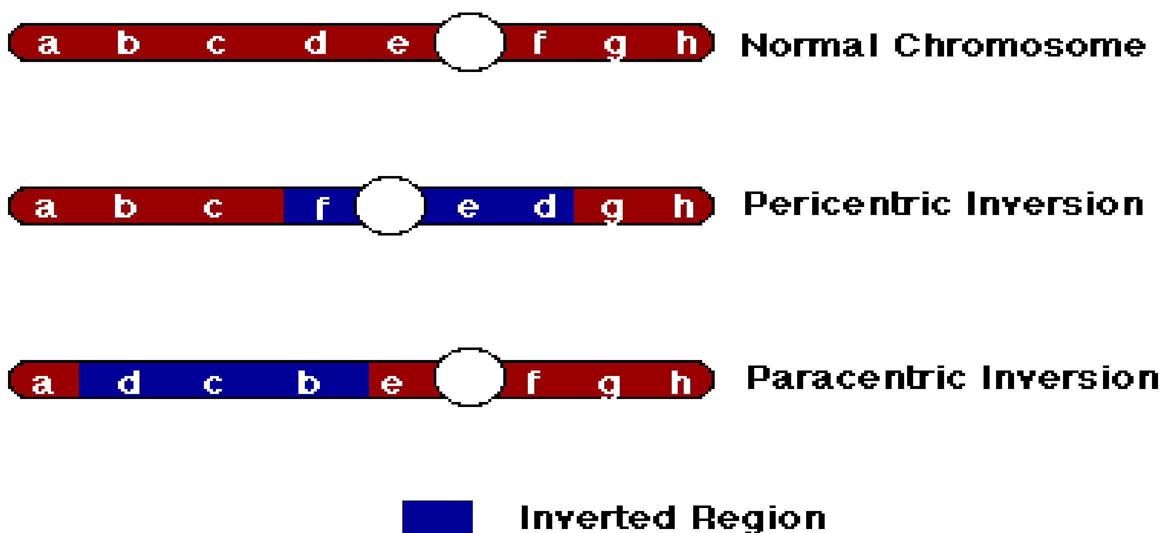


Figure 1.1: Simple pericentric and paracentric inversions. The ancestral gene order is symbolized by alphabetic order of letters. The corresponding inverted structures originate through breaks between c-d and f-g (respectively, between a-b and d-e), a turn of 180 ° and the subsequent union of the fragments with the rest of the chromosome.

Cytological methods for detecting inversions have been predominantly employed in *Diptera* because these insects present specialized cells that undergo repeated rounds of DNA replication without cell division, forming giant chromosomes (polytene chromosomes) that could thus be studied at a greatly enhanced resolution. These chromosomes have characteristic banding patterns, which can be used to identify structural changes, and often form an inversion loop in individuals containing 2 different gene arrangements on the same chromosome (Figure 1.2).

The polythene chromosome preparations allow detection of the frequency of inversions in natural populations, localization of the break points of the inversions, and identification of genetic markers by *in situ* hybridization. This technique is still the most accurate way of localizing genetic markers in

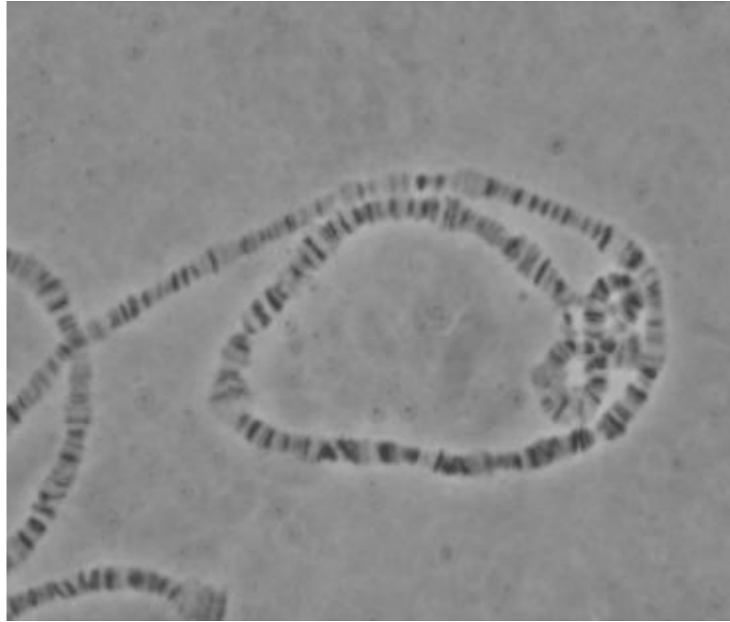


Figure 1.2: Polythene chromosomes of heterokaryotypic individual of *D. subobscura* with gene arrangements O_{ST}/O_{3+4}, where the inversion loop is easily observed.

species whose genome has not been still sequenced, since the comparison to closely related species is usually not appropriate because normally there is no synteny (i.e. the physical co-localization of genetic loci on the same chromosome) and the order of genes varies (SANTOS *et al.* 2010). The molecular techniques have allowed the identification of new inversions, for example by detecting unusual patterns of linkage or comparison of the order of markers between species. With classic as well as modern techniques, inversions have been identified in a wide range of species of microorganisms, plants, insects and animals (HOFFMANN, SGRO and WEEKS 2004). Finally, it is expected that new techniques of high throughput sequencing (next generation sequencing) will allow obtaining new genomes and more rapid and accurate identification of inversions originating in different lineages (reviewed in HOFFMANN and RIESENBERG (2008)).

There is still no agreement on what is the mechanism that generates inversions. In some species it has been observed that the formation of inversion is related with the ectopic recombination caused by *transposable elements* (TE) (CÁCERES *et al*. 1999). Thus, the formation of an inversion requires two sequences that are repeated on the same chromosome in inverted orientation. However, because of the lack of TE “marks” near the breakpoints of some inversions new mechanisms have recently been proposed to explain their origin, such as the *staggered breaks model* (CASALS and NAVARRO 2007). This model suggests that the first step is the formation of staggered cuts in DNA

chains, particularly in two different points of the same chromosome. The inversion occurs because the repair mechanisms tie the 5 'ends with ends 3' of the opposite breakpoint. At the end, the resulting spaces are refilled with the consequent formation of inverted duplications at each end of the inversion. Depending on the length of staggered cuts, the duplicated regions could be detectable or not. Finally, inversions can be artificially induced by X-rays (SPERLICH *et al.* 1977). Inversions are quite frequent mutations, for example, approximately three quarters of *Drosophila* species are polymorphic for paracentric inversions (KRIMBAS 1992). Recent studies comparing the genomes of these species showed the presence of several reversed regions that are species-specific, suggesting that rearrangements play an important role in the evolution of species genomes (AYALA and COLUZZI 2005). At the intra-specific level it is common to find polymorphism for inversions, where the inverted chromosomes segregate with their non-inverted homologs (HOFFMANN and RIESENBERG 2008). In *Drosophila*, inversions tend to be large and cover a significant portion of the chromosome (KRIMBAS and POWELL 1992). In addition, the number of inversions that have different species varies greatly within the genus *Drosophila* (HOFFMANN and RIESENBERG 2008).

1.1.2. RIGID AND FLEXIBLE CHROMOSOMAL POLYMORPHISMS

In a comparative study of chromosomal polymorphisms in different species of *Drosophila* Dobzhansky classified these polymorphisms as flexible or rigid (DOBZHANSKY 1962). In species that have flexible chromosomal polymorphisms, the frequency of different arrangements in natural populations varies according with fluctuations in environmental factors. For instance, the majority of *D. pseudoobscura* populations demonstrate the same pattern: the frequencies of chromosomal arrangements vary depending on the temperature in natural populations as well as in the laboratory. The change in temperature also causes cyclic seasonal oscillations of different arrangements in this species. There are also notable differences between nearby locations but they have very different biotic conditions because they are located at distinct altitudes. In contrast, in the species that have rigid polymorphism the seasonal deviations of the frequencies of the different gene arrangements are not observed, or they are insignificant. The nearby populations but with different biotic conditions do not differ as well, and in extreme cases, such as *D. pavani* (BRNCIC 1957), differences between geographically separated populations are not observed either.

As this work is related to *Drosophila subobscura* the important question is: is its chromosomal polymorphism for inversions rigid or flexible? DOBZHANSKY (1962) considered that it was rigid, based on observations of KUNZE-MÜHL, MÜLLER and SPERLICH (1958), which coincided with those of STUMM-ZOLLINGER (1953) and PENTZOS-DAPONTE (1964). In all these studies the seasonal variation of chromosomal polymorphism of *D. subobscura* was analyzed in a particular locality and the significant changes in the frequencies of the different gene arrangements were not detected in any case. PREVOSTI (1964) analyzed the chromosomal polymorphism in populations of this species in the area of Barcelona, but in locations whose microevolutionary conditions were very distinct, and neither observed differences. But there were other studies, which observed seasonal or altitudinal variation of the chromosomal polymorphism in *D. subobscura*. BURLA and GÖTZ (1965) detected a significant seasonal variation of the frequencies of certain arrangements in nearby localities of Zurich. KRIMBAS (1964) also observed seasonal variations of chromosomal polymorphism in Greek populations of the species and attributed the distributions of certain arrangements to the effect of temperature. So, there is no doubt that the chromosomal polymorphism of *D. subobscura* varies over widespread geographic areas. The other proof of this fact is the latitudinal clines presented by some arrangements of the species, which will be considered further in this thesis. The populations of northern Europe, for example, have very poor polymorphism and could be considered in this sense as marginal populations. The Mediterranean populations have a very rich polymorphism and the Central European populations could be considered intermediate. RODRÍGUEZ-TRELLES, ALVAREZ and ZAPATA (1996) detected variation of polymorphism for inversions in chromosome *O* of *D. subobscura* at two levels: a variation in the short term, seasonal (spring, early summer, late summer and autumn) with seasonal changes repeated over the years of the most frequent gene arrangements and correlated with the seasonal variation of climatic factors as temperature, humidity, rainfall and sunshine; and a variation of some arrangements in a long period, i.e., directional.

1.1.3. LATITUDINAL, ALTITUDINAL AND SEASONAL VARIATION OF INVERSION POLYMORPHISM

Geographical changes

A naturally occurring pattern in species from the genus *Drosophila* has received much attention by evolutionary biologists: latitudinal/altitudinal gradients in chromosomal inversion polymorphisms

(e.g., DUBININ and TINIAKOV 1946; STALKER 1976; METTLER, VOELKER and MUKAI 1977; LEVITAN 1978; KNIBB, OAKESHOTT and GIBSON 1981; MENOZZI and KRIMBAS 1992; VAN'T LAND *et al.* 1999; SCHAEFFER *et al.* 2003; see KRIMBAS and POWEL 1992 for review). The exhaustive study of chromosomal inversions allowed detection in many species some interesting and revealing patterns of geographical distribution (POWELL 1997). The interpretation of these patterns was not without controversy. There were interpretations of selection based on environmental heterogeneity, drift and historical considerations.

i. Latitudinal clines

In some *Drosophila* species presenting polymorphism for inversions, this polymorphism is distributed geographically so that the frequency of some inversions varies in systematic way in relation to latitude forming *latitudinal clines*. *Cline* means a systematic change in the frequency of arrangement (or inversion) along a geographical gradient. The latitudinal clines can be measured by the slope of the regression line of (corrected) frequency of the arrangement (or inversion) in each of the analyzed locations. Thus, in *D. melanogaster* its 4 cosmopolitan inversions have well pronounced latitudinal patterns. When studying different populations of America, Europe and Australasia it was shown that these inversions have a negative correlation with latitude in all studied areas (LEMEUNIER and AULARD 1992). Latitudinal clines were also found in *D. robusta* (LEVITAN 1992) and *D. persimilis* (POWELL 1992).

The most convincing evidence that the latitudinal clines are caused by natural selection probably comes from the temperate (i.e., cold tolerant) species *Drosophila subobscura* Collins 1936. In this species latitudinal clines were also described for different arrangements in different chromosomes in Europe, the ancestral region of the species, as well as in South America and North America, both regions colonized by the species in the late 70's and early 80's (PREVOSTI *et al.* 1988; AYALA, SERRA and PREVOSTI 1989; MENOZZI and KRIMBAS 1992). The fact that concurrent clines occur in both hemispheres strongly suggests that they are established by natural selection.

ii. Altitudinal clines

Altitude is another factor that has been studied with regard to geographical changes in inversion polymorphism. In this case, the species that showed the clearest altitudinal clines was *D. robusta*.

In addition, the establishment of polymorphism in altitudinal clines is similar throughout all its geographic area, which is very good evidence that selection is responsible for the establishment of a cline (ETGES 1984; LEVITAN 1992).

Another example of natural selection modulating the altitudinal clines establishment was observed in the Hawaiian species *D. silvestris*. In this case it was found that populations, which had presented altitudinal clines and disappeared after a volcanic eruption, reestablished the altitudinal cline when recolonized the area (CARSON, LOCKWOOD and CRADDOCK 1990).

There are few published studies with regard to changes in chromosomal polymorphism correlated with altitudinal gradient in *D. subobscura*. The first study, conducted by MARTINEZ-SEBASTIAN, LATORRE and DE FRUTOS in 1984, compared the chromosomal polymorphism in three populations located at different altitudes in the Sierra de Gúdar with their differences for ecological and climatic conditions, but they did not detect any difference. However, BURLA, JUNGEN and BÄCHLI in 1986 detected differences in polymorphism when comparing the samples captured at different altitudes in a region of Switzerland. The differences were mainly in the *J* chromosome, which presented an altitudinal cline comparable with that presented latitudinally.

Other species demonstrating the altitudinal clines are *D. pseudoobscura* and *D. persimilis* (DOBZHANSKY and EPLING 1948), *D. flavopilosa* (BRNCIC 1972), *D. nasuta* (RANGANATHA and KRISHNAMURTHY 1978), *D. annanassae* (REDDY and KRISHNAMURTHY 1974) and *D. mediopunctata* (ANANINA *et al.* 2004).

Temporal changes

i. Seasonal changes (short-term)

Short-term changes in the frequency of various inversions were the first evidence that lead Dobzhansky to suspect that the chromosomal inversion polymorphism is subjected to the action of natural selection. His early works demonstrated seasonal changes in the chromosomal inversion polymorphism of *D. pseudoobscura* after the analysis of monthly samples in three different populations (DOBZHANSKY 1970).

The most comprehensive study to date in order to determine whether *D. subobscura* shows some sort of seasonal variation was carried out in the population of Mount Pedroso, in northwestern

Spain. In two studied periods it was detected that some inversions showed seasonality (FONTDEVILA *et al.* 1983; RODRÍGUEZ-TRELLES, ALVAREZ and ZAPATA 1996). In general, there was an increase of the $O_{\underline{3+4}+7}$ arrangement in summer and a decrease in spring and autumn, while O_{ST} behaved in an opposite way. These seasonal changes present a behavior, which is consistent with clinal geographical distribution of these arrangements. It means that the arrangement O_{ST}, which is very common in populations of northern Europe, produced a minimum frequency in summer in the studied area, while the arrangement $O_{\underline{3+4}+7}$, typical in the South showed a maximum in this season. Direct effects of rising temperature in a given period have shown transiently shift to summer-like frequencies in the genetic constitution of populations with a posterior rapid recovery of inversion frequencies to their normal seasonal range (RODRÍGUEZ-TRELLES, TARRÍO and SANTOS 2013).

Moreover, these arrangements showed a strong relationship with climatic factors such as maximum daily temperature and relative humidity. There is also a study that compares the seasonal variations in a population near Barcelona. In this work, DE FRUTOS and PREVOSTI (1984) did not detect variations in the chromosomes *A*, *E* and *O*, but they did for frequencies of the inversions in the chromosomes *J* and *U* though the data were insufficient to see if they followed a cyclical pattern, although seasonal fluctuations of inversion frequencies in all chromosomes has been recently reported (RODRÍGUEZ-TRELLES, TARRÍO and SANTOS 2013).

Seasonal variations were also found in *D. persimilis* (DOBZHANSKY and AYALA 1973), *D. robusta* (LEVITAN 1992) and *D. funebris* (SPERLICH and PFRIEM 1986).

ii. Long-term changes

For many of the species mentioned here, there are registries of inversion frequencies in populations covering more than 50 years. This allowed an analysis to see whether this polymorphism has been changing over time.

One of these species is *D. pseudoobscura* whose polymorphism although in general remained stable over time, had varied in certain populations. For example the *PP* and *TL* arrangements increased in their frequencies in a given area (from British Columbia to California). One possible explanation for this phenomenon was that this variation was associated with the environmental changes, but the conducted experiments did not give any conclusive evidence (ANDERSON *et al.* 1991). An alternative explanation was that the inversions themselves had evolved. This means that

the increase of the frequency of these inversions was due to a change in their genetic content (by mutation, gene conversion or crossover), which increased their selective value.

Other species with variations of chromosomal polymorphism in the long term are *D. robusta* (LEVITAN 1992), *D. melanogaste*r (UMINA *et al.* 2005) and *D. subobscura* (SOLÉ *et al.* 2002; BALANYÀ *et al*. 2004). One of the first long-term studies in *D. subobscura* was conducted by GOSTELI (1990a,b), who compared the levels of polymorphism from the data obtained in the mid 80's with another of the early 60's in a Swiss population. The conclusion from this study was that standard inversions of the chromosomes *A*, *J*, *U* and *O* had decreased in the frequency while the arrangements J_1, U_{1+2}, O_{3+4} and O_{3+4+8} had increased. The studies in Mount Pedroso (RODRÍGUEZ-TRELLES, ALVAREZ and ZAPATA 1996; RODRÍGUEZ-TRELLES and RODRÍGUEZ 1998) can also be considered as long term as they compared the data obtained in the years 1976-1980 with that of 1988-1991, where it was observed a decrease in the frequency of arrangements O_{ST} and O_{3+4+8}, and significant increase in O_{3+4}. Similar results were obtained by ORENGO and PREVOSTI (1996) comparing data from Barcelona population in an interval of 29 years. In general, the standard arrangements had declined in their frequency while the chromosomal inversions typical for southern populations had increased.

More recently, BALANYÀ *et al.* (2006) in a study where different environmental factors such as temperature were related with the data of inversion polymorphism collected during 40 years in 26 populations from Europe, South America and North America, demonstrated that the chromosomal polymorphism has varied over time and that inversions typically considered as warm-climate (which represent the cline S-N; MENOZZI and KRIMBAS 1992) had been increasing in frequency as the global temperature increased on the planet. Thus, the temperature could be considered as a selective force in adaptation by chromosomal polymorphism. Unfortunately, when this hypothesis was tested in the lab, the results were not conclusive (SANTOS *et al*. 2004; 2005). It was an attempt to study the role of temperature in thermal adaptation examining how the frequencies of chromosomal inversions and other genetic markers evolved in Chilean populations of *D. subobscura* maintained at three different temperatures (13, 18 and 22 ° C), which comprise a large part of the thermal range of the species. The results showed that, despite some inversion frequencies varied according to expected patterns observed in nature, they didn’t seem to be caused only by a direct effect of temperature. However, it should be noted that in this experiment

the flies were kept at constant temperatures, which does not accurately reflect what happens in nature. Thus, inversions may not be favored by the total temperature, but by thermal fluctuations experienced by individuals in different populations.

The reasons for the flexibility and different types of variation of inversion polymorphisms lay in their adaptive character to the several environmental conditions listed above. The main mechanism, which allows selection to act over the inversions is the ability of the latter to suppress the recombination (PEGUEROLES *et al.* 2010b).

1.1.4. INVERSIONS AS SUPPRESSORS OF RECOMBINATION

The study of chromosomal inversions began in the early XX century, when Sturtevant and colleagues first observed that the gene maps of *D. melanogaste*r and *D. simulans* were reversed (STURTEVANT 1921; STURTEVANT 1926). Besides it was found that chromosome crossovers were reduced in females *heterozygous* for inversions (when the individual has one non-inverted homolog chromosome and the other with inversion) but not in *homozygous* ones, either for inverted on non-inverted chromosomes (STURTEVANT and PLUNKETT 1926). The researchers assumed that if the inversion were relatively small, it would prevent chromosome crossovers because of the difficulty in pairing of the homologous chromosomes. However, if the inversions were long enough to allow chromosome crossovers the result would be the formation of acentric and dicentric chromosomes and the consequent formation of unbalanced gametes and unviable zygotes. However, Beadle and Sturtevant did not detect the right proportion of viable eggs that they expected to find and, therefore, suggested that the meiotic products in the eggs of *Drosophila* are linearly arranged, and only one of the two more distal nuclei constitute the functional nucleus of the egg (STURTEVANT and BEADLE 1936). Therefore, both internal (which would contain the unbalanced gametes) and one of the external nuclei form the polar bodies (Figure 1.3). CARSON (1946) showed cytologically this mechanism working in the fly *Sciara*. In other species such as wheat, unbalanced gametal products are not removed, causing great problems of infertility (STRICKBERGER 1976). Furthermore, the inversion could be large enough to produce an even number of chromosome crossovers and in this case viable gametes will be formed. Therefore, as originally discovered by Sturtevant, inversions suppress recombination in the inverted fragment in heterozygous (heterokaryotype) combination. The heterozygotes for inversions can be observed using the giant salivary gland chromosomes of

Diptera as indicated above, because these chromosomes are formed by superposition of many chromatids of each homologue, and, moreover, the two homologs are closely paired (somatic pairing).

In the late '70s, Paul A. Roberts, in a review on the genetics of chromosomal aberrations, explained that the chromosomal inversions reduce recombination for two reasons. On one hand, the loop of the inversion in heterokaryotypes is a physical handicap that partially inhibits the formation of chiasms, especially near the break points and, moreover, the odd number of chromosome crossovers leads to the formation of unviable gametes (ROBERTS 1976). If one of the two homologs has an inversion and the other doesn't, to achieve a correct somatic pairing both homologs must form a loop (inversion loop) in the area of the reversed fragment.

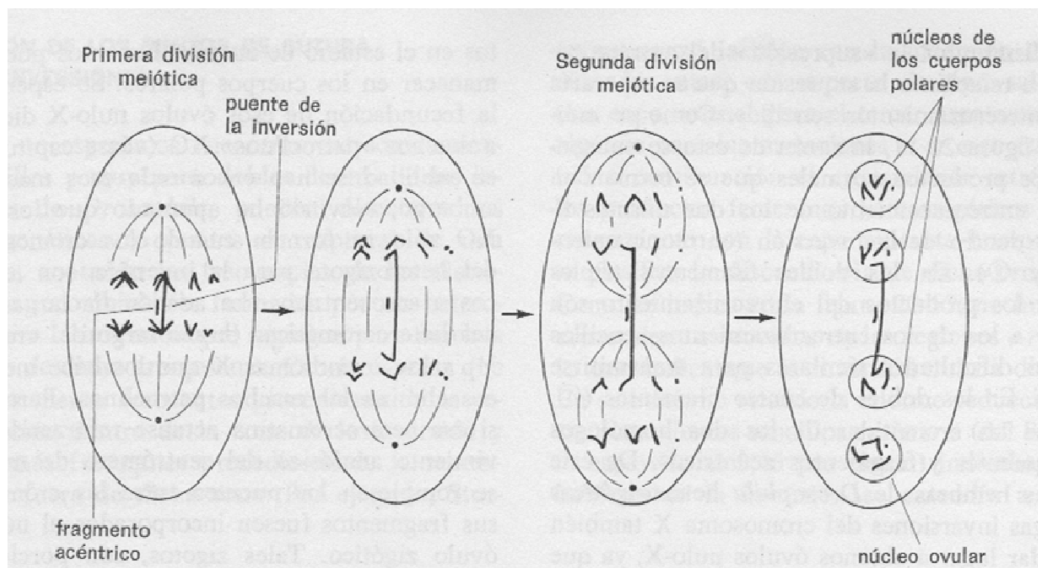


Figure 1.3: Scheme of the linear distribution of the meiotic products in the Drosophila ovaries by STURTEVANT and BEADLE (1936). Figure 22.11 of STRICKBERGER (1976).

Figure 1.4 presents a diagram of the formation of this loop in a heterozygote for inversion *AR* and *ST* of *D. pseudoobscura*. In *Drosophila* females heterozygous for paracentric inversions if crossovers occur in the inverted fragment during meiosis the recombinant chromatids are acentric or dicentric and they transform into the polar bodies but not into the oocyte (Figure 1.5). In *Drosophila*, chromosome crossovers do not occur in males. Therefore, there are no recombinant chromatids in the heterokaryotypic gametes and the fertility of these individuals is normal. This mechanism cannot exist in the case of pericentric inversions, which may explain the fact that these inversions are very rare in natural populations. If there is a double crossover inside an inversion,

however, the recombinant chromatids will be normal; but the probability of this event is very small and proportional to the length of the inversion. In general, crossovers are strongly reduced near the break point and their possibility increases gradually moving away from it. However, within the inversions restrictions are stronger as only an even number of crossovers produce balanced gametes, while outside the inversions any crossover would produce recombinant progeny.

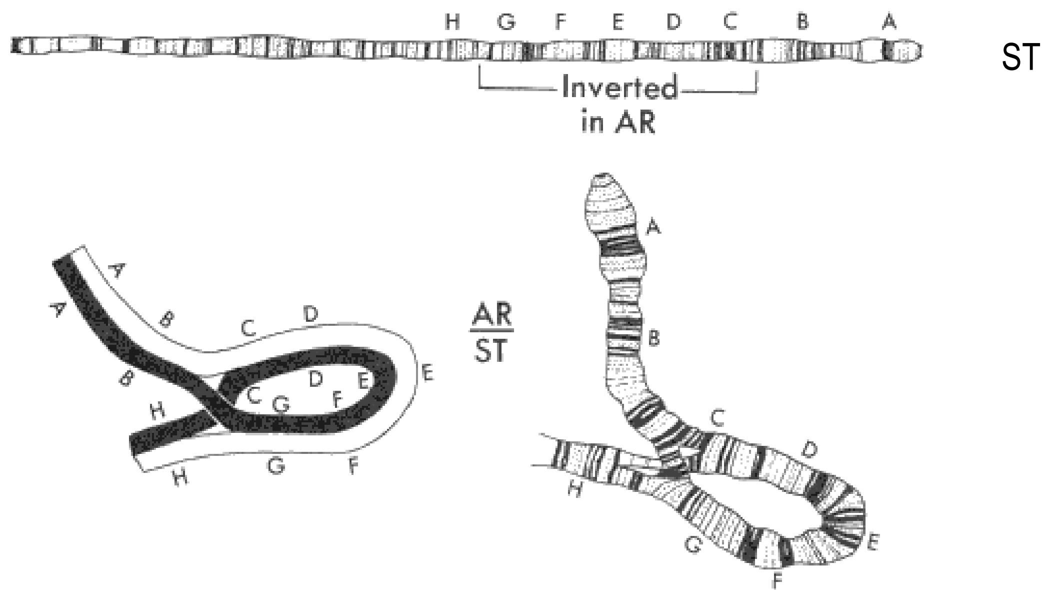


Figure 1.4: Diagram of the inverted region between the *AR* and *ST* arrangement of *D. pseudoobscura* and the inversion loop formed in a heterokaryotype (*AR/ST*). From ROBERTS (1976).

Moreover, reduction of recombination outside the inversions is not symmetrical to the two break points, but varies depending on the species. In some species it has been observed that the reduction is greater at the distal end of inversion (ROBERTS 1962, PEGUEROLES *et al.* 2010b) whereas in other species, reduction in the proximal end is more frequent (CARSON 1953; STUMP *et al.* 2007).

On the other hand, it was described that the presence of inversions increases the rate of recombination in other regions of the genome (WHITE and MORLEY 1955; KRIMBAS and POWELL 1992). For example, the rate of recombination in *D. virilis* increases in more remote areas of inversion (KOMAI and TATAKU 1940). Moreover, the Schultz-Redfield effect implies that the presence of inversions in a particular chromosome increases the rate of recombination in other chromosomes (SCHULTZ and REDFIELD 1951). This effect has been observed in *D. melanogaster*

(STEINBERG 1936), in *D. virilis* (KOMAI and TAKAKU 1942), *D. robusta* (CARSON 1953) and *D. subobscura* (PEGUEROLES *et al.* 2010b).

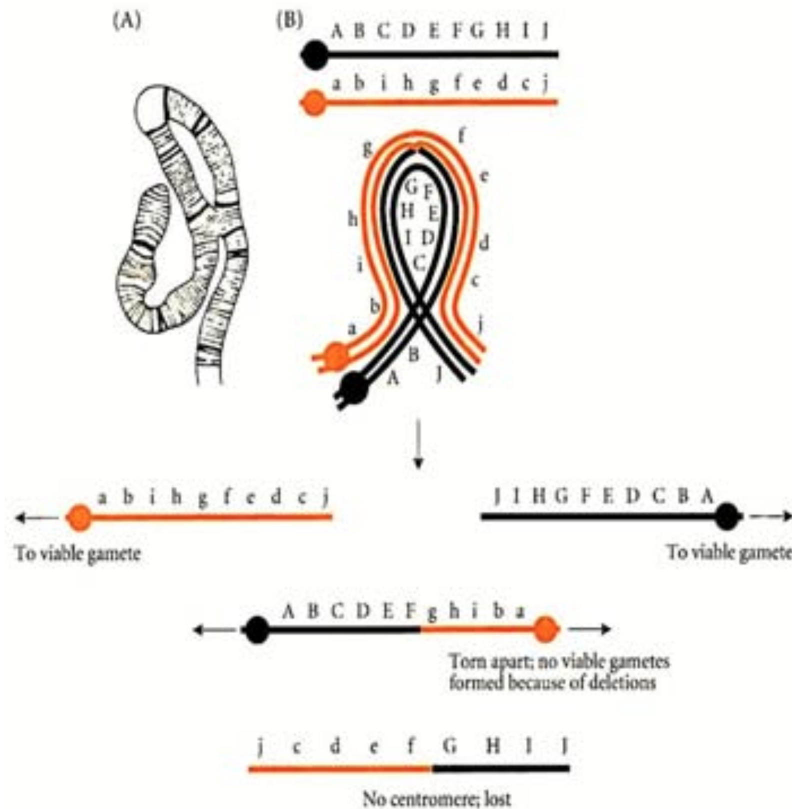


Figure 1.5: **Scheme of the mechanism of inhibition of recombination by inversion** (from GRIFFITHS *et al.* 2000 with modifications).

Arcadi Navarro and coworkers carried out a theoretical study in 1997 to see the effect of inversions on the gene flux, either caused by an even number of crossovers or by gene conversion (NAVARRO *et al.* 1997). They used two models, but one that took into account the phenomenon of interference fitted best to the empirical data. This study arrived at three main conclusions. First, recombination is completely inhibited near the break points, but gene conversion is not. Second, in the central part of the inversion it is expected that the rate of gene flux would be higher, mainly due to recombination. Third, in short inversions gene flux is uniformly distributed as gene conversion is the main force, while in large inversions more gene flux is expected in central areas of the inversion than near the break points due to double crossovers. This conclusion is related to the observation that interference plays a more important role in small inversions (CHOVNICK 1973). However, in areas not affected by inversions, recombination by double crossovers will be stronger in

determining patterns of nucleotide variability than gene conversion (NAVARRO *et al.* 1997). Other authors such as Ary Hoffmann suggest that gene conversion may be reduced around the break points, where meiotic pairing is altered (HOFFMANN, SGRO and WEEKS 2004; HOFFMANN and RIESENBERG 2008). Given that both gene conversion and double crossovers need chiasm formation, it is difficult to believe that near the break points only recombination by crossover is strongly inhibited.

The pioneering study on the characterization of the *Drosophila* gene conversion was done in 1973 by Arthur Chovnick. He proposed that estimates of gene conversion are essential to properly characterize the role of inversions as barriers of gene exchange in natural populations, and assumed that in heterozygous individuals the genetic conversion would be important in the regions where homologous chromosomes are correctly paired allowing the exchange (CHOVNICK 1973). Studying the gene *rossy* in *D. melanogaster*, which is located in the center of inversion *In(3LR)*, he came to the conclusion that the gene conversion in homozygotes had the same magnitude as in heterozygotes. He pointed out that the location of that gene in a central part of the inversion might be the key for obtaining this result. He also argued that interference affects double crossovers but not gene conversion. Finally, he also concluded that the formation of double crossovers and gene conversion that occur with higher rates than mutation reduces the genetic differentiation between chromosomal arrangements.

The reduction of recombination within the inversion can facilitate the process of speciation. For instant, the inversions can facilitate the accumulation of genes that contribute to reproductive isolation between populations connected by gene flow (NOOR *et al.* 2001). There are several models that explain how the chromosomal rearrangements accelerate speciation. The model of the hybrids dysfunction proposed by AYALA and COLUZZI (2005) states that inhibited recombination in individuals heterozygous for inversion produces a decrease in fitness. Under these conditions, natural selection favors the appearance of mutants that reduce the probability of exchange and so, eventually, will complete reproductive isolation between two species. There are examples that prove this model in the Australian grasshoppers studied by WHITE (1968).

The other model is the so-called suppression of recombination model proposed by COLUZZI (1982). Under this model of speciation it is expected that the molecular differences between species are concentrated only in the inverted segments, which differentiate them. The evidence for this model

was found in *D. pseudoobscura* and *D. persimilis* (NOOR *et al.* 2001) which are two sympatric species that differ only in two major fixed paracentric inversions in the *X* and 2 chromosomes. Moreover, the genes associated with the mechanisms of isolation between the two species are placed only inside the inversions.

Therefore, the inhibition of recombination makes inheritance of the inversions as essentially simple Mendelian units, so that the different arrangements of the same chromosome can be compared with the different alleles presented in a gene. In addition, the removal of recombination can cause that different arrangements present adaptive differences in individuals who wear them.

1.1.5. THE GENETIC BASIS OF FITNESS VARIATION IN THE CHROMOSOMAL ARRANGEMENTS

Dobzhansky and his colleagues made detailed studies on the chromosomal polymorphism for inversions in natural populations of the North American species *D. pseudoobscura*. In this species the chromosomal polymorphism for inversions is often limited to only one of the four major chromosomes, chromosome 3. Not only simple inversions were found, but also complex chromosomal arrangements formed by overlapping inversions. In *D. pseudoobscura* each chromosomal arrangement is called by the name of the location where it was found for the first time: thus, together with the standard arrangements (*ST*) the Chiricahua (*CH*), the Arrowhead (*AR*), etc. were described. Figure 1.6 shows the geographical distribution of chromosomal arrangements *ST*, *AR* and *PP* (Pikes Peak). To the west of the U.S. the arrangement *ST* is relatively common, but it is rare in Arizona and was not observed in Texas. The arrangement *PP* shows a contrasting pattern in its distribution, while arrangement *CH* has the highest frequency in Arizona and New Mexico (SINNOT, DUNN and DOBZHANSKY 1958). We see that these arrangements are distributed as longitudinal clines (west-east). Dobzhansky was convinced that the different chromosomal arrangements were maintained in populations by selection for the heterozygotes.

Dobzhansky was eager with the utility of population cages to study the evolutionary dynamics of inversions (WRIGHT and DOBZHANSKY 1946). He founded population cages of *D. pseudoobscura* containing different frequencies of two chromosomal arrangements segregating in natural populations (DOBZHANSKY 1948). Parents of the flies used to establish each of the cages were collected in the sites of Piñon Flats, Keen Camp and Mather (California). Piñon Flats and Keen

Camp are close to each other (13 km) but the biotic characteristics of both sites were very different. Mather is about 500 km to the north to the Sierra Nevada. A mixture of flies *from the same locality* was introduced to each cage with a certain proportion of two chromosome arrangements.

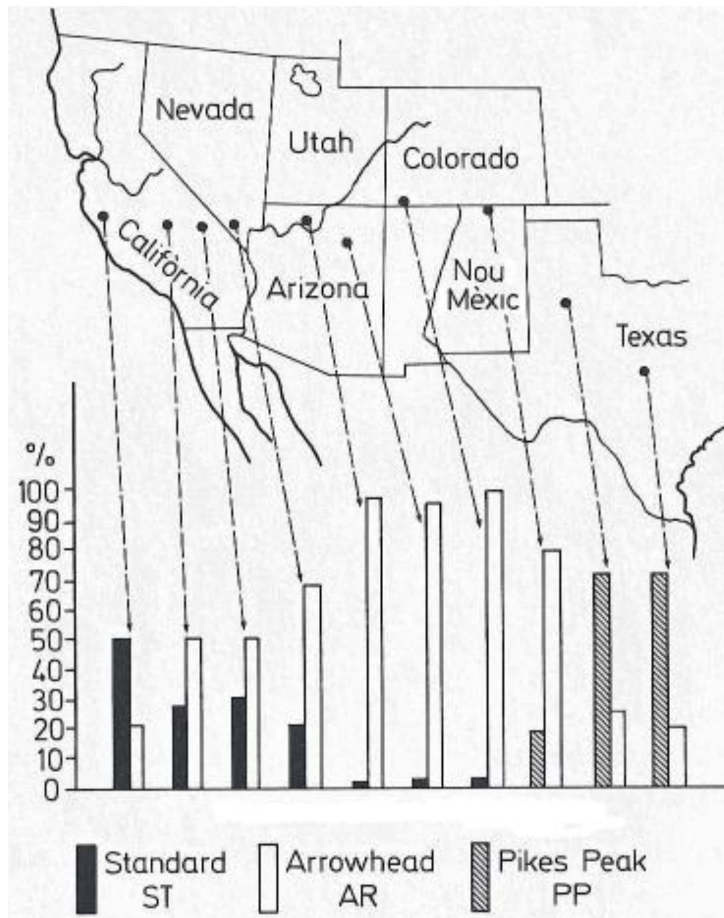


Figure 1.6: Frequency of chromosomal arrangements *ST* (standard), *AR* (Arrowhead) and *PP* (Pikes Peak) of *D. pseudoobscura* at different locations in the U.S. (SINNOT, DUNN and DOBZHANSKY 1958, p. 278).

Every two or three months the samples of eggs were taken from the cage and larvae were kept until they had the optimal conditions; then the giant chromosomes of salivary glands of these larvae were analyzed and the frequencies of the two arrangements were determined. Therefore, the researchers knew the frequencies of each chromosome that were in the cages at different time intervals. Figure 1.7 shows the results obtained for cages 39 and 40, started with different proportions of *AR* and *CH* arrangements from *the same* locality. In both cases a polymorphism was reached without lost of any of the two arrangements. In the case of cage 39, the arrangement *AR* had an initial frequency around 80%, and the cage 40 about 20%. In both cases the frequency

curve approached an equilibrium frequency of *AR* around 60%. Regardless of geographical origin, the heterokaryotypes have higher fitness than the corresponding homokaryotypes. Accordingly, the process of natural selection doesn't determine the elimination of an arrangement and the establishment of another; instead of this it reaches an equilibrium in which both arrangements are maintained with a defined frequency. However, the adaptive value of different arrangements depended on the geographic origin of the flies. Thus, homokaryotypes AR/AR were far superior to homokaryotypes *CH/CH* if the involved chromosomes were from the Piñon Flats population and not if they were, for example, from Mather. This demonstrated that *the chromosomes with the same arrangements but found in different locations had different genetic content*. In certain locations, the genetic content of chromosomes with different arrangements were mutually coadapted by natural selection, which was the cause of the observed heterosis (the greater fitness of heterokaryotypes).

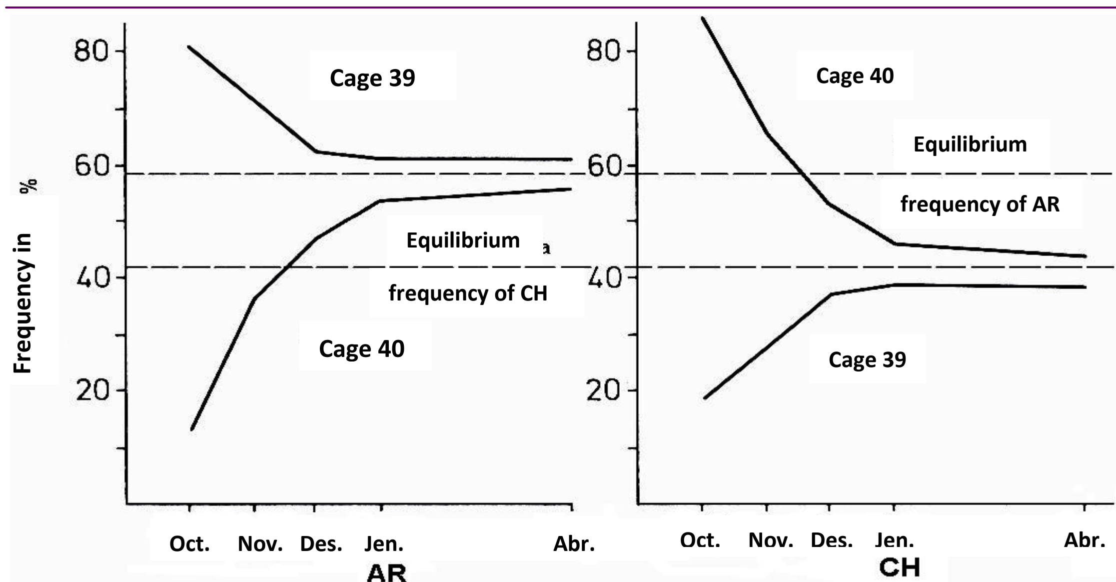


Figure 1.7: Seasonal variation of the frequencies of chromosomal arrangements *AR* (Arrowhead) and *CH* (Chiricahua) in experimental populations of *D. pseudoobscura*. Regardless of the initial chromosomal frequencies with which cages were founded, the equilibrium of frequencies was reached: 60% *AR* and 40% *CH* (cages founded with individuals from a specific natural population; SINNOT, DUNN and DOBZHANSKY 1958, p. 278).

1.1.6. HYPOTHESES ON THE MAINTENANCE OF THE INVERSIONS IN NATURAL POPULATIONS

What is the reason for this heterotic effect observed in heterokaryotypes? Why heterokaryotypes are positively selected? The hypothesis of coadaptation tries to answer these questions. The term *coadaptation* has two aspects. First, the alleles that interact within an inversion are coadapted to each other to produce a particularly adapted genotype (epistatic effects on fitness). Second, at

least initially as the new inversion will be mainly in heterozygous combination, its *block* of alleles, which are held together because of the lack of recombination, must be coadapted with alleles of alternative chromosomal arrangement, so that the heterokaryotype has a higher fitness.

Thus, the presence of gene complexes in the inverted segments of the chromosomes, adapted to different environmental conditions, could explain the rapid effect of natural selection on the inversions and the origin of the latitudinal clines in natural populations of *D. subobscura* (PREVOSTI *et al.* 1985). Despite it is clear that the chromosomal inversions of *D. subobscura* are adaptive; it is not yet known with certainty what mechanism maintains them at certain frequencies. Here we will briefly describe the main hypotheses that explain a selective role of inversions.

The experiments performed by Dobzhansky in population cages provided more evidence for the coadaptation hypothesis based on the genetic content of the inversions (WRIGHT and DOBZHANSKY 1946). As it has been seen, these experiments showed that each chromosomal arrangement in a population is specifically coadapted with alternative arrangements of the same population. Therefore, at least in *D. pseudoobscura*, the adaptive character of chromosomal polymorphism for inversions is very specific, and supports only those chromosomal arrangements that are from the same natural population: a stable polymorphism could be reached when the experiments in population cages were started with the arrangements A and B obtained from *the same* population. When the chromosomal arrangements A and B came from *different* populations the obtained results were unpredictable.

In *D. subobscura* there was an attempt to find evidence that supports the theory of coadaptation, but without positive results (KRIMBAS 1993). For example, MCFARQUHAR and ROBERTSON (1963) and PENTZOS-DAPONTE and SPERLICH (1965) didn´t found neither heterosis (superiority of hybrids) in F_1 and F_2 nor a breakdown of coadapted complexes. Finally, PREVOSTI (1967) observed in the F_2 a possible break of coadapted complexes referring to the wing length but did not found superiority of F_1 hybrids. Besides the coadaptation theory established by DOBZHANSKY (1950), there are several hypotheses, which explain the maintenance of inversions, that have been summarized in two recent articles (HOFFMANN and RIESENBERG 2008; SANTOS 2009) and are detailed below.

Marvin Wasserman in 1968 suggested that the polymorphism for inversions can also be maintained in populations if the heterokaryotypes present a greater biological effectiveness than

both homozygotes (with or without inversion) due to epistatic interactions of *supergenes* combinations or complexes of coadapted genes (WASSERMAN 1968). In addition, Wasserman took into account that structural homozygosity does not mean homozygosity for genes (CARSON 1958). We spoke about homokaryotypes referring to homozygotes for the same arrangement, that in natural populations differ in genetic content unless they provide direct parental relationships. Thus, if more than one combination of supergenes exists for the same chromosomal arrangement in a population, the homokaryotypes present lower biological effectiveness due to the disruption of favorable epistatic combinations by recombination, which among heterokaryotypes would not be affected. This recombination will produce a reduction in overall biological effectiveness of favorable combinations.

This hypothesis could be best fit to the observed data for the chromosome *O* of *D. subobscura* as the linkage disequilibrium studies of seasonal variations of allozymes found evidence for epistatic interactions between loci especially within the O_{ST} arrangement (SANTOS 2009). But due to the high dispersal ability of this species (AYALA, SERRA and PREVOSTI 1989) and the presence of geographical clines in frequencies for O_{ST} inversion, this hypothesis could not serve for the explanation of this phenomenon. The alternative explanations are that different geographic populations could be differentiated by some arrangements and periodic fluctuations mingling populations can induce cyclic changes in linkage disequilibrium due to migration rather than to epistasis (SANTOS 2009).

A more recent hypothesis is that of *local adaptation*, proposed by KIRKPATRICK and BARTON in 2006. According to these authors, the inversions are favored in natural populations since the reduction of recombination in heterokaryotypes maintains a series of alleles linked to inversion and locally adapted, while the epistatic relations between them are not necessary. Without inversions, favorable combinations would be lost by recombination. As epistasis between alleles is not a prerequisite, this hypothesis is more general. The fate of such inversion is almost fixation, and the polymorphism is explained by migration or by deleterious alleles captured in the inversion.

In different studies on *D. subobscura* it was not found any difference for a certain arrangement on the *O* chromosome between two geographically distant populations (ROZAS and AGUADÉ 1990; ROZAS *et al.* 1995, 1999). Besides, a large genetic difference was found between different

inversions of the chromosome *O* in individuals from the same population (ROZAS *et al.* 1999; MUNTÉ *et al.* 2005; SÁNCHEZ-GRACIA and ROZAS 2011; PEGUEROLES *et al.* 2013), suggesting that the model of local adaptation could be appropriate to explain the maintenance of polymorphism in this chromosome. It has also been proposed that the inversion itself could be the target of selection, because of the mutation generated just at the breakpoint. For example, in humans it has been observed that the expression of a specific gene is truncated due to the position effects or direct disruption of the gene (TADIN-STRAPPS *et al.* 2004). The maintenance of these inversions will depend on the biological effectiveness of the offspring.

Adaptive inversion: if the inversion is a carrier of one or more adaptive alleles, natural selection will act on the entire inversion. Underdominance occurs when heterokaryotypes have a lower biological effectiveness than homokaryotypes (with or without inversion). The decrease in biological efficiency may be due to different factors, such as the structural problems in the pairing of heterozygotes or by genetic incompatibilities (KIRKPATRICK and BARTON 2006). A very high frequency of simple crossovers within the inversions has been observed in several plant species, which leads to appearance of unbalanced gametes and sterility problems (RIESENBERG 2001). This process does not maintain the inversion polymorphism in the population, but rather results in the fixation or loss of inversion. Overdominance takes place when the heterozygotes for an inversion have a greater biological effectiveness than any of the homozygotes. The difference with the three mechanisms previously described, where the heterozygote is also positively selected, is that this mechanism causes the superiority of the heterozygote because the alleles forming the heterozygote are deleterious in homozygous state. This means that different inversions may accumulate deleterious recessive alleles and therefore, deleterious effects disappear in heterozygous individuals.

Finally, inversion may also be neutral. In this case, the probability of its dispersal or loss in natural populations depends on their effective size, the genetic drift and migration. This hypothesis has been advocated in cases where significant deviations from the Hardy-Weinberg equilibrium were observed in the inversion polymorphism (ARADOTTIR and ANGUS 2004).

1.2. ADAPTATION TO ENVIRONMENT

Environmental effects influence to a large extent the chances of survival and reproduction of different genotypes in the population. Differential fitnesses underlie the adaptation of species to the environment by progressively sieving unfit individuals arisen at each generation. Climatic conditions determine the geographic boundaries of many species, and temperature is one of the most important variables that limit the distribution and abundance of species (PARMESAN 2006). In eukaryotes, active life is possible between few degrees below zero and fifty degrees above zero. If these limits are exceeded an organism can die due to the freezing of cytoplasm (forming ice crystals that destroy membranes) or the denaturation of proteins in the case of extreme thermal shock. Temperature is one of the most important environmental variables affecting morphological, physiological, behavioral and life-history traits of organisms.

Elevated concentrations of atmospheric greenhouse gases have changed global climate, raising the Earth's surface temperature by 0.74°C in the past century (IPCC 2001). The impact of climate change on global biodiversity is an active area of research, and several studies have reviewed the evidence for the capacity of terrestrial organisms, populations, communities and ecosystems to cope with current climate change, and the upscaling of their responses, from the molecular and genetic level to the levels of community, ecosystem and biosphere. Generally, the adaptation to new environmental conditions triggers physiological and behavioral responses which may have a genetic basis and eventually alter the genetic composition of populations (HOFFMANN and PARSONS 1991). Examples of these adaptive responses were already found at the level of phenology of species. In the last 50 years, many species have extended their distribution ranges poleward. In one study where the information on species, such as density, phenology, morphology and behavior, was related with global warming (ROOT *et al.* 2003), it was determined that 80% of species showed temperature-related changes. These changes had occurred in the expected direction taking into consideration the physiological limits of species. In addition, traits related to migration, reproduction and development have shifted over time at the same rate than the average increase in spring temperatures (PARMESAN 2006). Most responses are attributed to phenotypic plasticity, but recent studies have shown that climate change is leading to genetic changes in the populations of some organisms, mostly involving adjustment to the length of the seasons (BRADSHAW and HOLZAPFEL 2008). Also, the temporal adaptive changes in different species of

Drosophila have been observed at the genetic level in general (LEVITAN 2003; UMINA *et al*. 2005), and particularly in *D. subobscura* (RODRÍGUEZ-TRELLES and RODRÍGUEZ 1998; BALANYÀ *et al.* 2006; 2009).

Species of the genus *Drosophila* have been extensively used as model organisms for studying evolutionary responses to extreme temperatures (MAYNARD-SMITH 1956), as their thermal niches and habitat requirements are quite variable, ranging from species with restricted distributions to cosmopolitan species (HOFFMANN, SØRENSEN and LOESCHCKE 2003).

1.2.1. PHENOTYPIC VARIATION

The presence of phenotypic variation in nature, which then is shaped by the action of selection, is the unavoidable condition for biological evolution. Formally, we can say that selection is the result of three conditions: the existence of phenotypic variation among individuals in a population; the existence of differences in survival and/or reproduction of different phenotypes, i.e. their fitness; and heritability of these differences in fitness (LEWONTIN 1970; CADEVALL 1988; SOLER 2002).

On the other hand, the phenotype is formed by the expression of many different genes, and also by the interactions of genotype with the environment. Thus, the phenotypic variability can be divided into two components: the genetic component or the particular set of genes possessed by the individual and the environmental component, which is the set of all non-genetic causes that influence the phenotypic value. The genotype gives a genotypic value to the individual, but the environment affects this value and renders the final phenotypic value.

In addition, environmental variation can have two effects on the phenotypic variability: *i)* general or external environmental effects, caused by factors of influence (e.g. temperature, salinity, density) that are shared by groups of individuals and *ii)* specific or internal environmental effects caused by residual deviations of the phenotype that could be specified by the genotype bases as well as developmental noise and their interaction with the general environmental effects. Such deviations are unique for individuals and largely unpredictable. The variance associated with these specific environmental effects can be estimated by using fully inbred lines in which there is no genetic variation.

When environmental conditions change, organisms and populations must also be modified to resist the pressure of these changes, so that the development of individuals could be altered to a greater or lesser degree. The following three processes are considered to be control mechanisms of phenotypic variability.

Canalization is considered as strength of morphogenetic constrain (WADDINGTON 1942; GILBERT 2003) by which the development seems to be damped and slight deviations from the genotype or slight perturbations in the environment does not necessarily lead to the production of abnormal phenotypes. Evolutionary geneticists define it as the tendency of characters or traits to develop a reduction in variability (GIBSON and WAGNER 2000; MEIKLEJOHN and HARTL 2002). This process allows the production of a specific or basic phenotype under different environmental and genetic conditions and therefore decreases the inter-individual variance within groups, reducing the genetic susceptibility to environmental conditions. Therefore, the canalization can be estimated by studying the inter-individual variation.

There are two ways to conduct a study of canalization. DWORKIN (2005) defines the first one as the norm of reaction or property of the genome, and the second as an approximate variation. Both definitions lead to different metrics in the study of canalization. Although there is no specific design for canalization experiments, there are experimental treatments that manage these types of studies:

> ***i. Control of the amount of genetic variation among lines or populations****:* Canalization is a property of the genotype and as each individual has a unique genotype, to investigate the canalization of a population under various environments may be somewhat difficult for the differential response of different genotypes. Therefore, it is important to control the genetic variation within lines. In most genetically manipulated systems, and in the species that are produced clonally, it is possible to get individuals closely related by inbreeding or controlled genetic crosses (such as procedures of chromosome extraction in *Drosophila*). Thus, the same genotype can be tested under multiple environments and also it is possible to make a reasonable sampling of the genotype.
>
> *ii.* ***Need for the independent and multiple sampling (through genotypes and not individuals):*** Since each line (i.e. *Drosophila* isochromosomal inbred lines) represents a simple genetic sample, the measurement of multiple individuals within the line increases

essentially the sampling of a simple measure, giving the best estimate. When using multiple lines it is important to consider the independence of the lines. Another way to improve the estimation is the replication of the data, where the same genotype is resampled and the samples are treated as independent.

iii. ***Control of the gene pool by comparing treatments:*** This requirement is essential in the studies of genetic canalization where different chromosomes are compared by their effects of canalization. For genetically manipulated systems it is possible to have different lines, which are absolutely identical except for the markers under study. If the genetic background is not under control, the results of the observed effects of loci (or markers) will be confusing.

The developmental stability is the ability of organisms to dampen random noise that arises spontaneously as a result of stochastic variation in cellular processes that are involved in the development of morphological structures (KLINGENBERG 2004). The genetic control of developmental stability is closely linked to non-additive genetic variation of morphological characters of interest. Dominance and epistasis play an important role in the genetic architecture of developmental stability. The molecular and cellular processes during development are inherently variable, but contribute to the reliable assembly of the intricate organization of the body plan. The mechanisms leading to this level of reliability of the phenotype are known as *developmental stability*. However, the nature of these mechanisms is still not well known. The developmental stability is a complex of a wide class of phenomena of developmental buffer, which also include canalization against genetic and environmental effects.

This degree of resistance against possible disturbances is difficult to measure, so it is easier to quantify its contrary, the developmental instability, which is understood as the uncertainty that leads to the morphological variability even when the genetic and environmental conditions remain constant. It can be conveniently measured as random differences between left and right sides (bilateral asymmetry) in the body or parts of an organism (intra-individual variation). In organisms whose parts or sides are bilaterally symmetrical the fluctuating asymmetry provides an easy way to study the instability of development. Both sides and parts share the same genome and are usually developed under nearly identical environmental conditions and therefore the variation of asymmetry around the average is due to random fluctuations of the developmental processes, and

can be used as a measure of developmental instability (KLINGENBERG 2003).

Plasticity is an ability of an individual genotype to express one phenotype under a series of environmental circumstances and another phenotype under other circumstances. These are alternatives, which permit populations to adapt to the changing environmental conditions and to increase the variability between groups of individuals. Studying independent lines in which multiple individuals have been sampled for one phenotypic value in different environments having similar environmental (internal) variance, it was demonstrated that the plasticity and the canalization are displayed as opposite characteristics of the same phenomena (NIJHOUT and DAVIDOWITZ 2003). Therefore, the plasticity of a phenotype is indicated by a change of the mean of a character through different environments.

In the frame of this work with respect to *D. subobscura* and its thermal adaptation we will consider different phenotypic characters such as body size, thermal preference and thermal resistance and their relationships with chromosomal inversions.

1.2.2. THERMAL TRAITS

As it was mentioned above, temperature is an important climatic variable that determines the distribution and abundance of living organisms, and one of the factors affecting the evolutionary history of individuals. The environmental temperature in many cases can exceed the physiological limits and consequently can be potentially lethal. To cope with the changes in temperature, the organisms may present behavioral (thermal preference) as well as physiological (thermal tolerance) responses.

Thermal preference

The thermal preference (T_p) is defined as the body temperature an organism chooses when provided with a range of potential temperatures. The body temperature of ectotherm animals is strongly linked to the temperature of environment where they live. In many cases, many of them have serious limitations in its regulation through physiological adjustments and therefore adjust their body temperature by behavior (HUEY and PASCUAL 2009). For example, moving from sunny

areas to shaded sites (DILLON *et al.* 2009). The thermoregulatory behavior can be adaptive in two complementary ways: (i) it may help the individual to avoid extreme temperatures, hot and cold, as both can be lethal (NORRIS 1967); (ii) it can increase the time while the individual maintains their physiological optimal temperature (HUEY, HERTZ and SINERVO 2003). In fact, the thermal preference of the species is assumed to closely correspond to the temperature at which the species maximize biological performance.

Thermal tolerance

In the same manner as the thermal preference gives an idea of the temperature at which the fitness of the individual is potentially maximized, the upper knockout temperature (T_{ko}) or maximum thermal tolerance gives the upper temperatures that the species under consideration can tolerate.

Ectotherm animals constitute the vast majority of the terrestrial biodiversity and are likely to be particularly vulnerable to global warming because their basic physiological functions such as locomotion, growth and reproduction are strongly influenced by temperature. The ability of ectotherms to perform these functions at different temperatures is described by a curve of thermal performance (HUEY and STEVENSON 1979), which relates the biological efficiency with temperature. Thus, the curve rises gradually from a minimum critical temperature (CT_{min}) to the optimum temperature (CT_{opt}), and then falls rapidly to the maximum critical temperature (CT_{max}) (Figure 1.8).

The critical temperatures CT_{min} and CT_{max}, defined by the body limits, have been measured in several ectotherm species and generally covary with latitude, reflecting adaptation, at least in part, of ectotherm species to the climate. Therefore, these curves show the direct effect of temperature on the biological efficiency of the organism, and thus provide a framework to elucidate the physiological effects of global climate change in an empirical manner (DEUTSCH *et al.* 2008).

Stressful conditions act as evolutionary force that contributes to the adaptation of natural populations (reviewed in HOFFMANN, SØRENSEN and LOESCHCKE 2003). In this sense, thermal resistance is one of the most important mechanisms that allow organisms to face the increase in temperature variation. For *Drosophila*, it was found that selection favors thermal resistance in

nature (KREBS and FEDER 1997), and that there are correlations between the expression of thermal stress proteins and gradients of environmental stress (FEDER and HOFMANN 1999).

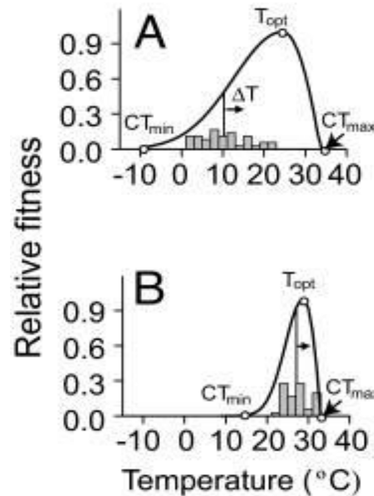


Figure 1.8: **Thermal efficiency curves for species of temperate areas (A) and tropical species (B). In both curves the CT_{min}, CT_{max} and T_{opt}** (the figure corresponds to fig.1 in DEUTSCH *et al.* 2008).

Relationship between thermal traits and chromosomal polymorphism: evidence for coadaptation

Ectotherms pose a suit of behavioral and physiological strategies to cope with spatiotemporal variation in ambient temperature (ANGILLETTA 2009). For instance, behavioral adjustments (e.g. modifying daily activity patterns and selecting favorable microclimates; STEVENSON 1985, HUEY and PASCUAL 2009) can buffer the impact of sub-optimal air temperatures, and are the main means of thermoregulation in small insects (CASEY and HEGEL 1981; KINGSOLVER and WATT 1983; WILLMER and UNWIN 1981). Although such adjustments can enable ectotherms to maintain relatively constant body temperatures (T_b) at different seasons and/or latitudes (STEVENSON 1985; HUEY, HERTZ and SINERVO 2003, HUEY and PASCUAL 2009), the observation of cyclical seasonal changes in genetic markers putatively related to thermal adaptation (DOBZHANSKY 1970; RODRÍGUEZ-TRELLES, ALVAREZ and ZAPATA 1996), and the clinal variation in thermal stress tolerance in some *Drosophila* species (COYNE, BUNDGAARD and PROUT 1983; HOFFMANN, SØRENSEN and LOESCHCKE 2003; HOFFMANN 2010), suggests that behavioral thermoregulation may be insufficient to fully compensate shifts in air temperature.

If behavioral thermoregulation is not fully compensatory and climate variation influences the actual T_b and physiological performance of organisms distributed over broad latitudinal ranges (i.e., performance falls below its optimum during cooling and warming), then temperature is more than just a key environmental factor that affects development, growth, and survival of individuals (SINCLAIR *et al.* 2003; HOFMANN and TODGHAM 2010): it is likely the main selective agent that drives directly or indirectly the evolution of clinal patterns in genetic, phenotypic, and life history traits. Furthermore, the divergence of thermal optima in the different subpopulations according to the T_b experienced by the organism is expected to bolster a covariance between behavioral shifts (thermal preference) and performance (HUEY and BENNETT 1987; HUEY, HERTZ and SINERVO 2003). This is related to the idea of "coadaptation", where natural selection is supposed to favor the harmonious adjustment among the suit of co-evolving traits (DOBZHANSKY 1970). Parallel clines on different continents or along independent temperature gradients can thus offer an invaluable opportunity to study thermal coadaptation since the role of temperature in driving those clines is quite compelling.

Some widespread latitudinal clines in *Drosophila* also provide an additional advantage for studies of thermal coadaptation: there is a relatively well-known historical record following the colonization of a new geographical region (e.g. AYALA, SERRA and PREVOSTI 1989; HOFFMANN and WEEKS 2007). Perhaps the best example is that of *Drosophila subobscura,* a native Palaearctic species that invaded the Americas about 30 years ago, spread rapidly on both continents, and clinal patterns for traits and genetic polymorphisms appeared soon (PREVOSTI *et al.* 1988; HUEY *et al.* 2000; BALANYÀ *et al.* 2003; REZENDE *et al*. 2010). For instance, North American populations have rapidly evolved decreased desiccation resistance with increasing latitude as expected, which matches the pattern found in Old World populations and suggests that strong selection for thermal-related traits along latitudinal gradients is taking place. On the other hand, in South America this trait has an opposite pattern: higher desiccation tolerance is observed in colder areas (GILCHRIST *et al.* 2008). Contrasting outcomes were also observed for other clinally varying traits wing cell size and cell number (CALBOLI, GILCHRIST and PARTRIDGE 2003), and wing shape (AYALA, SERRA and PREVOSTI 1989; GILCHRIST, HUEY and SERRA 2001) where the role of temperature remains elusive, which apparently suggests that selective pressures vary in the different clines. But the actual causes of this variability are unclear. An alternative explanation, however, is that evolution can sometimes be constrained by antagonistic genetic correlations (i.e., genetic correlations among traits that are not

in accord with the direction of selection (BETRÀN, SANTOS and RUIZ 1998; ETTERSON and SHAW 2001) arising from linkage disequilibrium between alleles at different loci, and patterns of linkage disequilibrium can vary among populations or seasons (FONTDEVILA *et al.* 1983; RODRÍGUEZ-TRELLES 2003). In this context, we now know that contrasting wing shape clines in *D. subobscura* came out as a correlated response of the world-wide parallel inversion clines (BALANYÀ *et al.* 2003) because inversion-shape relationships in native and colonizing populations are opposite (presumably due to the different associations between inversions and particular alleles which influence the trait), probably as a result of the bottleneck effect that occurred during the colonization of America (FRAGATA *et al.* 2010). Besides, different patterns of linkage disequilibrium could result from variability in migration rates between genetically differentiated populations in the various latitudinal clines (NOSIL *et al.* 2006). In summary, conflicting outcomes between old and rapidly evolving new clines should probably not be viewed as a nuisance, but as reminder that an appropriate knowledge of the underlying genetic architecture is required to further understand why (or why not) those inconsistencies arise.

More specifically, if behavior "drives" the subsequent parallel evolution in morphology and physiology as predicted (HUEY, HERTZ and SINERVO 2003; DUCKWORTH 2009), it is essential to analyze the genetic basis of thermal preference and temperature-related traits to see whether or not thermal coadaptation can happen along a cline. It has been recently undertaken a within-population large-scale study in our laboratory to analyze the association between chromosomal inversion polymorphisms that show parallel latitudinal clines in native and colonizing populations of *D. subobscura*, with the thermal preferences and knockout temperatures of their carriers (REGO *et al.* 2010). The main results can be summarized as follows: (*i*) "cold-adapted" or "cold-climate" gene arrangements (i.e., those gene arrangements in all five major acrocentric chromosomes that show a negative correlation coefficient with maximum temperatures along the cline, or a positive correlation coefficient with latitude in Palaearctic populations; MENOZZI and KRIMBAS 1992; KRIMBAS 1992) were associated with a lower T_p and T_{ko}, in accordance with the natural patterns; (*ii*) different chromosomes were responsible for the bulk of the genetic variation in T_p (chromosomes *A* and *O*) and T_{ko} (chromosome *E*); and (*iii*) T_p and T_{ko} were phenotypically uncorrelated, which agrees with the observation that different independently segregating chromosomes were mainly responsible for the corresponding associations. Taken at a face value, behavioral thermoregulation and performance were indeed "coadapted" in the sense that cold-climate (warm-climate) gene

arrangements collectively favor a coherent response to colder (warmer) environments, but this was not due to a covariance of behavior and physiology.

There were, however, two potential limitations in the study. First, each individual fly was scored for only one chromosome of its diploid set and, hence, dominance effects (if any) where hidden in the analysis. Second, both intra- and interchromosomal contributions were mixed because the assayed flies had the genetic background from the sampled wild population. Although it could be claimed that this protocol is closer to what happens in nature, the problem is that there was a lot of noise in the experiment, which might have precluded a better characterization of the underlying genetic effects. Thus, although the amount of genetic variation on T_p and T_{ko} explained by the combined effect of all chromosomes carrying at least one cold-climate gene arrangement was statistically significant, it only accounted for 1% of the total phenotypic variation (REGO *et al.* 2010).

In this thesis we examine coadaptation (in the "genetic covariance" sense) of *T*p and T_{ko} by taking advantage of the fact that the polymorphic inversions on chromosome *O* appeared to be associated with behavioral thermoregulation in *D. subobscura* (REGO *et al.* 2010), and that this is the only chromosome that can be used to measure the expression of associated traits in replicated inbred or outbred genotypes. Chromosome *O* harbors several genes involved in the heat shock response; in particular, gene *Hsp*68 (maps on sections *O*(88E)-*O*(89B)) and relatively close to the proximal breakpoint of inversion O_8; MOLTÓ *et al.* 1992, CUENCA *et al.* 1998), and gene *Hsp*70 (maps on section *O*(34A) and it is included in the warm-climate gene arrangement $O_{\underline{3+4}}$; MOLTÓ *et al.* 1992; CUENCA *et al.* 1998). *Hsp70* appears to be the primary protein involved in thermotolerance in *D. melanogaster* (PARSELL, TAULIEN and LINDQUIST 1993) though apparently not in other *Drosophila* species (KREBS 1999), and *Hsp70* allele frequencies show latitudinal clines and change in response to thermal evolution in the laboratory (BETTENCOURT *et al*. 2002). In addition, correlated responses to selection for knockdown resistance at 39ºC have also been found for *Hsp68* (MCCOLL, HOFFMANN and MCKECHNIE 1996). Previous work showed that *D. subobscura* flies carrying *O* chromosomes derived from replicated thermal lines (SANTOS *et al.* 2004; 2005) that had evolved in the laboratory at warm temperatures (22ºC) had a higher total net fitness than its cold-adapted (13ºC) counterparts; that is, a significant shift in thermal optima was observed (SANTOS 2007). All in all, it seems that there is indeed a room for the coevolution of behavior and performance in *D. subobscura*.

1.3. *DROSOPHILA SUBOBSCURA* AS A MODEL SPECIES FOR STUDY OF CLINAL ADAPTATION

“All animals are equal, but some animals are more equal than others”
George Orwell

1.3.1. *DROSOPHILA SUBOBSCURA*

What makes *Drosophila subobscura* an ideal model for study of clinal variation? There are a number of reasons, which will be consistently explained below. *Drosophila subobscura* (Figure 1.9) is a Palearctic species that recently (about thirty-five years ago) colonized the two American continents. J. E. Collin described this species in the 30s of the last century. It is a *Diptera* species from the group *obscura*, which is distributed almost all over Europe, except central and northern Scandinavia. It has also been observed in North Africa to Sahara and in the Middle East to Iran, but its distribution in the former USSR is not precisely known. Finally, it can be found on the Canary Islands, Azores and Madeira (Krimbas 1993; Figure 1.10).

Very rich polymorphism for inversions was observed in all acrocentric chromosomes of the species. More than 60 different inversions, which are grouped in more than 90 chromosomal arrangements (different complexes of inversions) in more than 150 natural populations, have been described and analysed so far (Krimbas 1992). The *O* chromosome is the longest chromosome in *D. subobscura* (23.3% of the euchromatic portion; i.e. 28 Mb out of a total of 120 Mb; Laayouni *et al.* 2007), and is homologous to arm 3R in *D. melanogaster* (Powell 1997). It is a chromosome that has greater polymorphism for inversions, as up to 26 different inversions have been described for this chromosome (Balanyà *et al.* 2003). The existence of the homokaryotypic strain *ch-cu* characterized by morphological recessive markers on the *O* chromosome cherry eyes (*ch*) and curled wings (*cu*), and highly homogeneous genetic background fixed for the standard chromosomal arrangements in all major acrocentric chromosomes but chromosome *O* where it is fixed for arrangement $O_{\underline{3+4}}$ ($O_{\underline{3+4}}$, J_{ST}, U_{ST}, E_{ST} and A_{ST}) (Figure 1.12a; Koske and Maynard-Smith 1954), allows analyzing the chromosomal polymorphism in this species (Lankinen and Pinsker 1977).

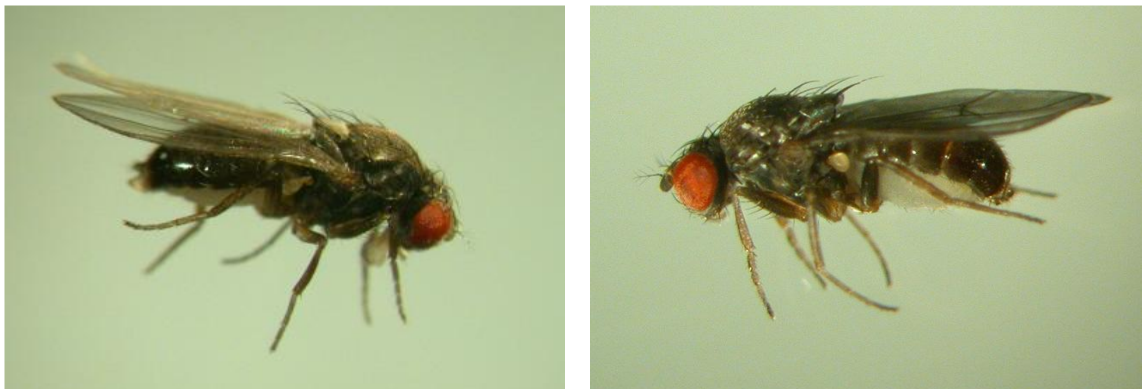


Figure 1.9: The images of *D. subobscura*, on the left a female and a male on the right (from PEGUEROLES 2010).

Figure 1.10: Geographic distribution of *Drosophila subobscura* (from RODRÍGUEZ–TRELLES, RODRÍGUEZ and SHEINER 1998).

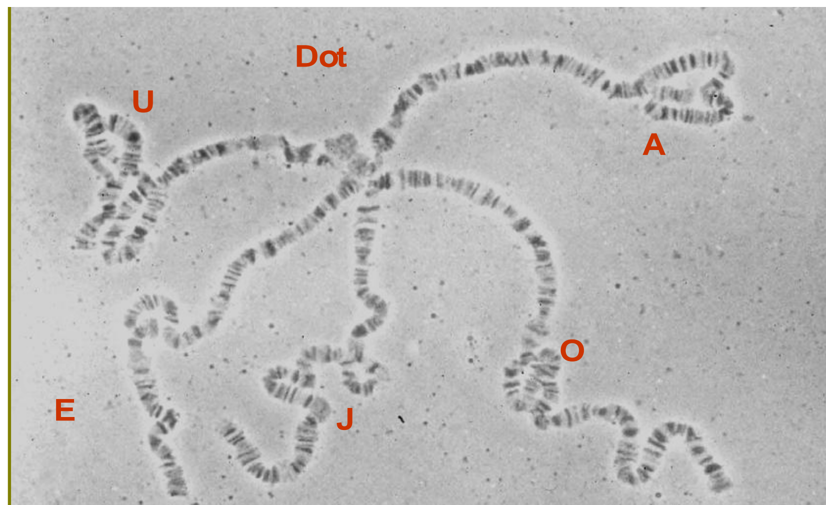


Figure 1.11: ***D. subobscura* has six pairs of mitotic chromosomes, five acrocentric and one dot.** MAINX, KOSKE and SMITAL (1953) identified five large chromosomes by the vowels of the alphabet, *A* being attributed to the sex chromosome.

The chromosome *O* is the only one for which a balancer stock (*Va*/*Ba*: *Varicose*/*Bare*; Sperlich *et al.* 1977, Figure 1.12b) is available (a balancer is a specially constructed chromosome that carries a dominant morphological marker that is lethal in homozygous state and has multiple inversions to suppress recombination), which allows obtaining homokaryotypic and heterokaryotypic lines for the chromosome *O* following a pattern of suitable crosses (Mestres *et al.* 1990; Mestres, Sanz and Serra 1998). This stock carries the dominant lethal genes *Varicose* (*Va*; produces irregular thickenings at the junctions of wing veins in heterozygotes) and *Bare* (*Ba*; a mutant that reduces the number of macrobristles on scutum, scutellum and head in heterozygous flies) on the *O* chromosome and was derived from the *ch-cu* strain. The *Va* balancer chromosome carries two (X-ray induced) overlapping inversions (named $O_{VIII+210}$) plus the naturally occurring arrangement $O_{\underline{3+4}}$. About two thirds of the *O* chromosome (segment *SII*) is covered by the $O_{VIII+210}$ inversions; most of the rest (segment *SI*) is covered by the O_{ST} arrangement. Thus, there is no need to conduct the difficult consanguineous crosses, essential for establishment of homozygous lines of any other chromosomes. The strain *Va/Ba* was obtained by Diether Sperlich (Sperlich *et al.* 1977) and its main features can be found in Mestres and Serra (2008). It has been observed that the majority of inversions in *D. subobscura* are not found alone in nature, but form complexes with other inversions. These complexes may be formed by overlapping inversions, such as $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+23}}$, or by non-overlapped inversions as $O_{\underline{3+4}+1}$ and $O_{\underline{3+4}+7}$. The continuous line below the numbers indicates that inversions are overlapped. Figure 1.13 includes a photograph of a heterokaryotype for one of these arrangements on chromosome *O*, and an explanatory scheme that facilitates the interpretation. Some of these arrangements are rare and found only in specific areas, but at least two or more arrangements of each chromosome are widely distributed throughout the range of the species and their frequencies have a clinal variation, which is correlated with latitude (Prevosti 1966).

In *D. subobscura*, the choice of standard gene arrangements returns to the chromosomal map published by Mainx, Koske and Smital (1953), who used a strain of *D. subobscura* from the Swiss town of Küsnacht, available in the laboratory at that time and which was homokaryotypic for all chromosomes. Later Kunze-Mühl and Müller (1958) improved this map, dividing it into 100 sections and 405 subsections and also showing the break points of the inversions. Krimbas (1993) revisited this map by adding break points for new inversions described later and it is normally used today (Figure 1.14).

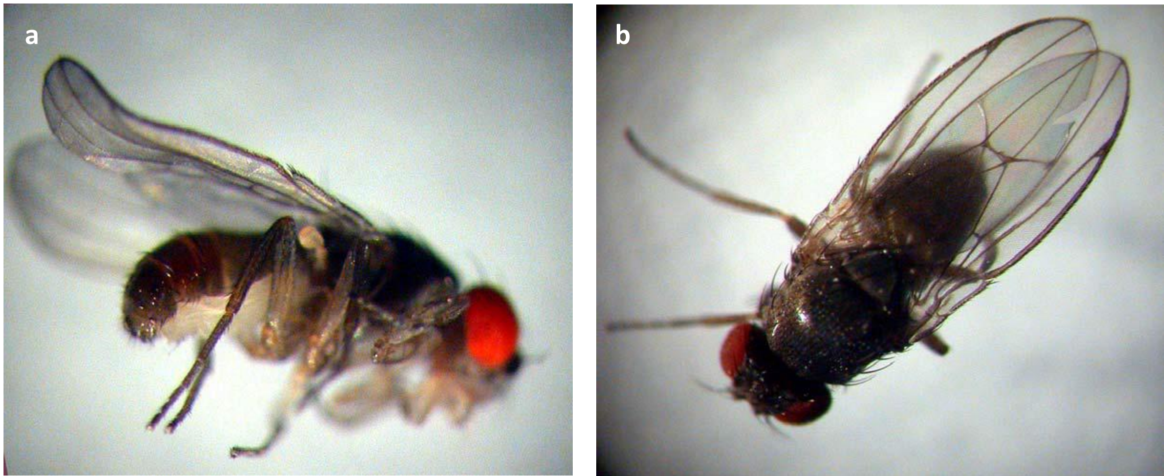


Figure 1.12: **Mutant flies from different laboratory stocks.** a) a fly from the *ch-cu* strain. The curled wings and cherry-colored eyes are observed; b) a fly from the *Va*/*Ba* balancer stock, which has "varicose" wing veins and reduced number of macrobristles.

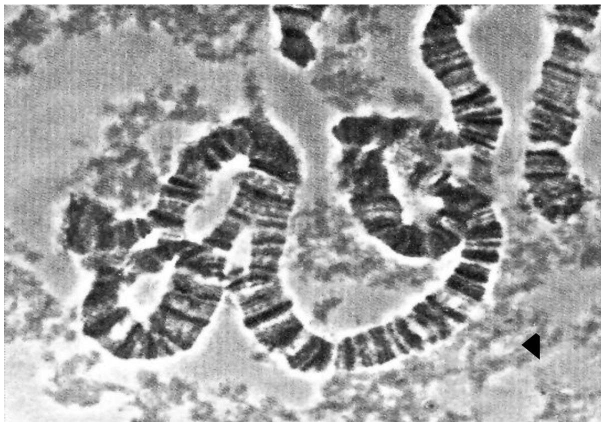
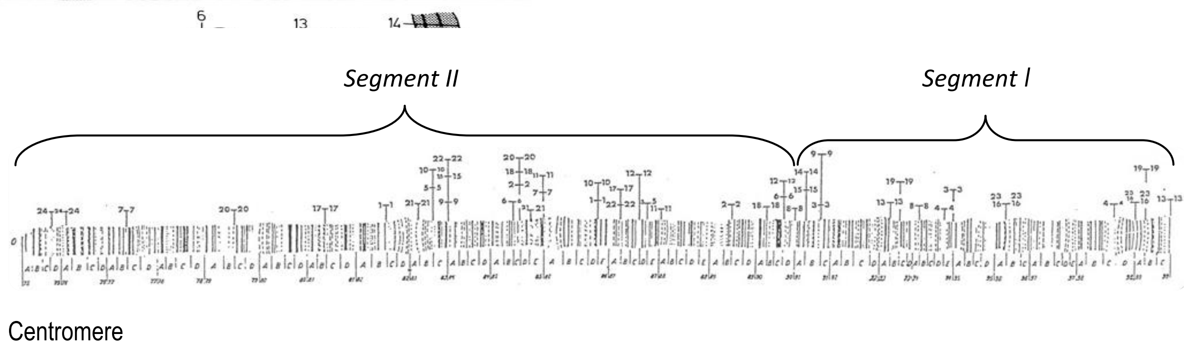


Figure 1.13: **Complex loops of inversions in the giant chromosomes of *D. subobscura*** (corresponds to the Figure 1 from SERRA 2009). The picture corresponds to an $O_{\underline{3+4}+1}/O_{ST}$ heterokaryotype. In the scheme, the trajectory of the normal chromosome is represented by a solid line and the trajectory of chromosome inversions O_1 and $O_{\underline{3+4}}$ with a dashed line. The origin of complex arrangement $O_{\underline{3+4}+1}$ by a series of successive inversions is explained at the bottom of the diagram.



Centromere

Figure 1.14: The cytological map of the *O* chromosome of *D. subobscura*. The 25 sections (with their subsections) and the breakpoints of the main chromosomal inversions are represented.

An early comprehensive summary of the inversion polymorphism of *D. subobscura* from native Palearctic populations revealed several patterns in the geographical distribution of chromosomal arrangements, with the so-called standard arrangements in the five acrocentric chromosomes increasing in frequency with latitude (KRIMBAS and LOUKAS 1980). The standard arrangements are

common in northern populations but its frequency decreases gradually towards the south. Other chromosomal arrangements which differ from the standard by one or more inversions are not observed in northern populations and its frequency increases towards the south. The authors clearly concluded, however, that disentangling adaptive explanations from purely historical processes that could generate those patterns (i.e., northward migration of the populations of *D. subobscura* after the last glacial period) would require further research.

Luckily enough, an unintentional natural experiment was going on at that time, when *D. subobscura* was first discovered in South America at the locality of Puerto Montt (Chile) in February 1978 (BRNCIC and BUDNIK 1980) and explosively colonized a broad latitudinal range (BRNCIC *et al.* 1981; PREVOSTI *et al.* 1985). In 1982, it was detected in the North American west coast (BECKENBACH and PREVOSTI 1986). The two colonization events were not independent but sequential (PASCUAL *et al.* 2007), so that the colonization of *D. subobscura* represents a natural experiment with two replicates, which is a very interesting case for studying evolutionary questions (BRNCIC and BUDNIK 1980; AYALA, SERRA and PREVOSTI 1989). *Drosophila subobscura* has two peaks of growth throughout the year, more important in spring and another with lower intensity in autumn (SERRA, PEGUEROLES and MESTRES 1987; PASCUAL *et al.* 1993; ARGEMÍ *et al.* 2003). It is a species with a great adaptive flexibility, since it has been collected in regions with different climatic and habitat conditions (MONCLÚS 1964; CODINA and PÉREZ 1980). The high abundance, which makes their capture easier, and the rich polymorphism for chromosomal inversions of these flies make this species to be a reference in the field of evolutionary genetics, as it is its American counterpart *D. pseudoobscura*. In ecological terms *D. subobscura* is a quite unknown species although considered a generalist (SHORROCKS 1982), but instead it is easily developing in the laboratory. Its life cycle is about 20 days under optimum conditions of larval density and temperature (18°C) (ORENGO and PREVOSTI 1994). Therefore, all these features make it a species widely used in genetic, ecological and evolutionary studies and biology in general (KRIMBAS 1993; POWELL 1997).

1.3.2. ADAPTIVE VALUE OF CHROMOSOMAL POLYMORPHISM FOR INVERSIONS IN *D. SUBOBSCURA*

If the evolutionary trajectories are rapid or slow, gradual or graded and predictable or contingent it has been a much-debated topic in Evolutionary Biology. The evaluation of the microevolutionary trajectories of replicated sets of natural populations in a geographic scale is rarely feasible

(BALANYÀ *et al.* 2003). The cases in which this can be done include those species that have recently invaded different geographical areas. In this case a way of demonstrating that there was a fast, uniform and predictable microevolutionary change, in geographic scale, is to check whether invasive populations are rapidly evolving independent geographical clines parallel to the existing stocks of the original range of the species. If, on the contrary, the populations develop invasive idiosyncratic clines, then we should give more importance to genetic drift as a factor responsible for the observed variation.

Drosophila subobscura is a species suitable for evaluating microevolutionary trajectories (AYALA, SERRA and PREVOSTI 1989). The inversion polymorphism in this species has been extensively studied in over a hundred natural populations in the Palearctic region (KRIMBAS and LOUKAS 1980; SPERLICH and PFRIEM 1986; KRIMBAS 1992; MENOZZI and KRIMBAS 1992; KRIMBAS 1993; BALANYÀ *et al.* 2006). As have been mentioned, many chromosomal arrangements present latitudinal clines in the Palearctic region. In addition, time-series data obtained in several European sites indicate that the frequencies of some arrangements, typical for warm latitudes, have increased in frequency over the years (ORENGO and PREVOSTI 1996; RODRÍGUEZ-TRELLES and RODRÍGUEZ 1998; SOLÉ *et al.* 2002). The most obvious climatic factor that consistently varies with latitude and, hence, seems to be the *prima facie* gradient causing the clines in *D. subobscura* is temperature. Thirty five years ago the species colonized large areas (> 15º latitude) in both North and South America. Colonizing populations have been exposed to the similar climatic gradients that the original European populations. Therefore, the populations of the Old World provide the basic patterns of evolutionary trajectories represented by the latitudinal variation of the chromosomal inversions frequencies. Although the American coasts just got a small number of colonizers (MESTRES *et al.* 2005, PASCUAL *et al.* 2007), its dispersion led to the formation of latitudinal clines for chromosomal inversions on the American continents. Both colonized areas have the same gene arrangements (PREVOSTI *et al.* 1988), which were transported by the sample of individuals-colonizers (two of the chromosome *A*, two of the *J*, three of the *U*, five of the *E* and six of the *O*). These gene arrangements are the most common in the Palearctic region, except for the O_5 inversion. Additionally, it was described that the arrangements O_5 and $O_{\underline{3+4}+7}$ have heterotic effect in American populations (MESTRES *et al.* 2001).

In fact, until then it was not clear whether latitudinal clines that have many chromosomal arrangements in Europe were due to natural selection (adaptation) or could be explained by

historical causes (migration). Considering an example of this later possibility we can suppose that a chromosomal arrangement appears in any location. In principle, if the arrangement is not lost, it extends between the descendants of those parent individuals in which the mutation was produced. Therefore, the highest frequency of that arrangement will be in the area of its origin. Obviously the arrangement can be extended to other areas further away, by migration, but then we expect to see the decreasing of its frequency moving away from the source area, i.e., we get a cline of its frequency, which would not be due to selective factors, but to the demographic "history" (also derived processes) of the arrangement. Thus, colonization of America by *D. subobscura* has provided a unique opportunity to determine whether its chromosomal polymorphism is adaptive.

A few years after colonization, Prevosti with colleagues analyzed the chromosomal polymorphism in colonizing populations (in 1981 in South America and North America in 1985-86); there was a surprisingly rapid evolution at geographic scale (PREVOSTI *et al.* 1985; 1988). Even in these early tests, some chromosomal arrangements had already developed latitudinal clines in their frequency, equivalent to those existing in the area of Palearctic origin. A second analysis of South American populations, in 1986, suggested that these clines were still evolving and were more similar to those existing in Europe (PREVOSTI *et al*. 1990). The appearance of these clines in the New World cannot be explained by historical causes and, therefore, the explanation must lay in their adaptive nature. The evolution of these clines was therefore very fast and seemed predictable in the sense that knowing the European clines, one can "predict" the possible development of similar clines in new geographical areas with similar climatic gradients. However, the new clines did not consistently continue to converge on the Old World baseline. The recent survey of Chilean populations of *D*. *subobscura* in our laboratory shows that inversion clines have basically faded or have even changed sign with latitude (CASTAÑEDA *et al.* 2013). In this study the authors tried to explain the observed patterns testing the hypothesis that this fading of inversion clines might be due to the Bogert effect; namely, that flies' thermoregulatory behavior has eventually compensated for environmental variation in temperature thus buffering selection on thermal-related traits. An important exception may involve temperature extremes (HOFFMANN 2010), which could be the real selective agent underlying chromosomal inversion clines in *D*. *subobscura* (REGO *et al.* 2010; CALABRIA *et al.* 2012; HUEY and PASCUAL 2009). Interestingly, of the three distribution areas of *D*. *subobscura* included in the time-series analysis of inversion clines (fig. 1 in PREVOSTI *et al.* 1988), extreme thermal events have been recorded in the Palaearctic region and in North America but

they are not so frequent in South America (ALEXANDER *et al.* 2006). The results were consistent with the idea that active behavioral thermoregulation might buffer environmental variation and reduce the potential impact of thermal selection on other traits such as chromosomal arrangements. Another evidence of adaptive nature of inversion polymorphism in *D. subobscura* to temperature conditions comes from the recent change of inversion frequencies on three independent continents correlated with global climate change (BALANYÀ *et al.* 2006).

Taking into account that the paracentric chromosomal inversions in *Drosophila* inhibit recombination, maintain favorable combinations of genes and they are not negatively selected because they do not reduce the biological effectiveness, its permanence in the population will depend on its genetic content (POWELL 1997). Thus, the presence of gene complexes in the inverted segments of the chromosomes, adapted to different environmental conditions, could explain the rapid effect of natural selection on the inversions and the origin and establishment of the latitudinal clines in natural populations of *D. subobscura* (PREVOSTI *et al.* 1985). So, despite it seems to be clear that the chromosomal inversions of *D. subobscura* are adaptive; it is not yet known with certainty which of the mechanisms described in section 1.1.6 is the closest to the truth.

The low level of gene transfer between inversions (*gene flux*) for genes located inside the inverted region observed in some *Drosophila* species is in agreement with both Coadaptation and Local Adaptation hypotheses (LAAYOUNI *et al.* 2003; SCHAEFFER *et al.* 2003; MUNTÉ *et al.* 2005; HOFFMANN and RIESEBERG 2008). However, despite the fact that Dobzhansky detected a lower fitness of heterozygous individuals from different populations of *D. pseudoobscura* in laboratory experiments (DOBZHANSKY 1950), molecular studies failed to detect genetic differentiation within inversions sampled from different populations (SCHAEFFER *et al.* 2003). In *D. subobscura,* high genetic differentiation between European populations was detected when chromosomal arrangements were used as markers, since their frequency widely varies between populations (KRIMBAS 1993). However, given that these chromosomal arrangements are under strong selection (PREVOSTI *et al*. 1988; BALANYÀ *et al*. 2006), gene flow between populations would likely be underestimated using the inversions themselves as markers. Interestingly, low levels of genetic differentiation between European populations of *D. subobscura* were observed using molecular markers such as RFLPs (ROZAS *et al.* 1995) and microsatellite *loci* (PASCUAL *et al.* 2001). Consequently, *gene flow* (defined as allele exchange between populations) and *gene flux* (defined

as allele exchange during meiosis in heterokaryotypic females) could be changing the genetic content of inversions from widely separated populations. Thus, the analysis of candidate genes undergoing selection in *D. subobscura* could allow contrasting different hypotheses explaining the maintenance of inversion polymorphism in populations.

1.3.3. CHROMOSOMAL ARRANGEMENTS AS THE GENETIC MARKERS OF GLOBAL CLIMATE CHANGE

Climate change is altering the geographic distributions, the abundances, the phenology and the biotic interactions between organisms. It also can alter the genetic composition of species, but the quantification of this requires genetic time-series data. The historical records of the frequencies of chromosomal arrangements allow evaluating the genetic sensitivity to the climate change and to other environmental factors. In the case of the species *D. subobscura*, there were time-series data (from thirteen to forty years) of the frequencies of chromosomal arrangements and the data of temperature in twenty-six sites (thirteen of Europe, seven of North America and six from South America). BALANYÀ *et al.* (2006) checked if the temperatures had increased over the years in those places and also if the frequencies of chromosomal arrangements typical for warm latitudes changed. They wanted to study whether the magnitude and direction of genetic changes (changes in the frequencies of the arrangements) were parallel to changes detected in temperature, and whether this happened on the three continents. Historical data of gene arrangement frequencies in the twenty-six analyzed locations were obtained from the observations published by different authors.

Between 1997 and 2004 the researchers had obtained data from the same populations (or very near) for which the historical data were available. In all samples they analyzed the content of chromosomal arrangements for each of the five acrocentric chromosomes of the species. Instead of analyzing the changes of the frequencies of the individual arrangements, a chromosome index (Ch_{PC1}) was used. To determine whether there had been a change in temperature during the time between the historical and recent samples collection an index of temperature (T_{PC1}) was also developed, based on average monthly temperatures recorded at the weather stations located nearby to the places of the sample collections during four years prior to each sampling. The temperature index, T_{PC1}, as expected, negatively correlated with latitude on the three continents (Figure 1.15). It has increased significantly between the historical and recent samples, and this also

happened on the three continents, which is consistent with the global warming trends. In fact, T_{PC1} has increased in twenty-two of twenty-six analyzed localities. The changes were larger in Europe, which may be due to the fact that a longer period of time between the collections of historical and recent samples has gone in the Old World, and because of the existence of a wider range of climates. The chromosome index is inversely related not only with latitude (Figure 1.16), but also with T_{PC1} on three continents, so it serves as a genetic indicator of the local climate. In twenty-four of twenty-six analyzed sites, the chromosome arrangements associated with warmer latitudes (high values of Ch_{PC1}) have increased in frequency between the historical and recent samples. In a given locality, the frequencies of arrangements and the temperature have become more "equatorial". When the magnitude of these changes shifted in terms equivalent to degrees of latitude, the observed changes in the frequencies of arrangements and the temperature on three continents may be considered equivalent to the displacement of the historical samples ~ 1 ° of latitude closer to the equator.

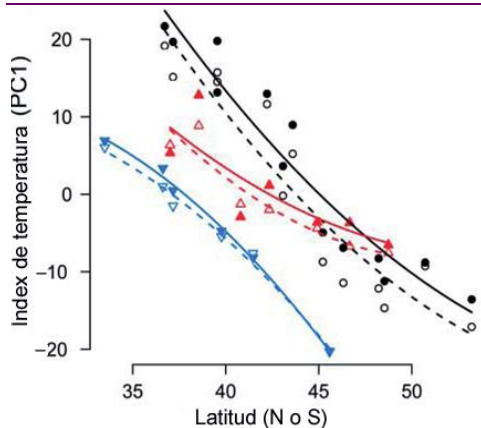


Figure 1.15: The rate of temperature (T_{PC1}) is inversely correlated with latitude in the twenty-six localities analyzed on the three continents, and it has increased significantly between the historical samples (dashed regression curves) and recent samples (black circles, curves regression constant). The European sites are represented by black symbols, those of North America, red, and South America, blue. The regression curves correspond to orthogonal polynomials of second degree.

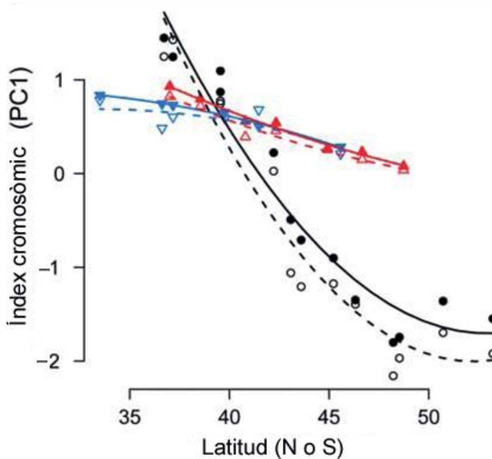


Figure 1.16: The chromosome index (Ch_{PC1}) is inversely related to latitude and has increased between the historical samples (dashed regression curves) and recent samples (continuous regression curves). The European sites are represented by black symbols, those of North America by red, and South America by blue. The regression curves correspond to orthogonal polynomials of second degree.

1.3.4. THERMAL ADAPTATION IN THE LABORATORY

Previous work in our lab has investigated the putative role of temperature *per se* by examining the evolutionary responses of chromosomal inversions and genetic markers in three replicated populations of *D. subobscura* cultured at three temperatures (13, 18, and 22º C) spanning much of the tolerable range for this species (SANTOS *et al.* 2004; 2005; LAAYOUNI *et al.* 2007). A quick and consistent shift in gene arrangement frequencies in response to thermal selection regime was observed, but the trends were generally inconsistent with simple climatic-based explanations of worldwide latitudinal patterns and suggested *in situ* shifts of inversion frequencies in response to global warming. Of course laboratory experiments are not the best way to reconstruct natural clines, but the results suggested that temperature alone does not easily account for the worldwide clines of inversion polymorphism in *D. subobscura*. As mentioned above, it appears that the latitudinal clines of *D. subobscura* respond to temperature changes. For this reason, the next experiment was accomplished to obtain candidate genes for thermal adaptation (LAAYOUNI *et al.* 2007).

Global gene expression was measured from the 3-fold replicated laboratory thermal selection stocks using cDNA microarrays with *D. melanogaster* clones (LAAYOUNI *et al.* 2007). A total of 306 (6.6%) cDNA clones were identified as 'differentially expressed' (following a false discovery rate correction) after contrasting the two furthest apart thermal selection regimes (i.e., 13°C *vs*. 22°C). Analysis of functional categories defined by the Gene Ontology project pointed to an overrepresentation of genes involved in carbohydrate metabolism, nucleic acids metabolism and regulation of transcription among other categories. The location of differently expressed genes was approximated with respect to *D. melanogaster* genome project by considering conservation of chromosomal elements among *Drosophila* species. Furthermore, 88 markers were physically mapped by in situ-hybridization to the polythene chromosomes of *D. subobscura* showing that a larger than expected number mapped inside inverted chromosomal segments. In view of the rapidly and consistently evolved latitudinal clines in chromosome inversion polymorphism following the New World invasion by *D. subobscura*, and the shifts in inversion frequencies in response to laboratory thermal adaptation and to climate change, these results were not at all unexpected (LAAYOUNI *et al.* 2007). The problem is, however, that linkage with inversions will highly complicate the identification of chromosome regions that are targets of selection. Nevertheless it is interesting

to determine which of the identified candidate genes may play a role in thermal adaptation of *D. subobscura*, analyzing them at the nucleotide level and considering each chromosomal arrangement independently.

1.3.5. STUDIES OF INVERSIONS IN *D. SUBOBSCURA* AT THE MOLECULAR LEVEL

The effect of inversions on the nucleotide variability depends mainly on two factors: age of the inversions and gene flow (NAVARRO, BARBADILLA and RUIZ 2000). At the same time, gene flow can be seen influenced by the adaptive value of inversions (HOFFMANN and RIESENBERG 2008) and also by the size of inversions (NAVARRO *et al.* 1997). In *D. subobscura* molecular studies were performed to try to resolve these issues; the vast majority of those have focused on *Segment I* (*SI*) of the chromosome *O* and on the sex chromosome (*A*). In *D. subobscura*, the nuclear *rp49* (ribosomal protein 49) gene region, which maps at section 91C very close to one of the breakpoints of the O_3 inversion in *segment SI* of the chromosome *O*, presents high genetic differentiation between the $O_{\underline{3+4}}$ and O_{ST} chromosomal arrangements, which allowed determination of gene conversion tracts (ROZAS and AGUADÉ 1994). When including two further chromosomal arrangements, $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+23}}$, in the_analysis of the *rp49* gene region, significant differences were found in all comparisons between the arrangements, although the highest were in the comparisons involving the inversion O_{ST} (ROZAS *et al.* 1999). They also quantified the gene conversion using the method developed by Esther Betràn and coauthors (BETRÀN *et al.* 1997). It was found that the gene flow was not completely removed inside the inversion, and its magnitude depended on the distance of the marker from the inversion breakpoint. Both RFLP analysis (ROZAS and AGUADÉ 1990; ROZAS *et al.* 1995) and direct sequencing (ROZAS and AGUADÉ 1993, 1994; ROZAS *et al.* 1999) have shown that the $O_{\underline{3+4}}$ arrangement has a higher level of silent polymorphism at the *rp49* gene region than O_{ST}. On the other hand, comparison of this gene region between *D. subobscura* and the closely related species *D. madeirensis* and *D. guanche* indicates that the O_3 is the ancestral inversion (RAMOS-ONSINS *et al.* 1998).

The *Acph1* gene and *Fmr1* (located also in *segment SI*) have also shown high genetic differentiation between the arrangements O_{ST} and $O_{\underline{3+4}}$ (NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a; PEGUEROLES *et al.* 2013). A combined analysis of eight different genes located inside the arrangements O_{ST} and $O_{\underline{3+4}}$, showed that the levels of variability were similar regardless of the

location of the genes used, with a high differentiation between arrangements and significant levels of intra- and inter-locus linkage disequilibrium (MUNTÉ *et al.* 2005). The high genetic differentiation, even in the central parts of the inversion, is attributed to the absence of even number of crossovers or to the fact that the recombinant chromosomes were eliminated by natural selection to maintain coadapted gene complexes. Finally, some studies have been conducted on the gene *Odh* of *D. subobscura*, located in the *Segment II* inside of inversions O_1, O_5 and O_7 and out of the inversion O_2 but on the edge of its breakpoint, analyzing a fragment of 800 nucleotides. With this molecular marker it was intended to study new aspects of the American colonization by *D. subobscura* (MESTRES *et al.* 2004; GOMEZ-BALDÓ *et al.* 2008). Only 11 different haplotypes were found in the populations of the new continent, probably as a result of founder effect, and some of them were clearly associated with certain chromosomal arrangements (O_5, $O_{\underline{3+4}+2}$ and $O_{\underline{3+4}+7}$). In contrast, many different haplotypes were observed in the analyzed Palearctic populations (Barcelona, Spain and Mt. Parnes, Greece; ARAUZ *et al.* 2011), and only partial associations were detected in the cases of gene arrangement $O_{\underline{3+4}+1}$ and the O_5 inversion. This latest inversion presented only two haplotypes, and one of them is completely associated with this inversion in the American populations (MESTRES *et al.* 2004; GOMEZ-BALDÓ *et al.* 2008). No differences between O_{ST} and $O_{\underline{3+4}}$ arrangements were detected for genes located in *Segment II* (PEGUEROLES *et al.* 2013).

In the first study of the sex chromosome of *D. subobscura*, the A_{ST}, A_1 and A_2 arrangements were analyzed for one region of 5.4 Mb, which includes the gene *yellow* (MUNTÉ, AGUADÉ and SEGARRA 2000). The analyzed fragment is located very near to the breakpoint of the inversion A_1, and outside of inversion A_2 and accordingly its genetic differentiation was significant only in comparison with the inversion A_1. Recently, another study on the sex chromosome was conducted, especially on the inversions A_{ST} and A_2, but this time using information from five nuclear genes (NOBREGA *et al.* 2008). The inversion A_2 is simple and of medium size (41.3 cm) with a genetic length considerably greater than the theoretical minimum to form even number of crossovers (20 cM). Although significant genetic differentiation was detected in five genes, it depended on the location of the gene, and the maximum difference was found for the gene nearest to the break point of the inversion. Therefore, despite the presence of gene flow, especially in the central parts of inversion, it was not sufficient to homogenize the genetic content of the two inversions.

According to Dobzhansky's coadaptation hypothesis, genetic differentiation should be expected between different inversions as well as between different populations for the same inversion. Several population studies have been carried out in *D. subobscura* at the molecular level. Four populations were analyzed by restriction mapping of the gene *rp49* in the gene arrangements O_{ST} and $O_{\underline{3+4}}$ (ROZAS *et al.* 1995). Three of these populations were from Europe and four from the island of Tenerife. There were no differences between the populations of the continent, and the only significant difference was due to the population of Tenerife. Two *Drosophila subobscura* populations from both sides of the Mediterranean region, Mt. Parnes (Greece) and Barcelona (Spain) were analyzed in order to get further insight into this subject (PEGUEROLES *et al.* 2013). In both populations, analyses were also focused on O_{ST} and $O_{\underline{3+4}}$ arrangements. Nucleotide diversity levels were obtained for six nuclear genes (*Pif1A, Abi, Sqd, Yrt, Atpα* and *Fmr1*) located across the *O* chromosome. Interestingly, the only gene located in *Segment I* inside the inverted region (*Fmr1*), showed high nucleotide differentiation between both arrangements. However, no significant genetic differences were detected within arrangements between Mt. Parnes and Barcelona populations for none of the six genes, indicating high levels of gene flow between populations. Similarly, these two populations shared different haplotypes for the *Odh* gene and in one case they even presented the same chromosomal arrangement $O_{\underline{3+4}+1}$ further supporting extensive gene flow among localities (ARAÚZ *et al.* 2011). In 2001 another population study was published at the molecular level, this time with 10 microsatellite loci, not necessarily related to inversions; where five European and two North American populations were compared (PASCUAL *et al.* 2001). As a result genetic differentiation was not found among European populations, which also suggested a high gene flow between populations. Similarly, high genetic similarity within chromosomal arrangements of the A, J and U chromosome were detected for microsatellite loci among latitudinally distant localities (SIMOES *et al.* 2012).

Several studies were conducted to reconstruct the genetic map of the *O* chromosome of *D. subobscura* and to characterize the recombination rates in different parts of this chromosome. The recombination rate was estimated in two types of homokaryotypic lines carrying O_{ST} and $O_{\underline{3+4}}$ inversions and genetic map was made using 13 microsatellite markers in the work of PEGUEROLES *et al.* (2010a). The obtained genetic maps present similar length: 184 cM and 196 cM for O_{ST} and $O_{\underline{3+4}}$ respectively. Despite having detected recombination rate homogeneity across the chromosome, they observed significant regional differences. Several recombination hot- and cold-

spots were detected, and their numbers were different in the homokaryotypic lines (O_{ST} and $O_{\underline{3+4}}$). This variability could be attributed to differences between the genetic content of the two arrangements or to differences between the lines.

The same authors analyzed whether recombination is inhibited inside and outside inversions in three types of *D. subobscura* heterokaryotypes coming from three distant populations by using 12 microsatellite markers distributed along the *O* chromosome (PEGUEROLES *et al.* 2010b). Heterokaryotype descendants were always in higher frequency than homokaryotypes, whether recombinant or non-recombinant, even though the strains came from very distant populations supporting both the local adaptation hypothesis (KIRKPATRICK and BARTON 2006) and, in part, the coadaptation hypothesis (DOBZHANSKY 1950).

1.3.6. CHROMOSOMAL ARRANGEMENTS SELECTED FOR THE PRESENT STUDY AND THEIR ORIGIN

The *O* chromosome is the longest with 25 sections (KUNZE-MÜHL and MÜLLER 1958). It was subdivided in two segments: the distal segment *SI* (sections 91–99), and the proximal segment *SII*, which occupies the remaining two-thirds of the total length of the chromosome *O* (Figure 1.14). Our studies were focused on the three most frequent overlapping arrangements found in the first segment: $O_{ST,}$ $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$.

The arrangements of the *SI* exhibit clear latitudinal clines in Europe, O_{ST} is the prevalent arrangement in northern populations, while $O_{\underline{3+4}}$ is more frequent in southern populations. The $O_{\underline{3+4}}/O_{ST}$ system of *Drosophila subobscura* presents several distinctive features that make it especially suitable to detect the action of selection on chromosomal polymorphism through the study of nucleotide variation. First, the $O_{\underline{3+4}}$ and O_{ST} chromosomal arrangements differ by two overlapping inversions (inversions 3 and 4) that arose independently on the ancestral O_3 arrangement (RAMOS-ONSINS *et al*. 1998), which is now not found alone in *D. subobscura*, but it is present in the closely related species *D. madeirensis* and *D. guanche* (Figure 1.17). This independent origin (which could be regarded as sampling a single O_3 chromosome twice) would result in an initial lack of nucleotide variation within each arrangement and in the initial presence of fixed differences between arrangements. Second, the existence of parallel latitudinal clines for these arrangements, both in Europe (KRIMBAS 1992) and in the recently colonized areas of North

and South America, would support their adaptive character (PREVOSTI *et al.* 1988). Third, the $O_{\underline{3+4}}$-O_3-O_{ST} complex would conform to the Wallace rule of triads for partially overlapping inversions (WALLACE 1953; KRIMBAS 1992). According to this rule, elimination of the central member of a chromosomal triad would contribute to more efficiently maintaining longer coadapted gene complexes, since genetic exchange would be greatly reduced between the two external arrangements. It is worth considering that O_3 and O_4 inversions are never found alone in natural populations of *D. subobscura*. Fourth, there is evidence of strong genetic differentiation between O_{ST} and $O_{\underline{3+4}}$ at loci near the distal break point of inversion O_3 (ROZAS and AGUADÉ 1993, 1994; NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a). And fifth, the rather old age of O_{ST} and $O_{\underline{3+4}}$ (ROZAS and AGUADÉ 1994) suggests that recombination may have eroded the initial association between nucleotide variants and chromosomal arrangements. These features, and particularly the derived character of both arrangements and their age, differentiate the O_{ST}/$O_{\underline{3+4}}$ inversion system from others where variation at multiple regions has been surveyed (HASSON and EANES 1996; LAAYOUNI *et al.* 2003; MOUSSET *et al.* 2003; SCHAEFFER *et al.* 2003).

On the other hand, $O_{\underline{3+4+8}}$ does not show a clinal pattern, traditionally being restricted to the Mediterranean area and being the most abundant in Northern Africa (PREVOSTI 1974). However, in the last decades its distribution changed dramatically and recent surveys revealed frequencies as high as 22.6% in Groningen (BALANYÀ *et al.* 2004). Furthermore, differential basal expression of *Hsp*70 gene, candidate for thermal adaptation, was detected between carriers of these three arrangements from the same population (CALABRIA *et al*. 2012).

Chromosomal phylogeny in *D. subobscura* (KRIMBAS 1992) indicated that $O_{\underline{3+4}}$ is a central arrangement from which several arrangements, including $O_{\underline{3+4+8}}$, originated, giving rise to the $O_{\underline{[3+4]}}$ inversion complex. Therefore, it can be inferred unambiguously that this arrangement is younger than $O_{\underline{3+4}}$. The presence of such a complex system of gene arrangements in *Segment I* of the *O* chromosome of *D. subobscura* offers an opportunity to further analyze the putative effect of chromosomal inversions on nonsynonymous variation. Indeed, differences in the effective size of these arrangements are expected as a result of differences in their frequency, and even in their age. The effective size of young arrangements might be strongly affected by the extreme bottleneck produced in their origin, which should cause a reduction in the long-term effective size of the arrangement (NAVARRO-SABATÉ, AGUADÉ and SEGARRA 2003). All these factors taken together

make these three specific gene arrangements particularly suitable to test the association between temperature, candidate genes for thermal adaptation and clinal distribution of chromosomal inversions.

It has been classically considered that inversions have a unique origin, that is, that they are monophyletic (POWELL 1997 and references therein), due to the low probability of generating two simultaneous breaks at exactly the same positions independently in different chromosomes. In *Drosophila subobscura* the region including the *rp49* gene, which codes for a ribosomal protein and located very close to the proximal break point of inversion O_3 (AGUADÉ 1989) were envisaged as ideal to study the origin and differentiation of the different chromosomal arrangements present in that region of the *O* chromosome (ROZAS and AGUADÉ 1990). In that study levels of variation were compared between O_{ST} and O_{3+4} that differ by two overlapping inversions, and between O_{3+4} and O_{3+4+8} that differ by a single inversion. In the pooled data, variation, measured as nucleotide and haplotype diversity, was highest for O_{3+4}, which also harbors the highest number of restriction site and length polymorphisms. This suggested that O_{3+4} may be ancestral arrangement. This ancestral character of O_{3+4} *vs.* O_{ST} had already been proposed based both on the geographical distribution of the different chromosomal arrangements (PREVOSTI 1971, 1972) and on seasonal variation of linkage disequilibrium (FONTDEVILA *et al.* 1983) and supported in the further studies of the same authors (ROZAS and AGUADÉ 1993, 1994). The *rp49* region locates in a central position of the inversion loop between arrangements O_{3+4} and O_{3+4+8}, whose sequences did not cluster in the gene tree according to their gene arrangements as an evidence for the transfer of genetic information (both by double crossovers or gene conversion) between them (ROZAS and AGUADÉ 1993). SHARP and LI (1989) however suggested that *rp49* would be a slow evolving gene; in this case lower than average levels of variation within species would be expected. The study of RAMOS-ONSINS *et al.* (1998) definitely established the ancient character of O_3 inversion in relation to O_{ST} and O_{3+4}, basing on the phylogenetic trees reconstruction, analysis of parsimony-informative sites and comparison of D_{XY} values of the *rp49* gene as neutrally evolving.

The gene genealogy based on the *Acph-1* gene region (NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a), where all O_{ST} lines clustered together as all O_{3+4} lines also did, clearly supported their monophyletic character, which was also inferred from variation at the *rp49* gene region (ROZAS and AGUADÉ 1994). Therefore, both arrangements were affected at some time in the past by the

extreme bottleneck implied by their origin. Consequently, nucleotide polymorphisms present in each arrangement either have originated independently by mutation or have been incorporated by genetic exchange, most likely by gene conversion (ROZAS and AGUADÉ 1994; ROZAS *et al.* 1999; NAVARRO *et al.* 1997), between O_{ST} and $O_{\underline{3+4}}$, and even between any of them and O_3.

In the work published by ROZAS *et al.* 1999, where the origin and evolutionary history of three gene arrangements were considered, the topology of the gene tree for all O_{ST} and for all $O_{\underline{3+4}+X}$ sequences clearly resembled a star phylogeny, that is, a phylogeny where the tree is stretched near the terminal nodes and compressed near the root. That was the topology expected for populations or markers that have recently expanded from a very small size and are therefore in the transient phase to equilibrium. During this phase, the specific footprint left by the expansion should be detected in the pattern of nucleotide variation. No footprint would be detected, however, if the elapsed time since the expansion were long enough (e.g., more than 4*N* generations).

Due to the unique origin of inversions, a particular gene arrangement increases in frequency (and therefore expands) from one copy to its current frequency in the population. In this case, the expansion of at least some extant arrangements was probably associated with the extinction of the ancestral O_3 arrangement. The observed pattern of nucleotide variation would indicate, therefore, that the time since the origin of the particular inversion has not been long enough to reach equilibrium. The negative values of Tajima's *D* and of Fu and Li's *D* and *F* statistics and the Poisson shape of the pairwise difference distribution (and the corresponding small values of the raggedness statistic) observed by ROZAS *et al.* 1999 might also support this interpretation. Nevertheless, in populations of constant size, Tajima's *D* and the raggedness *r* statistics are a function of the intragenic recombination level. It was shown that the higher the recombination parameter, the lower the raggedness *r* values and the lower the variance of Tajima's *D* statistic. In general, either the raggedness *r* statistic or Tajima's *D* statistic was significantly different from the expected values. This allows to conclude, therefore, that the observed pattern of nucleotide variation within gene arrangement still reflects the expansion of the corresponding inversion since its origin (Figure 1.17).

Dating of chromosomal arrangements of *D. subobscura* has been made, using the nucleotide content of some sequenced genes (Table 1.1). The first obtained dating was for the gene arrangements O_{ST} and $O_{\underline{3+4}}$. The estimates were obtained using the silent variability (i.e.,

synonymous and non-coding positions) of more divergent sequences (removing those with gene conversion) and using *D. pseudoobscura* as an outgroup (ROZAS and AGUADÉ 1994). Then, using a different methodology, the age of the inversions O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4+8}}$ was recalculated. This time it was taken into account the average silent variability of all sequences assuming that the variation within a gene arrangement is not at equilibrium, removing those that had gene conversion (ROZAS *et al.* 1999). In addition, *D. guanche* and *D. madeirensis* were used as outgroups and two supposed times of divergence for these species from different studies were applied. The age of the arrangements O_{ST} and $O_{\underline{3+4}}$ was re-dated using the gene *Acph1* also located in the segment *SI* (NAVARRO-SABATÉ, AGUADÉ, M., SEGARRA, C. 1999a). The age was estimated using the method described in ROZAS *et al.* (1999), and obtained dating was quite similar. The age of the arrangements O_{ST} and $O_{\underline{3+4}}$ based on the mean silent nucleotide diversity of the gene *Sqd* and considering the divergence time of *D. subobscura* and *D. pseudoobscura* was quite similar (PEGUEROLES *et al.* 2013). Overall, older estimates were obtained using the information from the two more distant individuals, in relation to the values obtained using the mean silent nucleotide diversity (Table 1.1). Within each approach, the age estimates for $O_{\underline{3+4}}$ and O_{ST} were quite similar nonetheless, older coalescent times were detected for $O_{\underline{3+4}}$ for most of the studies carried out till now indicating that the methodology, genes and populations used might affect the outcome.

Because of the geographic distribution of the gene arrangements, some of them with a clinal distribution, the expansion hypothesis should be more appropriately contrasted with a stepping-stone model. However, both the estimated population size of *D. subobscura* (10^7; COMERON 1997) and the estimated times for the origin of the different inversions (Table 1.1) support the hypothesis that variation within gene arrangement has not yet attained equilibrium (i.e., the time of origin should be much lower than 4*N* generations).

On the other hand, the observed pattern of nucleotide variation could also reflect an expansion of the whole species, but in this case all loci in the genome would show the same pattern of variation. Nucleotide variation at the region encompassing the two *Acp70A* genes of *D. subobscura* (CIRERA and AGUADÉ 1998), which is located in a region not affected by inversions, did not show, however, negative values of Tajima's *D* statistic. When analyzing five additional genes not located within inversions (*Pif1A, Abi, Sqd, Yrt* and *Atp*) only one did not show negative values although

differences were detected between populations (Barcelona and Mount Parnes) with most of the genes supporting a recent population expansion only in the Greek population.

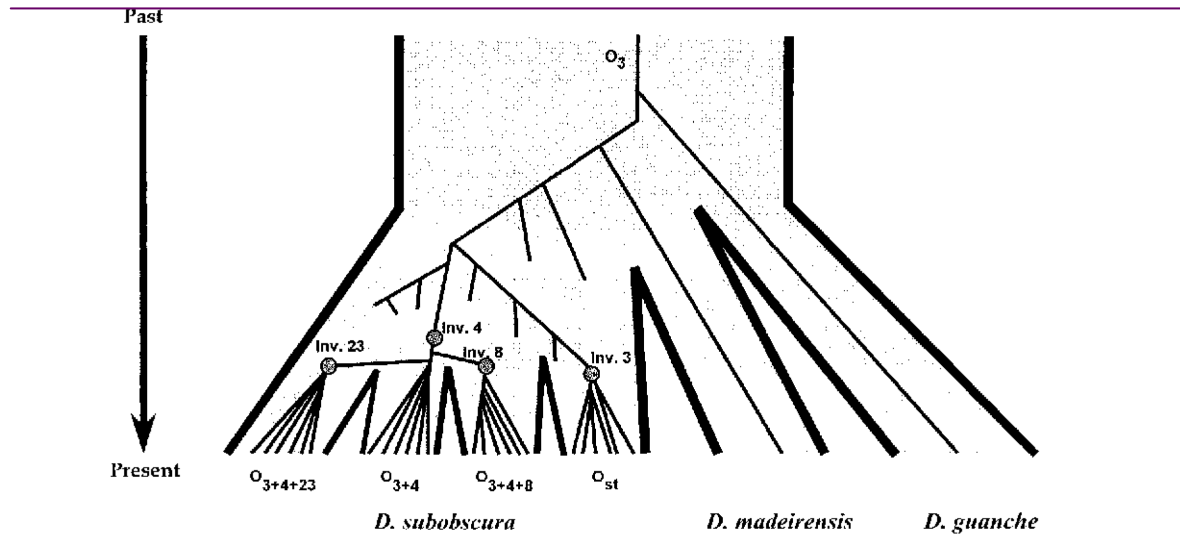


Figure 1.17: Schematic representation of the history of the studied inversions and of the *rp49* genealogy in the different gene arrangements of *Segment I* of the *O* chromosome of *D. subobscura*. The genealogy, which is stretched near the terminal nodes and compressed near the root (star phylogeny), tries to reflect the expansion of the extant gene arrangements studied. Circles indicate the origin of inversions. Corresponds to the Figure 6 from ROZAS *et al.* 1999.

Table 1.1: The age of gene arrangements O_{ST}, O_{3+4} and O_{3+4+8} inferred from variability of genes inside of inversions of *Segment I* of the *O* chromosome with different approaches: a) from mean nucleotide diversity; b) from the two most divergent individuals.

Age of arrangement, Myr			Gene	Reference
O_{ST}	O_{3+4}	O_{3+4+8}		
0.58	0.73	-	*rp49a*	ROZAS and AGUADÉ 1994
0.24	0.33	0.34	*rp49b*	ROZAS *et al.* 1999
0.26	0.31	-	*Acph-1b*	NAVARRO-SABATÉ *et al.*1999b
0.49	0.44	-	*Fmr1b*	PEGUEROLES *et al.* 2013
0.77	1.01	-	*Fmr1a*	PEGUEROLES *et al.* 2013

PART 2

OBJECTIVES

You say you want a revolution
Well, you know
we all want to change the world.
You tell me that it's evolution,
Well, you know
we all want to change the world...
You say you got a real solution
Well, you know
we'd all love to see the plan...

John Lennon and Paul McCartney
"Revolution 1" (1968)

2

OBJECTIVES

The aim of my thesis is related to one of the most intriguing aspects of latitudinal clines; namely, the adaptive role of inversions. The demonstration of the adaptive value of the chromosomal polymorphism is a paradigm in evolutionary theory, but the selective mechanisms involved in its maintenance (e.g. multiple niche selection, overdominance, etc.) are still a matter of dispute. Deciphering of these mechanisms involves molecular studies of the genic content of chromosomal rearrangements and its recombinational dynamics. Traditionally three selective hypotheses have been advanced to explain the maintenance of the chromosomal polymorphism, according to the level of operation of natural selection: chromosomal, individual genes and coadapted genes ("supergenes"). It has been difficult to distinguish among these hypotheses, although some relevant results on seasonal variation favour the coadaptation hypothesis. I focused on this problem using two main approaches. First we wanted to distinguish between karyotypic selection (chromosomal), stating that each rearrangement is adapted to a particular environment, and supergenic selection, that asserts that interacting epistatic genes included in inversions are responsible for local adaptation. The clinal variation of *D. subobscura* chromosomal polymorphism and the availability of molecular markers gave us an excellent opportunity to solve this enigma. If there is an association between haplotypes and inversions, and if this association is identical along the cline, we cannot dismiss that the associated haplotype is maintained due to a historical hitchhiking by the rearrangement. On the other hand, if there is haplotypic clinal variation inside each inversion, it is most probable that this association can be explained by its selective value responsible of local adaptation.

The specific objectives are detailed below:

1. To sample seven European populations located along the latitudinal gradient covering the whole range of *Drosophila subobscura* and focusing on the Iberian Peninsula. Some of the populations

had been previously analyzed and therefore the collections were conducted at similar time to avoid possible effects of seasonal variation.

2. To measure the inversion frequencies from seven European populations (Málaga, Valencia, Perelló and Barcelona (Spain), Montpellier and Dijon (France), Groningen (The Netherlands) forming a latitudinal cline in order to compare them with the previous data.

3. To localize the candidate genes for thermal adaptation on the *O* chromosome of *D. subobscura* by *in situ* hybridization.

4. To describe the gene flux among three chromosomal arrangements of *Segment I* of the chromosome *O* (O_{ST}, $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$), genetic flow between populations of Málaga and Barcelona and to detect the footprint of selection associated with these arrangements, analyzing their genetic content and diversity by sequencing two candidate genes (*larp, Fmr1*) for thermal adaptation and their regulatory regions as well as four other genes (*Acph-1, Ast, trus* and CG5961), which map inside of inverted segments and of some of them.

5. To characterize the effects of gene arrangement (warm-adapted *vs*. cold- adapted), inbreeding and developmental temperature on the temperature preference and heat stress resistance using isochromosomal lines derived from population of Barcelona.

PART 3

MATERIAL AND METHODS

MATERIAL AND METHODS

3.1. MATERIAL COLLECTION

We focused on seven natural populations of *Drosophila subobscura* that are known to be representative for the Palearctic cline in chromosome arrangements.

In October 2007 flies were collected from a natural population near to Barcelona (la Font Groga, Km 8 l'Arrabassada road: 41° 43' N, 2° 13' E; elevation 389 meters). The collecting area preserves a good example of the Mediterranean forests of pine (*Pinus pinea*) and oak (*Quercus ilex*) trees, and typical brushwood. This area, Tibidabo hill, has been a common place for collecting *D. subobscura* flies for many genetic studies (QUINTANA and PREVOSTI 1990a, b; ORENGO and PREVOSTI 1996, 1999, 2002; MESTRES *et al.* 1994; ARAÚZ *et al.* 2008). The outbred stock consisted of 81 males and 205 females and was used to isolate independent *O* chromosomes.

The rest of Iberian populations from Valencia (89 males and 34 females), Málaga (39 males and 130 females) and Perelló (59 males and 93 females) were collected in October 2008, and were also used for the *O* chromosome isolation. The population from Valencia was sampled in a forest outside the residential area of La Canyada, basically formed by pines (*Pinus pinea*). We tried to repeat the capture made in Riba-Roja in 1998, but the riverbed was completely destroyed by a recent flooding. Málaga population was captured in abandoned nurseries belonged to INFOCA (Centro Operativo Provincial de Lucha contra Incendios Forestales, de la Junta de Andalucia), in the N-340 between Málaga and Torremolinos. The vegetation was composed mainly of pines (*Pinus pinea*), oaks (*Quercus ilex*) and eucalyptus (*Eucalyptus camaldulensis*). The population of Perelló was sampled in a patch of the forest located at 11km from the road between El Perelló and Rasquera. The forest consisted of pines (*Pinus pinea*) and its undergrowth (*Rubus* spp.).

Populations from Groningen (Netherlands; 180 males and 148 females), Montpellier (France; 154 females and 37 males) and Dijon (France; 117 males y 185 females) were collected from the end of August to the beginning of September in 2009, and males were crossed with the *ch-cu* strain to

determine chromosomal arrangements. The population of Montpellier was captured at Park at Mas Prunet. An eighteenth century urban park of 9000 m^2 consisted of evergreens and deciduous trees. The capture of the population from Dijon (France) took place in the Park Colombière, the largest park created in the sixteenth century in Dijon and renovated in 1978, consisted of both deciduous and perennials trees. The sample of Groningen was captured in the park Noorderplantsoen, an English garden style park located in the area of the old fortifications.

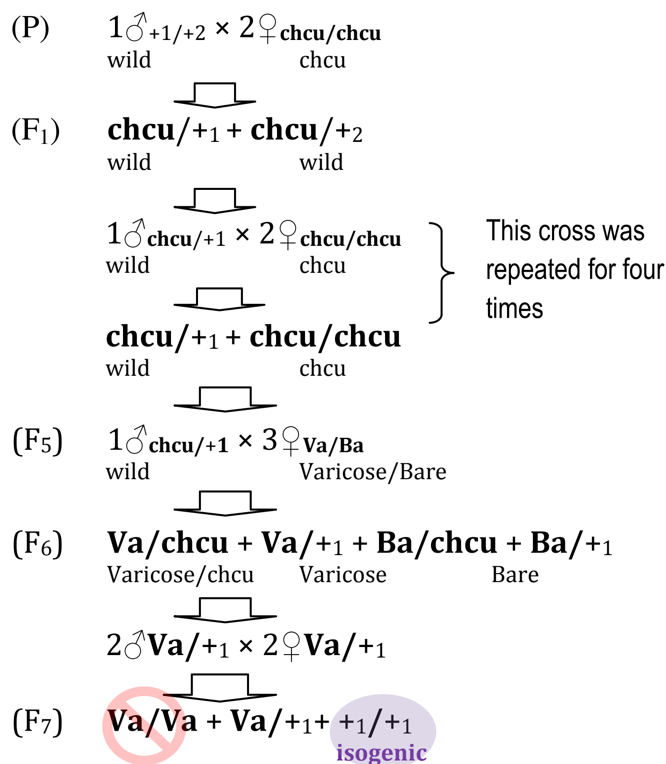
Flies were netted over baits consisting of fermenting bananas. To prepare the bait, the banana were put in a bowl, cut into small pieces and mixed with dry yeast (MONCLÚS 1964), as it was demonstrated that the smell is very attractive for these dipters. In each site of capture the traps were placed in 10-15 m from each other for about 10 meters in total. Once placed, they were left for an hour before the first round of capture and they were emptied every half an hour by a capture hose. The first counting of captured drosofilids was made directly in the field. The females of *D. subobscura* were placed in individual tubes to establish the isofemale lines and males were kept in the bottles with the culture medium until the arrival to the laboratory, where they were used for crosses. Once in the laboratory, both wild males and males descended from wild females were crossed individually with 3-4 virgin females from the *chcu* strain (KOSKE and MAYNARD-SMITH 1954; Figure 3.1).

3.2. ESTABLISHMENT OF THE HOMOKARYOTYPIC LINES

In *D. subobscura* it is difficult to recognize the gene arrangements in all possible homozygous/heterozygous combinations. Therefore we needed to cross wild-type males with *chcu* strain to visualize the wild karyotype of their offspring. Besides, in order to sequence directly the candidate genes on the wild *O* chromosome and to conduct the experiments on thermal traits we needed to establish the homokaryotypic lines, which present two copies of a wild chromosome. The procedure used to obtain isochromosomal lines for the *O* chromosome in an otherwise homogeneous genetic background is shown schematically on the Figure 3.1. It was carried out by subsequent crosses of each sampled fly with the *ch-cu* and *VaBa* strains (see their description in 1.3.1 of Introduction). The scheme of MESTRES *et al.* 1990 and MESTRES, SERRA and AYALA 1995 was followed.

Thus, wild-type males were individually crossed to three or four virgin females from the *ch-cu* marker strain in order to characterize the *O* chromosome arrangement. Similarly, each wild female was placed in an individual vial for egg laying and one offspring male was crossed to *ch-cu.* A single F_1 progeny from each cross was backcrossed to *ch-cu* females, and the scheme was repeated for five generations to homogenize the genetic background. After generation F_2 the lines were identified for the gene arrangement carried by the wild *O* chromosome by microscope inspection of up to eight third-instar larvae salivary gland squashes to look for inversion loops in polythene chromosomes.

Figure 3.1: Scheme for obtaining homokaryotypic lines by crossings with *chcu* and *VaBa* strains. The genotype $+_1/+_2$ indicates initial combination of wild chromosomes.



This allowed identification of the isolated *O* gene arrangements carried by the founding wild-type males. A number of flies from each line were frozen at -80ºC for further DNA extraction. These flies were used to study nucleotide diversity at a number of candidate genes mapped on the *O* chromosome by *in situ* hybridization in order to know their position in relationship with the different inversions (see below). After five generations of crosses with the *ch-cu* marker strain the isochromosomal lines were obtained. Thus, a single male from each selected line was crossed to

four virgin females from the *Va*/*Ba* balancer marker stock (SPERLICH *et al.* 1977). Because expression of the *Ba* gene is highly variable and affected by modifiers located on the *O* chromosome (ALVAREZ *et al.* 1981), we relied only on the *Va* marker to obtain the isochromosomal lines. Even if some recombination had occurred between the balancer and the wild chromosome during the extraction procedure, knowledge of the haplotype of the balancer chromosome excluded any artefactual scoring of wild chromosome haplotypes. Homokaryotypic lines were established from the final crosses ♂$O_{Va\ ch+cu}$ /O_{++++} × ♀$O_{Va\ ch\ cu}$/O_{++++}, and some individuals from each line were also frozen at -80 ºC for further DNA and RNA extraction. The lines were kept at 18°C under constant 12:12 light: dark periods in 130-mL bottles containing 50 mL of David's killed-yeast *Drosophila* medium (David 1962) and at controlled density. Inside the bottle we also added the folded paper impregnated with acaricide to provide a dry surface where larvae could pupate and for prevention of mites proliferation. Each line was maintained in 5 bottles each staggered in 7 days, so that we have whenever individuals of four stages of development. All fly handling was done at room temperature using CO_2 anesthesia.

So, isocromosomal lines were obtained for populations of Málaga, Valencia, Perelló and Barcelona. Moreover for the population of Barcelona, the genetic background was homogenized before obtaining the isocromosomal lines. For the rest of Iberian populations we did all steps of the scheme described above, excluding the repetitive crosses with *ch-cu*, after first cross with this strain the males of F_1 were directly crossed with the *VaBa* strain. As for the populations of Montpellier, Dijon and Groningen, only crosses of wild males with *ch-cu* strain were performed for recognition of the chromosomal arrangements, but homokaryotypic lines were not established for them.

3.3. CHARACTERIZATION OF THE *O* CHROMOSOME GENE ARRANGEMENTS

Diptera offer an outstanding opportunity to study chromosomal inversion polymorphisms due to the polythene nature of the chromosomes found in its salivary glands, where many multiple DNA replications have occurred without mitosis.

Salivary glands were isolated from the third instar larvae (the stage just before pupation) for each line to determine the wild *O* chromosome gene arrangement. Before dissection the larvae were washed in 0.8% NaCl solution. Salivary glands were fixed, stained with filtered aceto-orcein solution on microscope slides, squashed in 55% lactic acid according to a routine technique and preparations of polythene chromosomes were examined with the phase contrast microscope at 40× (oil immersion).

3.4. Candidate gene localization

3.4.1. DNA isolation

DNA was isolated from individuals of the *ch-cu* marker strain. The following protocol is a modification of that described in Pascual *et al.* (1997). Each fly was homogenized in a 1.5 ml microcentrifuge tube containing 160μl of 10 mM Tris, 60 mM NaCl, 5% (wt/vol) sucrose and 10 mM EDTA; pH 7.8. One hundred microliters of 1.25% SDS, 300 mM Tris, 5% sucrose and 10 mM EDTA; pH 9, were then added. The mixture was incubated at 65º for 30 min, after which 60μl of 3 M potassium acetate was added and the mixture was kept at –20º for 20 min. After centrifugation for 15 min in Eppendorf centrifuge, the supernatant was transferred to the new tube and one volume of 2-propanol was added to it, mixed carefully and left standing at room temperature for 5 min, which was followed by a 10 min of Eppendorf centrifugation. The pellet was washed with 70% ethanol. Residual ethanol was removed by drying the precipitate in a desiccator for 30 min, after which the DNA was resuspended in 50μl of sterile distilled water.

3.4.2. DNA amplifications for gene localizations

Thirty genes were chosen as candidates for thermal adaptation for current investigation previously identified as differentially expressed in different thermal regimes (Laayouni *et al.* 2007). In this work 30 pairs of specific oligonucleotides (Sigma-Aldrich Inc.) were used for amplification of gene sequences (Supplementary Table S1). Some of the probes were done in our laboratory following the conditions reported by Williams *et al.* (1990) for specific PCR, which were optimized for use with *D. subobscura* template DNA. All reaction volumes were 25μl, overlayered with 50 ml of light

mineral oil (Sigma Chemical Co., St. Louis). Each reaction consisted of 1× buffer (ECOGAN SRL, Barcelona, España), 1.6 mM $MgCl_2$, 0.2 mM of each dNTP (Boehringer Mannheim, Indianapolis), 10 pM of each primer, template DNA (~10–20 ng), and 0.8 units of Taq polymerase (ECOGAN SRL). Two specific primers were added to each single reaction. Amplification was achieved in PTC-100 of MJ Research Inc. (Watertown, MA) thermocycler programmed as follows: a preliminary 5-min denaturation at 94°C; 30-35 cycles of 30 sec at 94°C (denaturation), 1 min at 50-60°C (annealing), and 1 min at 72°C (extension); and a final extension at 72°C for 5 min followed by storage at 4°C. PCR product sizes were checked in 0.8 % agarose gels (Pronadisa, Madrid) with Tris HCl acetate/EDTA (TAE) buffer for 1.5-2 hours at 70 V, constant voltage. Reaction products were analyzed alongside the small molecular weight marker VI and the large molecular weight marker λBste II (Boehringer Mannheim). Ethidium bromide-stained gels (0.5μg/ml) were visualized on a UV transilluminator and digitalized with a Bio-Capt (version 12.5) image management system.

Probes were labeled following non-radioactive procedure. 300ng–1μg DNA was diluted with sterile water till total volume of 15μl and incubated in 94ºC for 10 min, laid on ice to avoid renaturalization and centrifuged for a moment. Then 2μl of DIG DNA Labeling Mix (10x concentrated dNTP labeling mixture: 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 0.65 mM dTTP, 0.35 mM DIG-dUTP, alkali-labile, pH 7.5), 2μl of hexanucleotides and 1μl of Klenow enzyme were added. The mix was incubated in 37ºC for 20 hours, after that DNA was precipitated with 0.1 of total volume (2μl) of sodium acetate and 2.5 of volume (50μl) of cold ethanol and stored in -20ºC for 4 hours. Then, DNA was centrifuged in 4ºC for 30 min, the supernatant was discarded, the pellet was washed with 100μl of 70% ethanol and centrifuged again in 4ºC for 5 min. Supernatant was discarded and pellet was dried by vacuum and dissolved in 15μl of sterile water. The total yield from the labeling reaction (500ng–2μg) was quantified according to the instructions supplied by the manufacturer. DNA concentration in the probes was checked by Dot Blot detection. The remaining probes were obtained by Gemma Calabria in the laboratory of Evolutionary Genetics of the University of Barcelona by directly including Dig-dUTP in the PCR amplification reaction.

3.4.3. POLYTHENE CHROMOSOME PREPARATION AND *IN SITU* HYBRIDIZATION

The karyotype of *D. subobscura* consists of five acrocentric chromosomes and a dot chromosome (Figure 1.11). Following MAINX, KOSKE and SMITAL (1953) the large chromosomes in this species

are traditionally named as *A* (= X, the sex chromosome), *J* (= chromosomal element D of Mueller/Sturtevant/Novitski and homologous to arm 3L in *Drosophila melanogaster*; see POWELL 1997, p. 307), *U* (= chromosomal element B and homologous to arm 2L), *E* (= chromosomal element C and homologous to arm 2R), and *O* (= chromosomal element E and homologous to arm 3R). The five major acrocentric chromosomes and the dot chromosome are divided into 100 sections (*A*: 1 – 16; *J*: 17 – 35; *U*: 36 – 53; *E*: 54 – 74; *O*: 75 – 99; *Dot*: 100), and each section into 3–5 subsections (A, B, ...) (KUNZE-MÜHL and MÜLLER 1958).

Third instar larvae were grown at low densities at 18°C in a modified version of David's killed-yeast culture medium (DAVID 1962). Slides with salivary gland chromosomes suitable for *in situ* hybridization were prepared according to LABRADOR, NAVEIRA and FONTDEVILA (1990). Prehybridization, hybridization, and detection were carried out as described by DE FRUTOS, KIMURA and PETERSON (1989, 1990). To eliminate basic proteins the polytene chromosome preparations were incubated in 2×SSC at 65ºC for 30 min, then they were dehydrated by subsequent immersion into 70% ethanol at 65ºC and 95% ethanol at 35ºC, for 10 min in each. Hybridization consisted in the following steps. The dried slides were immersed into 0.1M of NaOH for 90 sec and washed in three containers with fresh 2×SSC solutions for 1 min in each to denaturize the chromosomes. Then, the preparations were dehydrated by immersion of slides subsequently into 30%, 50%, 70% and 95% ethanol for five minutes in each at the ambient temperature and air-dried. The hybridization solution was prepared on ice, including for each labeled probe 15μl of DNA, 200 μl of formamide, 4μl of 10% SDS and 100μl of 2× SSC and filled with sterile water till 400μl of total volume. The probes were denaturized by maintaining the solution at 95ºC for 10 min and placing it on ice for 10 min. 20μl of hybridization solution were placed on a siliconized coverslip and covered with a polythene chromosome slide. Hybridization preparations were stored at 37°C overnight and washed and the coverslip removed in 2×SSC for 5 min at the same temperature, twice for 5 min in ambient temperature and finally immersed into 1×PBS for 5 min at ambient temperature two times. Posthybridization was accomplished in ambient temperature and consisted in incubation of the slides in two containers for 3 min in each with 2×SSC/3×Denhardt's solution (20×SSC: 175.3g of 3M NaCl and 88.2g of 0.3M Sodium Citrate in 1L of H_2O, pH 7; 50×Denhardt's: 1g of Ficoll, 1g of polivinilpirrolidone and 1g of BSA in 100 mL of H_2O). After that, 100ml of Buffer 1 (0.5% of blocking agent dissolved in sterile water at 50-70ºC) per 20 slides was added and preparations were incubated for 30 min more at room temperature. Then they were washed for 5 min twice with Buffer

2 (10×: 242g of 100mM Tris, 175.32g of 150 mM NaCl diluted in 2L of H_2O, pH 7.5). The antibody was diluted with Buffer 1 in proportion 1:5000 and applied over slides, covering them completely, and they were incubated for 60 min in a shaker. Then, the slides were washed with Buffer 1 and Buffer 3 (24.22g of 100mM Tris, 11.68g of 100mM NaCl and 20.32g of 50mM $MgCl_2$ in 1L of H_2O, pH 9.5), twice in 200 mL of each solution for 15 min. 60μl of staining solution (45μl of NBT and 35μl of X-phosphate diluted in 10mL of Buffer 3) was applied over each preparation and covered with coverslip. Slides were stored in horizontal position in obscure place for 10 min. Then, they were washed with sterile water and dried. Chromosomes were observed by phase contrast with a Zeiss Axioscope photomicroscope at 40x magnification, and digitalized with a CyberLink Power Director 5 image management system.

3.5. SEQUENCING OF CANDIDATE GENES

3.5.1. SAMPLE SETTINGS

Chromosomal arrangements. In southwestern European populations the most frequent chromosomal arrangements for the chromosome *O* are O_{ST}, $O_{\underline{3+4}}$, $O_{\underline{3+4}+7}$ and $O_{\underline{3+4+8}}$ (SOLÉ *et al.* 2002). The first two show a clear contrasting clinal pattern in the original Palaearctic populations, with O_{ST} increasing and $O_{\underline{3+4}}$ decreasing in frequency with increasing latitude (MENOZZI and KRIMBAS 1992; KRIMBAS 1992). Arrangement $O_{\underline{3+4+8}}$ is also interesting because in historical times it was mainly restricted to the Mediterranean region, and it was the most abundant arrangement in northern Africa (SOLÉ *et al.* 2002). However, in the last decades its distribution has changed dramatically and recent surveys have revealed frequencies as high as 22.6% in Groningen, Netherlands, where it was previously absent (BALANYÀ *et al.* 2004). These three arrangements include overlapping inversions, exclusively located in *Segment I* of the *O* chromosome, producing inversion loops in heterokaryotypes of different lengths. For this reason, six gene regions that mapped inside these three chromosomal arrangements were selected and sequenced in two *D. subobscura* populations to study gene flux between populations and arrangements.

Populations. We restricted our analysis to two populations separated by more than 800Km of linear distance: Barcelona with the highest chromosomal arrangement diversity (see Results and Discussion below, Table 4.1) and with a good representation of all three arrangements in the north

of the Iberian Peninsula and Málaga representing the southern distribution of the species in Europe. Thirty isochromosomal lines of Barcelona (ten for each of the three gene arrangements) and 16 isogenic lines of Málaga were used for this study. To complete the Málaga sample larvae from fourteen heterogenic lines of the F_2 generation crossed with *ch-cu* strain, for which the chromosomal arrangement of the wild chromosome was identified, were cloned to isolate the wild DNA strain for the eight gene regions. The number of isogenic and heterogenic lines per each population, gene arrangement and gene region size is summarized in Table 4.2. The isogenic lines were directly sequenced while 23 sequences were obtained by cloning.

Gene regions. To study the genetic content of populations and inversions for the three most frequent chromosomal arrangements of *Segment I* we focused on eight regions. Two genes (*Fmr1* and *larp)* were selected because they showed significant differences in the expression pattern when comparing the two most extreme temperatures in the work of LAAYOUNI *et al.* 2007 and due to their cytological location (Figure 4.6). We also sequenced a segment of their regulatory regions to be able to compare the genetic differentiation between arrangements at coding and regulatory regions for the same gene. Additionally, to further investigate genetic differentiation between populations four genomic regions were also sequenced because in previous studies they had shown high levels of genetic variability and differentiation when comparing two of the inversions from a northwestern Spanish population and they are homogeneously distributed along *Segment I* (MUNTÉ *et al.* 2005). The selected genes included *Acph-1* (Acid phosphatase-1), which is a well studied gene in *D. subobscura* (see NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999 a, b; NAVARRO-SABATÉ, AGUADÉ and SEGARRA 2003)*; Ast* (Aspartate Aminotransferase) corresponding to the region P22 in MUNTÉ *et al.* (2005)*;* and finally *trus* (toys are us) and an additional genomic region with unknown function (region CG5961 in *D. melanogaster*) both corresponding to the region P154 in the paper of MUNTÉ *et al.* (2005). All non-regulatory regions include coding and noncoding parts (with the exception of region CG5961, which only includes coding sequence; see Figure. 3.2). The cytological location of the five sequenced gene regions is marked in relation to the three analyzed arrangements (Figure 3.3).

To sequence gene *Acph-1* we used the primers published in NAVARRO-SABATÉ, AGUADÉ and SEGARRA (1999a); for *Ast*, *trus* and *CG5961* primers were the same used in the work of MUNTÉ *et al.* (2005). For the other two genes and their regulatory regions we designed the primers as follows.

The sequences of the genes *Fmr1* and *larp* of *D. pseudoobscura* and *D. melanogaster* were downloaded from Flybase database (www.flybase.org) and a preliminary BLAST was performed against a preliminary draft of the *D. subobscura* genome (Barcelona Subobscura Initiative). The genome of *D. subobscura* is not yet fully assembled, but fortunately enough the homologous sequences of the genes *Fmr1* and *larp* were identified in the genomic contigs. Finally, for each gene the sequences of the three species were aligned to locate the start codon of the coding region and primers in conserved regions were designed with the online program Primer3 ver. 0.4.0 (http://frodo.wi.mit.edu; ROZEN and SKALETSKY 2000). In the same way the primers for the 5 ' regulatory regions were obtained for these two genes. The primers used for amplification and sequencing of the genes, the amplified fragment sizes and cytological location of each gene are given in Table 4.3 and Supplementary Table S1.

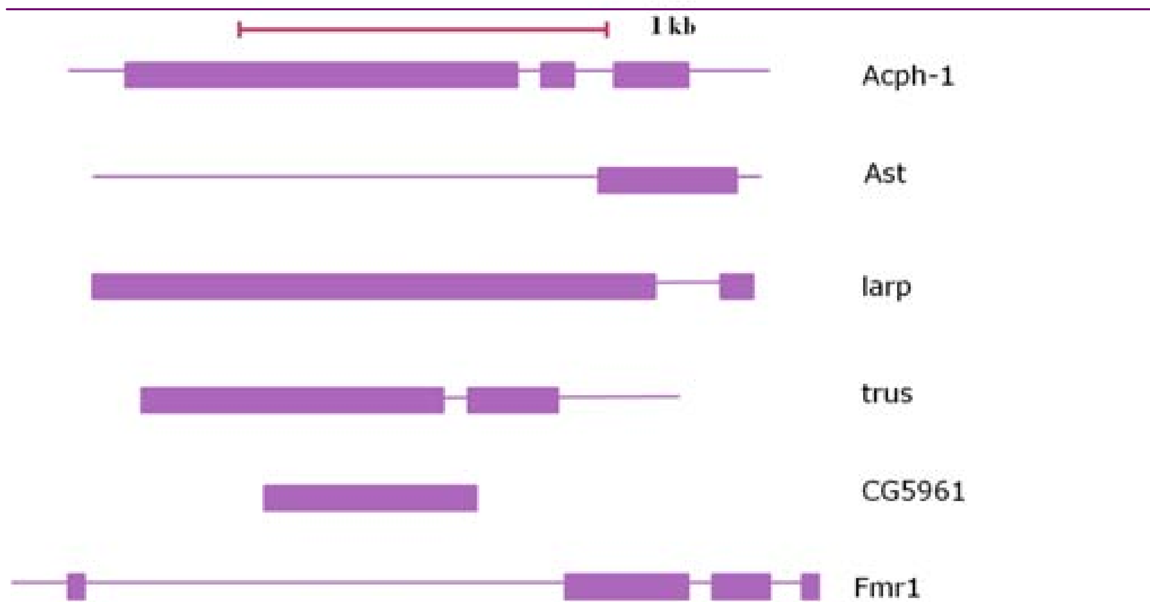


Figure 3.2: Scheme of the non-regulatory sequenced regions. Boxes and lines indicate coding and non-coding sequences, respectively.

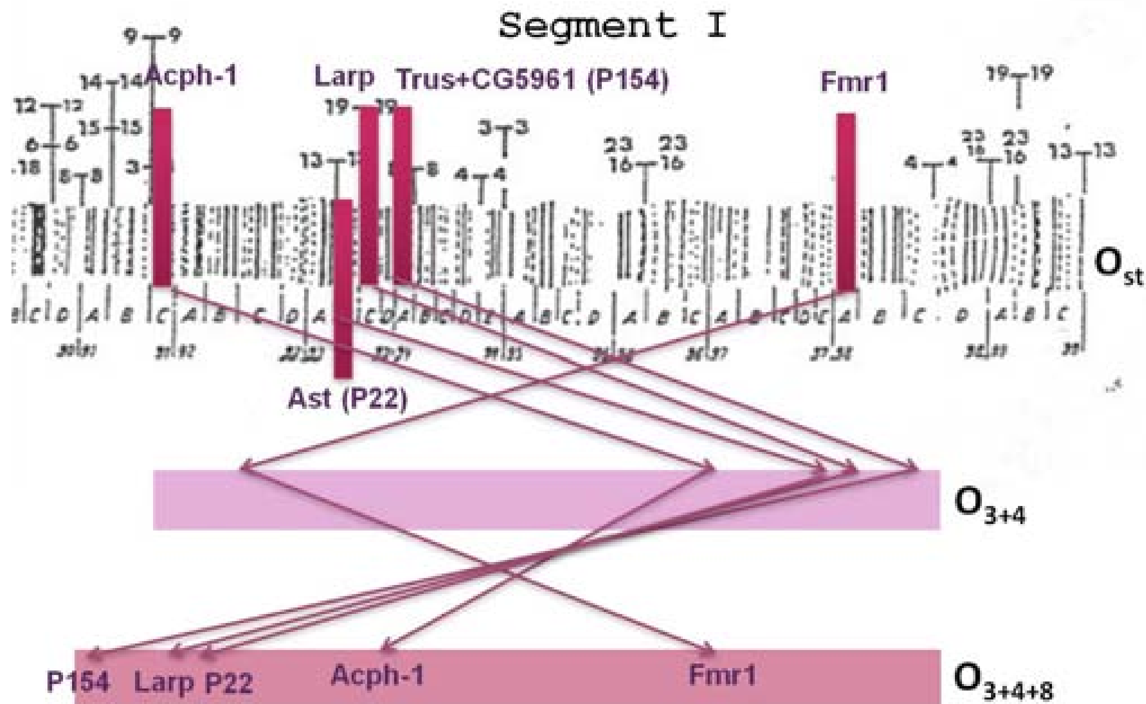


Figure 3.3: Cytological localization made by *in-situ* hybridization of six selected gene regions on *Segment I* of the O_{ST} chromosome in *D. subobscura.* Boxes and arrows indicate the relative position of each inversion and gene inside.

3.5.2. DNA SEQUENCING

The same protocol for DNA isolation indicated above was used for DNA extraction of individual flies for each line stored in 95% ethanol at -80°C. DNA was resuspended with 25µl of distilled autoclaved water. Approximately 10-20ng of DNA was added to a total volume of 20µl PCR mix. Each reaction tube contained 1×KAPA Taq buffer A (KAPABiosystems), 0.2mM of each dNTP (Boehringer Mannheim, Indianapolis), 10pM of each primer (synthesized by Sigma Aldrich®), and 2 units of KAPA Taq DNA polymerase (KAPABiosystems). Two specific primers were added to each single reaction. Amplification was run on a PTC-100 of MJ Research Inc. (Watertown, MA) thermocycler programmed as follows: a preliminary 5-min denaturation at 95°C; 35-40 cycles of 30 sec at 95°C (denaturation), 30 sec at 50-65°C (annealing), and 2 min 30 sec at 72°C (extension); and a final extension at 72°C for 2 min followed by storage at 4°C. PCR products were cleaned up with Qiagen PCR Extraction Kit following the manufacturer's instructions. Some PCR products were purified from the rest of reagents by enzymatic clean up with ExoSAP - IT® (Affimetrix): 8µl of

the enzymatic mix (0.025µl of Exo I (exonuclease I (20U/ml), which degrades the excess of primers, 0.250µl of SAP (alkaline phosphatase (1U/ml) that degrades excess of dNTPs) and 9.725µl of mQ H_2O) were added to 20µl of PCR. The following program was used for this step: 30 min at 37ºC and 5 min at 95ºC followed by storage at 4°C.

3.5.3. CLONING

Before sequencing, the PCR products of F_2 larvae of Málaga population, stored in 95% ethanol at -80°C, were cloned to isolate the wild *O* chromosome using the pGEM-T Easy ligation kit (Promega) following the manufacturer's instructions. The PCR products were ligated with pGEM-T 4 vector: 50-100ng of insert were mixed with 1µl of pGEM-T 4 vector DNA, 5µl of 5×T4 DNA ligase buffer and T4 DNA ligase (5U/µl) and filled up to 20µl with sterilized water. The reactions had been maintained at 4ºC overnight and then stored at -20ºC till transformation, which was performed using fresh DH5α *E. coli* competent cells.

Chemically competent cells (100µl) were mixed with 20µl of the ligation reaction. Reaction was incubated for 20 min on ice, heat shocked for 45 sec at 42°C and placed on ice for 2 min more. Nine hundred microliters of LB medium were added to the reaction, which was incubated for 1-1.5 h at 37°C. Cells were subsequently stricken onto LB agar plate containing ampicillin and Xgal. Plates were incubated overnight at 37°C.

Single white *E. coli* colonies were inoculated into 5ml of LB medium with ampicillin (20µg/ml) and incubated overnight with vigorous shaking at 37°C. The suspensions were centrifuged for 10 min at 4400 rpm and the supernatant was removed. The pellet was exposed to Miniprep procedure with Qiagen Miniprep Plasmid Kit following the manufacturer's protocol.

Sequencing reactions for cloned as well as for the isogenic lines were carried out in Macrogene Company of South Korea.

3.5.4. DATA ANALYSIS

Partial sequences were assembled with two programs: BioEdit 7.0.8.0 (HALL 1999) and Geneious ver. 5.4 (DRUMMOND *et al.* 2011). Complete sequences were multiply aligned with Clustal W

program (THOMPSON, HIGGINS and GIBSON 1994) included in BioEdit software and further edited with BioEdit 7.0.8.0 program (HALL 1999). When the sequences were aligned the exonic and intronic regions (Figure 3.2) for the gene regions *Acph-1*, *Ast*, *trus* and CG6159 were assigned following the work of MUNTÉ *et al.* (2005), while exons and introns of *larp* and *Fmr1* were determined by comparing the sequences of 12 genomes available on the UCSC page http://genome.ucsc.edu/, but especially of *D. pseudoobscura*, since it is the closest species with its complete genome sequenced. Once exons were annotated, the annotation was checked by blastx against UniProt protein database (http://www.uniprot.org/; THE UNIPROT CONSORTIUM 2012).

Analyses were performed for each region separately and for the single concatenated data set comprising those gene regions sequenced in the same lines, within the three most frequent chromosomal arrangements between the two populations and between arrangements. The DnaSP v5 program was used to perform most of the analyses and in all cases gaps were excluded from the analysis. Concatenation was performed with the program Concatenator ver.1 (PINA-MARTINS and PAULO 2008). Standard parameters of nucleotide polymorphism were estimated: the number of segregating sites in the sample (S), nucleotide diversity (π; NEI 1987), nucleotide diversity in synonymous sites and noncoding positions (π_{sil}; (NEI and GOJOBORI 1986) and heterozygosity per silent site (θ_{sil}; WATTERSON 1975). The nucleotide divergence per silent site (K_{sil}) was estimated in comparison to *D. pseudoobscura* according to NEI and GOJOBORI (1986). The differences of variability between arrangements and populations were examined using π_{sil} and the number of segregating positions (S) of each gene with Wilcoxon Matched Pairs Test incorporated in Statistica ver.9 (STATSOFT, INC 2009). The level of genetic differentiation between arrangements and between populations within arrangements was estimated as F_{ST}. The F_{ST} statistic is based on the average number of differences between sequences of the same population and the average number of differences between sequences from two different populations. Statistical significance of genetic differentiation was assessed with the p-value of the *S*nn statistics (-, not significant; *, $0.01<P<0.05$; **, $0.001<P<0.01$; ***, $P<0.001$; HUDSON, BOOS and KAPLAN 1992) using coalescent simulations made with 10000 replicates.

Physical distances between markers (Mb) were calculated under the assumption that all cytological bands of the *O* chromosome of *D. subobscura* contain the same DNA content as considered in MUNTÉ *et al.* (2005). This chromosome is equivalent in length to chromosome 2 of *D.*

pseudoobscura (30.8 Mb; PEGUEROLES *et al.* 2010 a, b) being the closest relative with its whole genome sequenced (CLARK *et al.* 2007). Assuming this length and the number of cytological bands from the KUNZE-MÜHL und MÜLLER map (1958) we have obtained a rough average estimation of the sequence length per band.

Neutrality tests (TAJIMA 1989; FU and LI 1993) were performed separately for the O_{ST}, $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ samples. To study possible deviations from neutrality Tajima's *D* (TAJIMA 1989) was calculated, which computes a standardized measure of the total number of segregating sites and the average number of mutations between sequence pairs. If the mutations presented in the sequences are neutral *D* is expected to be equal to zero. A negative Tajima's *D* signifies an excess of low frequency polymorphisms relative to the expectation due to either a selective sweep as a result of directional selection or a recent bottleneck with subsequent population expansion. A positive Tajima's *D* signifies indicates a decrease in population size and/or balancing selection.

Fu and Li's *D* test assumes that "old" mutations will tend to be found in the older part of the genealogy while "new" mutations will likely be found in the younger part of the genealogy. The older part of the genealogy consists mainly of *internal* branches, while the younger part mainly of *external* branches. A branch is said to be *external* if it directly connects to an external node, otherwise it is said to be *internal*. In the presence of purifying selection there will be an excess of mutations in the external branches because deleterious alleles are present in low frequencies. Also there is likely to be excess of mutations in the external branches if an advantageous allele has recently become fixed in the population, because then the majority of the mutations in the population are expected to be young. On the other hand, if balancing (overdominant) selection is operating at the locus, then some alleles may be old and so there may be deficiency of mutations in the external branches. Therefore, comparing the numbers of mutations in internal and external branches with their expectations under selective neutrality should be a powerful way to detect selection, which is the idea behind this test (FU and LI 1993). We have used *D. pseudoobscura* as outgroup, because without outgroup it is difficult to infer accurately the number of external branches.

Thus, *D* statistics indicates a possible deviation from neutralism in favor of selection. But both tests used here could be influenced by the evolutionary history of the sampled populations. To distinguish between demographical events and selection we analyzed whether the different

inversions in the two populations recently changed their effective size by computing the R_2 statistic which reflects the pairwise nucleotide difference distribution or mismatch distribution (RAMOS-ONSINS and ROZAS 2002), because it is more suitable for small sample sizes presenting recombination. The significance of this parameter was calculated from coalescent simulations with 1000 replicates considering the recombination rate *Rho,* estimated using the composite likelihood method of Hudson (HUDSON 2001) implemented in the RDP4 program (MARTIN *et al.* 2010; MCVEAN, AWADALLA and FEARNHEAD 2002), since the rejection of the null hypothesis of constant population size depends on the level of recombination implemented.

To detect the footprint of selection, we conducted a test of MCDONALD and KREITMAN (1991), based on the comparison between the ratio of synonymous and non-synonymous changes fixed between species and the relationship of synonymous and non-synonymous polymorphisms in the same species for coding positions. Under the assumption of neutrality, the ratio of synonymous changes between species should be the same as within species. The standard MKT for coding regions and extended for noncoding were used to detect selection (MCDONALD and KREITMAN 1991; EGEA, CASILLAS and BARBADILLA 2008; http://mkt.uab.cat).

To test noncoding regions, the coding portion of a gene considering only synonymous 4-fold degenerate sites was set up as putatively neutral and compared with noncoding part (introns and regulatory regions) of the same species to detect polymorphic sites and the orthologous sequence of *D. pseudoobscura* to account for divergence. The Neutrality Index (*NI*) was calculated as

$$NI = \frac{P_N/P_S}{D_N/D_S}$$

where P_n is a number of polymorphic non-neutral (non-synonymous) sites, P_s is a number of polymorphic neutral sites (synonymous), D_n - is the number of divergent non-neutral sites and D_s - is the number of divergent neutral sites. Then *α*, which is the proportion of adaptive substitutions (SMITH and EYRE-WALKER 2002) and ranges from -∞ to 1, was calculated as 1-*NI*,. If the ratio of fixed differences to polymorphisms is much higher for non-synonymous changes (i.e. $D_n/P_n \gg D_s/P_s$), resulting in positive value of *α*, this indicates that genetic changes have been subject to positive selection, which promote the fast fixation of advantageous variants. A negative *α* value indicates that there were fewer non-synonymous substitutions in evolution than expected given the

number of non-synonymous polymorphisms. That can be attributed to purifying selection preventing the fixation of harmful mutations (the number of divergent non-neutral changes D_n is lower than expected), but also an excess of non-neutral polymorphisms could be explained by balancing selection. If α is approximated to zero the null hypothesis of neutral equilibrium cannot be rejected.

Linkage disequilibrium (LD) between pairs of parsimony informative sites (and association between informative sites and chromosomal arrangement) was estimated according to Fisher's exact test and by the r^2 statistic (HILL and ROBERTSON 1968), and its statistical significance assessed by the Fisher test with Bonferroni correction for multiple comparisons (WEIR 1996). The overall level of LD was measured as ZnS (KELLY 1997) for parsimony informative sites (ZnS_i). Recombinant networks for each gene region and the concatenated data set were generated with Splitstree4 program (HUSON and BRYANT 2006). Gene conversion fragments were obtained using the method of BETRÀN *et al.* (1997) implemented in DnaSP v5 software (ROZAS *et al.* 2003).

The age of inversions was estimated for each gene region separately and averaging all of them as well as combining them in a concatenated dataset, using the mean silent nucleotide diversity of all individuals (ROZAS *et al.* 1999), except those identified as recombinants. The number of substitutions per site and year was estimated using the divergence per silent site between *D. subobscura* and *D. pseudoosbcura* assuming that the two species diverged 17.7 Myr ago, based on a large multilocus data set (TAMURA, SUBRAMANIAN and KUMAR 2004) and 8 Myr ago, based on only one gene (RAMOS-ONSINS *et al.* 1998) but suitable for comparison with previous studies. Finally, Sign tests were carried out using Statistica ver.9 for comparing the age of inversions inferred with the different gene regions.

3.6. MEASUREMENT OF THERMAL PREFERENCE AND THERMAL RESISTANCE

3.6.1. EXPERIMENTAL SETTINGS AND PROCEDURES

As was mentioned in a section above more than 200 isofemale lines were derived from *D. subobscura* wild flies collected near Barcelona (41°43'N, 2°13'E) in October 2007 and used to obtain isochromosomal lines for the *O* chromosome in an otherwise homogeneous genetic

background. Once obtained, the isochromosomal lines were genotyped for 13 microsatellite loci located on the *O* chromosome to check that no recombination events occurred during the different crosses. The 18 independent isochromosomal lines used in this study (see Experimental settings) were found to be homozygous for all the loci. The lines were kept at 18°C (12:12 light/dark cycle) in 130-mL bottles with low adult density (around 20 pairs/bottle) to standardize the rearing conditions before egg collections.

To obtain the experimental flies, all 54 crosses (inbred and outbred) were performed at 18°C by mating 4 days-old virgin males and females from the corresponding isochromosomal lines. After six days the males were discarded and the females (an equal number from each reciprocal cross in the outbred combinations) were transferred to egg-laying chambers containing fresh food and charcoal colouring. Eggs were placed in vials (45 eggs/vial containing 6 mL of food) at two rearing temperatures: 18°C and 22°C. Non-anaesthetized emerging flies were stored in bottles at low adult density and used to evaluate laboratory thermal preference (T_p) and knock out temperature (T_{ko}) for each cross (see below). All fly handling was done at room temperature using CO_2 anaesthesia only to sort virgin flies and to place females in the egg-laying chambers.

Six independent isochromosomal lines for each of the three arrangements (i.e., O_j^1,...,O_j^6; *j* = ST, 3+4, 3+4+8) were used in the present experiments. Extensive genetic differentiation of up to 4 Mb (i.e., about 15% of the euchromatic portion) has been detected among these arrangements (MUNTÉ *et al.* 2005). In other words, there are compelling reasons to think that the chromosome arrangements used in this work are genetically differentiated for *Hsp*70, and probably also for *Hsp68* since inversion effects can extend as far as 1000 kilobases outside from break points (ANDOLFATTO, DEPAULIS and NAVARRO 2001; LAAYOUNI *et al.* 2003).

Following SANTOS, IRIARTE and CESPEDES (2005) the experimental flies were obtained from 54 crosses, which are referred to as inbred (isogenic: $O_j^1 \times O_j^1$, $O_j^2 \times O_j^2$,..., $O_j^6 \times O_j^6$ with 18 crosses in total), or as outbred including both structural homokaryotypes ($O_j^1 \times O_j^2$, $O_j^2 \times O_j^3$,..., $O_j^6 \times O_j^1$ with 18 cyclically permuted reciprocal crosses in total) and heterokaryotypes ($O_j^1 \times O_k^1$, $O_j^2 \times O_k^2$,..., $O_j^6 \times O_k^6$; $j \neq k$; with 18 reciprocal crosses in total). Two developmental temperatures were used in the experiment to study potentially important effects of phenotypic plasticity: 18°C and 22°C. The

reason for this was the huge difference (about 7°C - 8°C) between our previous estimate of T_p (pooled average 16.6°C; REGO *et al.* 2010) in *D. subobscura* flies raised at 18°C, and that obtained by Huey and Pascual (23.7°C; HUEY and PASCUAL 2009) where flies were raised at 22°C. Even though the flies assayed came from different sources - south-western Europe in REGO *et al*. (2010), and North America in HUEY and PASCUAL (2009) -, which could account for the observed difference because thermal responses can vary between populations (YAMAMOTO 1994), it remains to be seen whether developmental plasticity can affect estimates of thermal preference and heat tolerance.

3.6.2. THERMAL PREFERENCE BEHAVIOUR IN A LABORATORY GRADIENT

Laboratory T_p was measured as previously described (REGO *et al.* 2010), adult flies (about 7 days old) were individually placed in separate lanes on an aluminium block (30 cm length × 31 cm width × 2.5 cm height) where a thermal gradient with temperatures ranging from 11°C to 29°C was generated by hot and cold plate at each end (see SAYEED and BENZER 1996). Temperatures along the gradient were measured on the aluminum block with thermocouples and were reproducible through the experiment. The range of temperatures achieved is encountered by active flies in the field (HUEY and PASCUAL 2009). A plexiglas cover with 30 separate lanes was placed on the block, creating suitable spaces for individual flies to freely move along the aluminum base (Figure 3.4). The relative humidity along the lanes was not measured, but condensation was not a problem in the experiment. The plexiglas plate was lightly dusted with quinine sulfate powder (a repellant for *Drosophila*; QUINN, HARRIS and BENZER 1974) to prevent flies from escaping the temperature gradient by resting on the walls or roof of the lane. Adult flies (15 females and 15 males) were gently aspirated from the vials, introduced into the lanes at room temperature (22°C–23°C) and given approximately 1 h to adjust. Afterwards the aluminum base was placed on the plates to generate the thermal gradient (~10 min) and each fly's position was recorded four times every 10 min: from 40 to 70 min (counting from the time when the thermal gradient was applied). To minimize circadian variations, four trials were run between 12:00 h and 18:00 h, which allowed assaying all flies (60 females and 60 males) from a batch of crosses on the same day. We used the median of the four measurements to estimate T_p of each fly. Measurements were performed in a room with a constant temperature (22°C - 23°C), and the flies were assayed under white light

illumination. The age of all flies tested was synchronized at ~10 days post-imaginal eclosion. This protocol renders a repeatable assessment of flies' thermal preferences (REGO *et al.* 2010). After the thermal preference assay, each fly was gently removed from the lane and individually placed in a vial with fresh food at 18°C for the subsequent assay of heat stress tolerance.

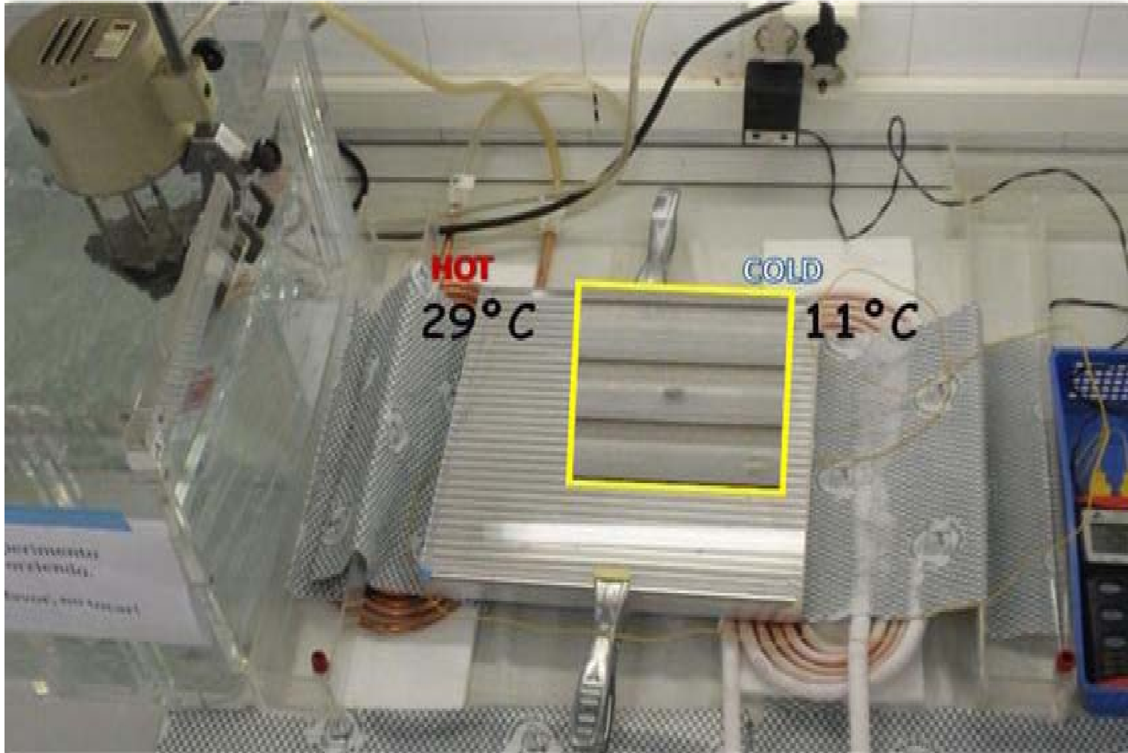


Figure 3.4: Alumium block covered with plexiglas lid connected to the thermocouples for thermal gradient generation. The cover contains 30 lanes with a fly inside of each.

3.6.3. HEAT RESISTANCE

One day after measurements of thermal preference flies were assayed for heat resistance also as previously described (REGO *et al.* 2010). Measurements were performed at ~14:00 h; that is, after 20–26 h from the time when their thermal preferences were recorded. Adults were individually placed in sealed empty vials and immersed in water-baths (60 flies per water-bath) at T_{min} = 24°C (Figure 3.5). Every 10 min individuals were scored for mobility (fly active or knocked out) and the temperature of the water was increased by ΔT = +1°C (it took ~2–3 min for the water bath to reach equilibrium). The procedure was repeated until the water-baths reached T_{max}, defined as the temperature when the last active fly was knocked out (T_{max} = 38°C was the upper limit in the

assays; median T_{max} = 33°C). For each fly T_{ko} was estimated as the temperature taken to knock it out (defined as the onset of muscle spasms; LUTTERSCHMIDT and HUTCHINSON 1997).

Figure 3.5: **Water bath with 60 vials for scoring thermal resistance.**

3.6.4. STATISTICAL METHODS

The experimental setup was devised to assay one male and one female from each cross and temperature per day (five blocks) for both T_p and T_{ko}, amounting to 1,080 flies in total. Some mishaps (e.g. individuals flew away or just died during the assays) were, however, unavoidable and the final data set contains a few more than or a few less than 10 flies in several crosses (the harmonic means of flies per cross and temperature were: T_p assay, 5.04 females and 4.80 males; T_{ko} assay, 4.89 females and 4.37 males). Statistical analysis with and without block design qualitatively yielded the same results. Therefore, to simplify matters blocks were not considered in the linear models below.

Consanguinity and temperature effects

Inbreeding and temperature effects were simultaneously analyzed by contrasting isogenic *vs*. outbred homokaryotypic flies reared at both developmental temperatures. The linear model used was:

$$T_{p(ijklmn)}=\mu+\kappa_i+C_{j(i)}+\tau_k+\iota_l+\varsigma_m+\kappa\tau_{ik}+\kappa\iota_{il}+\kappa\varsigma_{im}+\tau\iota_{kl}+\tau\varsigma_{km}+\ldots+\iota\varsigma_{lm}+\kappa\tau\iota_{ikl}+\kappa\tau\varsigma_{ikm}+\kappa\iota\varsigma_{ilm}+\tau\iota\varsigma_{klm}+\kappa\tau\iota\varsigma_{iklm}+\varepsilon_{ijklmn} \quad (1),$$

where μ is the overall grand mean, κ_i is the fixed effect of the karyotype (i =1, 2, 3), $C_{j(i)}$ is the random effect of the jth cross (j = 1, 2, ⋯, 6) within karyotype i, τ_k is the fixed effect of the developmental temperature (18°C or 22°C), ι_l is the fixed effect of inbreeding (isogenic or outbred homokaryotypic flies), ς_m is the fixed effect of sex, and ε_{ijklmn} is the residual error associated with the thermal preference (T_p) of the nth fly from the mth sex with the ith karyotype from the jth cross that was derived from the ιth group of crosses and assayed at the kth temperature. The covariate plate-hour was also introduced in the model to control for differences in circadian activity since several trials were conducted during each day. A similar linear model was used for knock out temperature, also introducing water-bath as a covariate since T_{ko} was assessed in different water-baths.

Notice that for the main effect "karyotype" the linear model (1) can be conveniently reduced to the following two-level nested ANOVA model:

$$T_{p(ijk)}=\mu+\kappa_i+C_{j(i)}+e_{ijk} \quad (2),$$

where the sum of squares for the error term e_{ijk} is simply the sum of the sum of squares for the remainder terms in (1). The usefulness of this model reduction is to efficiently perform randomization tests to test the null hypothesis about karyotype effects in a randomized (i.e., random assignment) experiment (EDGINGTON 1995). Permutation tests are far less sensitive to the presence of outliers than parametric tests. The null hypothesis of no karyotype effect was tested here after performing random permutations among replicate and selection temperature for the among selection temperature *F*-statistics. Each test used 10,000 random permutations.

Karyotype variation

To asses the effect of *O* chromosome karyotypes on T_p and T_{ko} we have focused on the outbred crosses, including both structural homo- and heterokaryotypes. The linear model used was similar

to (1) including the fixed effect of karyotype (κ_i; i = 1, 2, ⋯, 6), the random effect of cross within karyotypes ($C_{j(i)}$; j = 1, 2, ⋯, 6), the fixed effect of developmental temperature, and the fixed effect of sex. The covariate plate-hour was also introduced in the model. As above, a similar linear model was used for knock out temperature, also introducing water-bath as a covariate.

In the original Palaearctic populations chromosome arrangements $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ have a higher frequency at lower latitudes than arrangement O_{ST}, and the converse is true for higher latitudes (MENOZZI and KRIMBAS 1992). For this reason, the variation explained by the six karyotypes was further decomposed after pooling the first two arrangements into a single class ($O_{\underline{3+4}}{}^*$) as follows: between the two $O_{ST}/O_{\underline{3+4}}{}^*$ heterokaryotypes; among the three $O_{\underline{3+4}}{}^*/O_{\underline{3+4}}{}^*$ karyotypes; and among O_{ST}/O_{ST}, $O_{ST}/O_{\underline{3+4}}{}^*$, $O_{\underline{3+4}}{}^*/O_{\underline{3+4}}{}^*$. The karyotypic values for T_p and T_{ko} were also estimated in the additive-dominance scale (MATHER and JINKS 1977; MATHER and JINKS 1982) after pooling the two chromosome arrangements that share $O_{\underline{3+4}}$ (each comparison or contrast between two means has one degree of freedom).

The genetic correlation between T_p and T_{ko} can be approached as indicated in BETRÀN, SANTOS and RUIZ (1998). Assuming that the components of the between karyotypes sums of squares and cross-products (SSCP) hypothesis matrix ($\mathbf{H}_k$) are entirely genetic in origin, the correlation coefficient between the means of all six outbred karyotypes is given by:

$$r_k = \frac{H_k(1,2)}{\sqrt{H_k(1,1)\,H_k(2,2)}} \qquad (3),$$

where $\mathbf{H}_k$ (1, 2) is the off-diagonal element (sum of products of karyotype averages), and $\mathbf{H}_k$ (i, i) is a diagonal element (sum of squares of karyotypes averages) for the ith variable. This correlation coefficient is obviously an approximation to the genetic correlation because the $\mathbf{H}_k$ matrix also contains a fraction of the variation among the isogenic lines used to obtain the outbred flies (see section 3.6.1). The correlation coefficient can be tested as:

$$t = r\,\frac{k-2}{\sqrt{1-r^2}} \qquad (4),$$

where k is the number of karyotypes (SOKAL and ROHLF 1995). After pooling the arrangements that share arrangement $O_{\underline{3+4}}$ into a single class, we can now obtain the new hypothesis matrix $\mathbf{H}_p$. The correlation coefficient between the pooled averages can be estimated as:

$$r_k = \frac{H_k(1,2)}{\sqrt{H_k(1,1)\, H_k(2,2)}} \quad (5),$$

The square of this correlation can be interpreted as that fraction of the total variation among karyotypes that is explained by O_{ST}/O_{ST}, $O_{ST}/O_{\underline{3+4}}*$, $O_{\underline{3+4}}*/O_{\underline{3+4}}*$.

Computer software for statistical analysis

The computer programs used for statistical data analyses were MATLAB algebra program environment (ver. 7.0.4, MATHWORKS INC: 2007) together with the collection of tools supplied by the Statistics Toolbox. The statistical software packages STATISTICA ver. 9 (STATSOFT INC: 2009) and SPSS ver. 15 (SPSS INC: 2006) were also used.

PART 4

RESULTS AND DISCUSSION

The Red Queen shook her head.
"You may call it 'nonsense' if you like," she said, "but I've heard nonsense, compared with which that would be as sensible as a dictionary!"

Lewis Carroll,

Through the Looking-Glass

4.1. CHROMOSOMAL POLYMORPHISM OF *DROSOPHILA SUBOBSCURA* IN EUROPE

4.1.1 DISTRIBUTION OF GENE ARRANGEMENT FREQUENCIES

Table 4.1 and Figure 4.1 represent the results on the estimation of the frequency distribution of the O chromosome inversions. As in previous studies no significant differences between the polymorphism of the chromosome *O* obtained from wild males and males descended from wild females were observed (SOLÉ 2002; ARAÚZ 2009), the present study considers the global polymorphism. Considering all populations, a total of 14 different arrangements of more than 90 described for the chromosome *O* in the Palearctic region (KRIMBAS 1992) were found. There was no any arrangement undescribed before.

Table 4.1: Chromosomal polymorphism in seven European populations. The number of individuals (*N*) and the frequencies of each arrangement in percents per each population are given (our data, 2007-2009).

	Málaga	València	Perelló	Barcelona	Montpellier	Dijon	Groningen
N	101	93	122	201	148	149	152
O_{ST}	4.95	2.15	7.38	26.87	23.65	52.35	54.61
$O_{\underline{3+4}}$	31.68	32.26	43.44	32.34	34.46	24.16	21.71
$O_{\underline{3+4}+1}$	0.99	-	4.92	3.98	2.03	-	-
$O_{\underline{3+4}+2}$	2.97	2.15	1.64	1.99	-	-	-
$O_{\underline{3+4}+6}$	-	-	-	0.5	-	-	-
$O_{\underline{3+4}+7}$	52.48	46.24	22.13	4.98	1.35	0.67	-
$O_{\underline{3+4+8}}$	6.93	11.83	16.39	26.37	36.49	22.82	21.05
$O_{\underline{3+4}+12}$	-	-	-	-	-	-	1.97
$O_{\underline{3+4}+13}$	-	-	-	0.5	-	-	-
$O_{\underline{3+4}+17}$	-	-	0.82	-	-	-	-
$O_{\underline{3+4}+22}$	-	3.23	1.64	0.5	0.68	-	-
$O_{\underline{3+4+23}+2}$	-	1.08	0.82	1.49	1.35	-	-
O_6	-	-	-	-	-	-	0.66
O_7	-	1.08	0.82	0.5	-	-	-

The arrangements O_{ST}, $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ were found in all populations. The arrangement $O_{\underline{3+4}+7}$, which is a typical arrangement in Southern Europe, had very low frequency in the north of the Pyrenees and was absent in the population of Groningen. On the other side, the following

arrangements were only found in one of the populations: O_6 and $O_{\underline{3+4}+12}$ in the population of Groningen; $O_{\underline{3+4}+6}$ and $O_{\underline{3+4}+13}$ in the population of Barcelona; $O_{\underline{3+4}+17}$ in the population of Perelló.

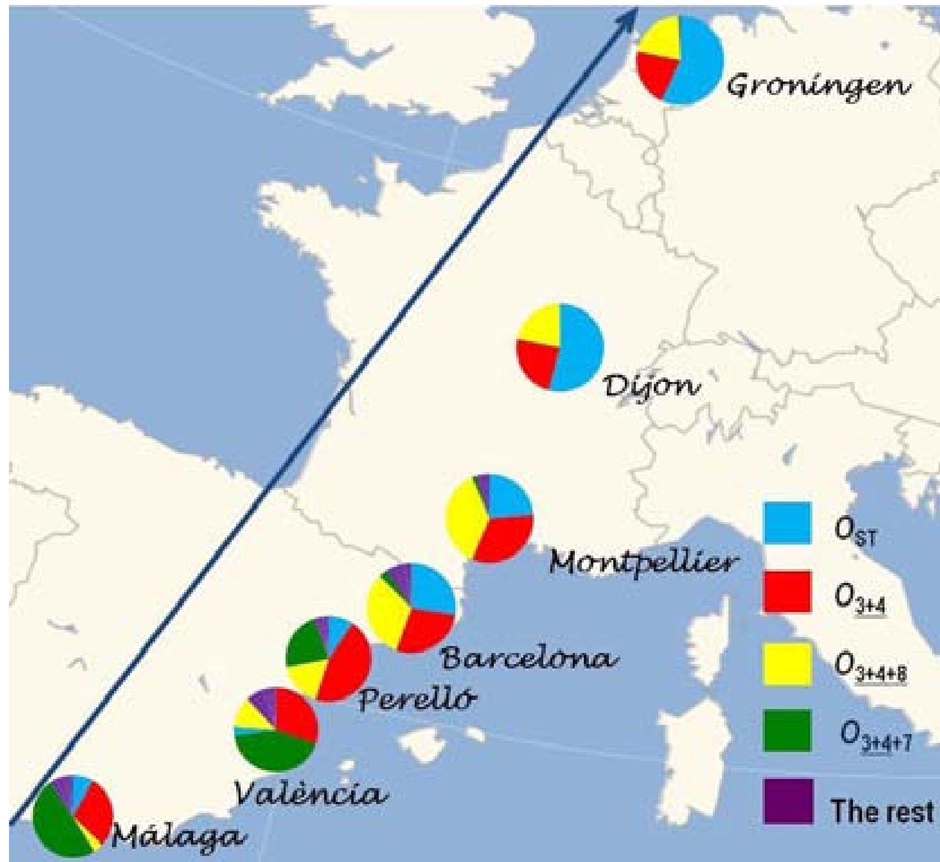


Figure 4.1: Distribution of inversion polymorphisms in Europe. The arrow indicates the temperature gradient.

Regarding to the chromosomal diversity, the population of Barcelona is more polymorphic because it has the highest heterozygosity inferred from inversion polymorphism, which is equal to 0.75, while the rest of the populations range between 0.61-0.73 and it present up to 11 different chromosomal arrangements, but it seems that in general, Mediterranean populations have chromosomal polymorphism higher than other populations, especially València, Perelló, Barcelona and Montpellier.

The frequencies of the most abundant gene arrangements in general correlate with the temperature gradient forming latitudinal clines as it could be seen on the Figure 4.1. O_{ST} arrangement is positively correlated with latitude and its frequency increases from the south to the north. At the same time the frequency of $O_{\underline{3+4}+7}$ shows a negative correlation with latitude and

reaches its maximum frequency in the south of Europe disappearing in the north. $O_{\underline{3+4}}$ shows almost the same frequency on the whole geographic area that it covers. At the same time, the frequency of $O_{\underline{3+4+8}}$ increases with latitude in the Iberian Peninsula until France, where it reaches 38% and then drops to 20-22% in the north of Europe, where it was absent in the initial samples (KRIMBAS and LOUKAS 1980).

A more detailed description of the clinal distribution of inversions in these populations and its calculation can be found in the CALABRIA (2012). To know whether the arrangements were distributed according to a latitudinal cline, the coefficient of regression between the transformed frequency of the arrangement in each population and latitude was calculated (O_{ST}: b = 0.893, P = 0.007; $O_{\underline{3+4}}$: b = -0.698, P = 0.081; $O_{\underline{3+4+8}}$: b = 0.487, P = 0.267; $O_{\underline{3+4}+7}$: b = -0,892, P = 0.017). These data support the visual observation of the clinal distribution of O_{ST} and $O_{\underline{3+4}+7}$ arrangements. In addition, comparing the new frequencies with the old, it is obvious that the frequency of O_{ST} decreases in Valencia and Montpellier, resulting in significant difference between old and recent samples (χ^2=24.30, P=0.0038 and χ^2=14.55, P=0.024, respectively), while it increases in Barcelona, where the difference between samples was also significant (χ^2 = 35.64, P < 0.0001). Instead, the $O_{\underline{3+4}}$ arrangement has a negative regression with latitude although not significant. SOLÉ *et al.* (2002) analyzed a total of 13 European populations, where this arrangement showed a regression coefficient b = -0.496 with P < 0.05. However, this arrangement presented no significant latitudinal cline in the colonized region (BALANYÀ *et al.* 2003). The $O_{\underline{3+4+8}}$ arrangement gave us a somewhat surprising observation. It was traditionally believed as of typical Mediterranean distribution, with frequency close to 100% in North Africa (KRIMBAS and LOUKAS 1980) and significant correlation with latitude (MENOZZI and KRIMBAS 1992). However, in recent studies the correlation is already not significant and also changes the sign of regression, as it was found in the Northern Europe with a frequency of up to 20% (SOLÉ *et al.* 2002). In our study regression with latitude remains positive (though not significant). This is mainly due to increase of its frequency in the sample of Montpellier in 2009, in comparison with the sample of the year 1999 (CALABRIA 2012).

Taking into account the work of MENOZZI and KRIMBAS (1992) the chromosomal arrangements can be divided into three groups: positively correlated with latitude, negatively correlated and uncorrelated significantly with latitude. According to this classification, we can consider that the

arrangement O_{ST} is typically cold-adapted (i.e., positively correlated with latitude), while the other three most abundant arrangements are considered to be warm-adapted. In the study of the changes in chromosomal polymorphism of *D. subobscura* over time, Solé *et al*. (2002) followed the same system and concluded that the arrangements typical for warmer latitudes have a tendency to increase in frequency while the arrangements typical for cold latitudes are decreasing, they related these changes to changes in temperature or other climatic factors probably related to the global warming.

Our results suggest that the O_{ST} arrangement has decreased in its frequency in Mediterranean localities while the frequency of $O_{\underline{3+4}+7}$ increased, which is consistent with what was previously described. The fact that our results did not show latitudinal clines for inversions $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ does not mean that they do not exist. Pooling the frequencies of $O_{\underline{3+4}}$ gene arrangement with that of $O_{\underline{3+4}+7}$ as it formed by non-overlapping inversions we resumed the results in the Figure 4.2 where the cline for $O_{\underline{3+4}}$ gene arrangement could be seen clearly with its abundance in the south and gradual decreasing of its frequency to the north (with a regression coefficient b = -0.93 and P = 0.003 (from Calabria 2012). Thus, it still may be the case that the inversion $O_{\underline{3+4}}$ is adaptive to warmer environments, but the $O_{\underline{3+4}+7}$ arrangement should provide some extra advantage, and that is why it is much more common in southern Europe.

Regarding to the arrangement $O_{\underline{3+4+8}}$, the absence of a cline does not mean that it is not adaptive. This arrangement had a great increase in recent years (Balanyà *et al.* 2004; Rezende *et al.* 2010), so it should give some advantage to its carriers. In the populations of Southern Europe it is unusual perhaps because of the competence with the more frequent $O_{\underline{3+4}+7}$, but from the Pyrenees, when the frequency of $O_{\underline{3+4}+7}$ decreases, this arrangement increased its frequency dramatically.

The chromosomal arrangements are adaptive, but their clines need not to be fixed in time, and may behave as dynamic structures over time. It has been assumed that the engine of adaptation was the temperature, but perhaps it was not the most important factor influencing the formation of clines as it was already noted in the work of Santos *et al.* (2004; 2005) and Castañeda *et al.* (2013), but many other factors may be decisive for distribution such as humidity, density and photoperiod, among others.

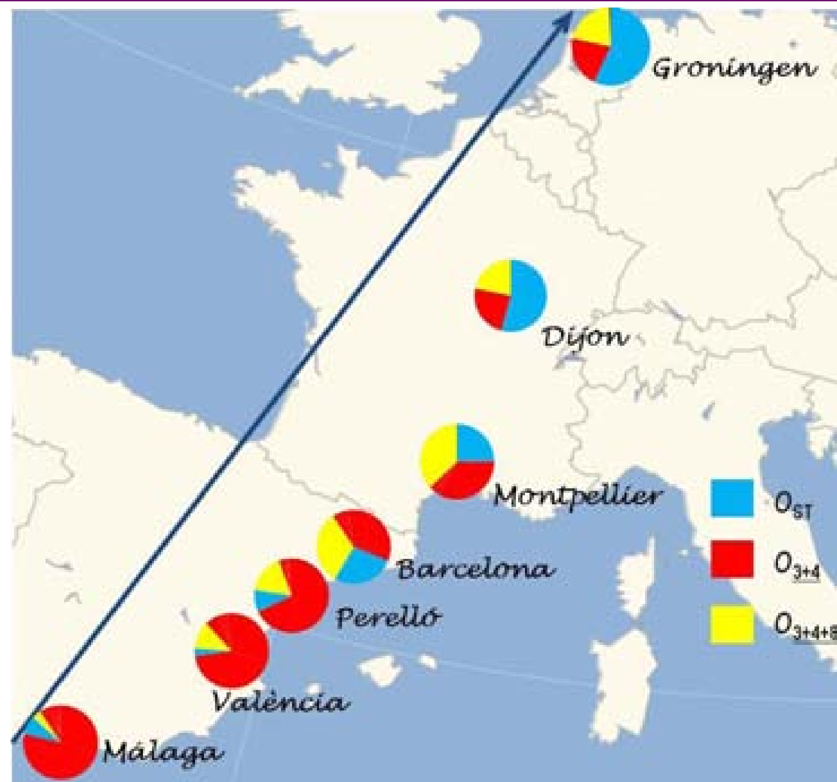


Figure 4.2: **Clinal distribution of the most abundant inversions.** The arrow indicates the temperature gradient.

4.1.2. TEMPORAL CHANGES IN CHROMOSOMAL POLYMORPHISMS IN EUROPEAN POPULATIONS

Here we could consider changes in inversion frequencies in the five out of seven populations of Barcelona, Málaga, València, Monpellier and Groningen for whom the previous data were available. By comparing the polymorphism between the two dates studied, significant differences in the composition for the population of València, Barcelona and Montpellier were found. The populations of Perelló and Dijon had not been sampled before, and therefore the comparisons could not be made for them so only the brief description of inversion proportions will be given.

Barcelona (Spain). In the population of Barcelona a total of 11 arrangements for the chromosome *O* have been found (Table 4.1) out of approximately 92 (produced from 66 inversions) recorded in the Palearctic region (KRIMBAS 1992, 1993; MENOZZI and KRIMBAS 1992). This population, which is considered central and hosts high genetic variability, has been well studied in the past, with regular revisiting of its gene arrangement frequencies (LOUKAS, KRIMBAS and VERGINI 1979; DE FRUTOS and

Prevosti 1984; Orengo and Prevosti 1996; Araúz *et al.* 2008). The most frequent gene arrangement for this region at the time of our study was O_{3+4} (32.34%). Almost at equal frequencies we have detected arrangements O_{3+4+8} and O_{ST} (26.37% and 26.87%, respectively). The remaining *O* chromosomal gene arrangements were recorded at very low frequencies. These results could seem somewhat unexpected in comparison with those from Araúz *et al.* (2008) for the year 2004 showed more typical composition of chromosomal polymorphism for the chromosome *O* in this population where arrangement O_{3+4} also had the highest frequency (31.62%), and arrangement O_{3+4+7} was quite frequent as well (17.09%), but arrangement O_{ST} was relatively scarce (11.11%) in comparison to the present collection. The graph on the Figure 4.3 illustrates the frequency dynamics for the four most abundant inversions on the *O* chromosome during 56 years from 1961 till 2007 in Barcelona. It can be seen that arrangement O_{3+4} is the most stable configuration in Barcelona, whose frequency has been maintained between 23.06 -32.34% during the last 35 years. The O_{3+4+7} gene arrangement increased in frequency during the period 1975-1982 from 18.2% to 35.5% (Loukas, Krimbas and Vergini 1979; de Frutos and Prevosti 1984). It was the most frequent arrangement until 1988 when its frequency peaked at 38.53% (Orengo and Prevosti 1996).

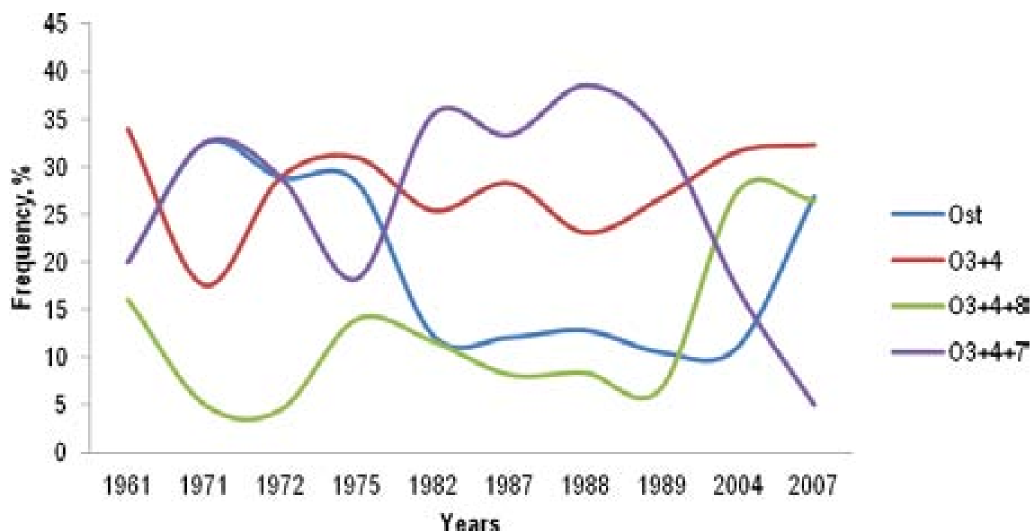


Figure 4.3: Temporal variation in gene arrangement frequencies in population of Barcelona. The graph is based on the data published in Prevosti 1964, 1966; Loukas, Krimbas and Vergini 1979; de Frutos and Prevosti 1984; Orengo and Prevosti 1996; Araúz *et al*. 2008 and our observations.

Thereafter it started to decline dramatically, and has almost disappeared nowadays. The frequencies of the O_{ST} and O_{3+4+8} gene arrangements show an opposite historical pattern to that of

$O_{\underline{3+4}+7}$. O_{ST} frequency rose steadily from 1961 to 1971 when its frequency peaked at 32.5% (Prevosti 1964, 1966; Loukas, Krimbas and Vergini 1979). It continued to remain quite high at the beginning of 70s until 1975, although its frequency was decreasing slightly during this period. From 1975 to 1982 it dropped to 12.3% and continued to decrease slightly until last observation in 1994 (Orengo and Prevosti 1996, Araúz *et al.* 2008, prior to our study). The percentage of $O_{\underline{3+4+8}}$ demonstrates a similar pattern from 1975 till the present. In 1975 there was a gradual decline from 13.9% to 6.74% in 1989, but then the frequency went up and in short period from 1989 to 1994 it reached 27.35%. In the last 13 years the frequency of $O_{\underline{3+4+8}}$ has been still steadily increasing. Contrary to what we found in València, the differences between two last observations (in 2004 and 2007) could not be explained by long-term changes in frequency, related with the increase in global temperature as the time between the samples were very short, however it might be related with short-term changes associated with climatology. The Barcelona sample was quite difficult to collect due to the weather conditions, since the fall of 2007 was very cold and to have the population sampled it was necessary to collect it during three days. Comparing average temperatures in October (data from the Fabra Observatory, situated in 1.5km from capture site) of two years differ by 2.5 degrees and it was cooler in 2007. These adverse climatic conditions could be related with the higher frequencies of the arrangement O_{ST}.

Málaga (Spain). The populations from Málaga, València and Perelló were collected in October 2008. We detected six gene arrangements in the population of Málaga after examination of 101 lines (Table 4.1). The arrangement $O_{\underline{3+4}+7}$ was the most frequent in this sample reaching 52.48%, which was the highest value for this gene arrangement in Spain at the time of our study. The $O_{\underline{3+4}}$ gene arrangement also had a high frequency there, attaining 31.68%. The frequencies for the arrangements $O_{\underline{3+4+8}}$ and O_{ST} were correspondingly very low, 6.93% and 4.95% respectively. This is in strong contrast to their frequencies in Barcelona in 2007, where they were the highest. Two other gene arrangements including 3+4 inversions were found in this population at very low frequencies: 0.99% for $O_{\underline{3+4}+1}$ and 2.97% for $O_{\underline{3+4}+2}$.

Figure 4.4 allow us to compare our present results with scarce long-term studies in the past and show the historical changes in inversion polymorphisms for this region. As can be seen the current frequency of $O_{\underline{3+4}+7}$ despite being the most frequent is not on its historical maximum, since at the very beginning of monitoring in 1963 a frequency of 79.9% was registered in Málaga (Prevosti

1966). So, in spite of a considerable decline until the next investigation in 1998 when it dropped to 38%, $O_{\underline{3+4}+7}$ has always been the most frequent arrangement in this region. The frequency of the $O_{\underline{3+4}}$ gene arrangement increased from 6.7% in 1963 to 34% in 1998 (SOLÉ *et al.* 2002) and almost did not change till 2008. O_{ST} leveled out over all three periods of observation, maintaining its frequency around 5-8%. The percentage of $O_{\underline{3+4+8}}$ was slightly increasing over time.

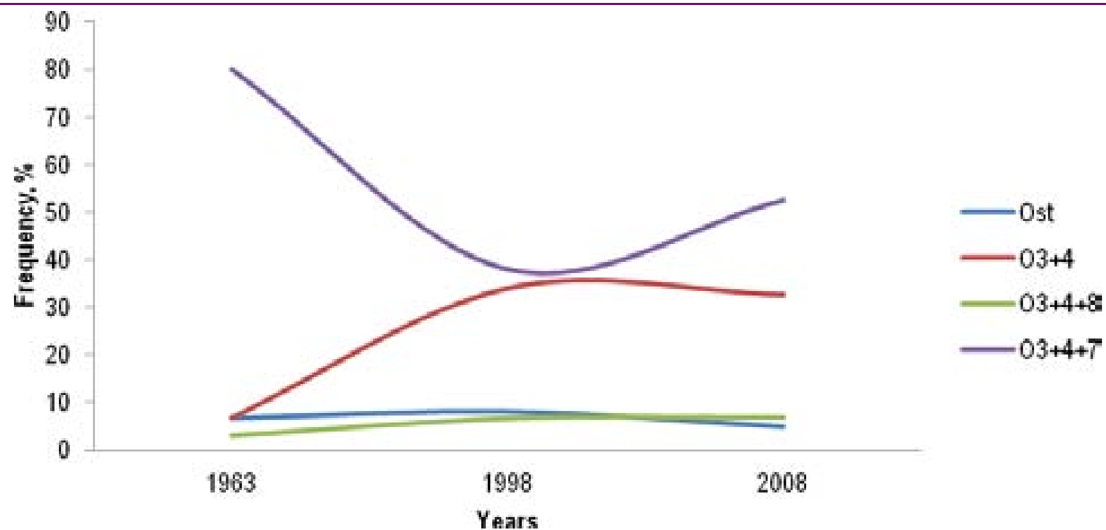


Figure 4.4: Temporal variation in gene arrangement frequencies in population of Málaga. The graph is based on the data published in PREVOSTI 1966, SOLÉ *et al.* 2002 and our observations.

València (Spain). We found eight gene arrangements in 93 lines in València (Table 4.1). The arrangement $O_{\underline{3+4}+7}$ was the most frequent in this population as in Málaga. The arrangement $O_{\underline{3+4}}$ was quite frequent also, reaching almost equal value in both populations. The frequency of $O_{\underline{3+4+8}}$ gene arrangement was significantly higher than in Màlaga, attaining 11.83%. The frequencies of the rest of gene arrangements attained very low values: 1.08% (O_7 and $O_{\underline{3+4+23}+2}$), 2.15% (O_{ST} and $O_{\underline{3+4}+2}$) and 3.23% for $O_{\underline{3+4}+22}$. As for the time-series observations, the arrangements containing 3+4 inversions ($O_{\underline{3+4}}$, $O_{\underline{3+4}+7}$, $O_{\underline{3+4+8}}$) increased substantially in their frequencies while O_{ST} dropped dramatically from 14-15% in 60s-90s to 2% nowadays (see Figure 4.5).

The differences in this population between the two more recent observations (1998 and 2008; χ^2 = 24.30, P = 0.0038; f. d. = 9) are due to a paucity of the arrangement O_{ST} in the population of 2008 at the expense of the arrangements $O_{\underline{3+4}}$ and $O_{\underline{3+4}+7}$. These differences could be due to the processes of microdifferentiation described in *D. subobscura* (KRIMBAS 1993), since the two sampling areas are physically remote. As it was mentioned in Materials and Methods, the area

where it was captured in 1998 was destroyed by rains and we made the sampling in an area of pine forest located about 9 km from Riba-roja. Another explanation could be to the fact that the captures were made in different seasons, both old populations in spring and the new population in autumn of 2008. Seasonal changes in chromosomal polymorphism have been observed and studied for this species (FONTDEVILA *et al.* 1983; RODRÍGUEZ-TRELLES, ALVAREZ and ZAPATA 1996; ZIVANOVIC 2007) and there have been changes in the frequency of some chromosomal arrangements. FONTDEVILA *et al.* (1983) studied the population of El Pedroso (Galicia) and seasonal changes were detected for arrangements O_{ST} and O_{3+4+7}, the last peaked in summer and its frequency decreased in spring and autumn, while the former showed opposite behavior. In addition, for the same population RODRÍGUEZ-TRELLES, ALVAREZ and ZAPATA (1996) found a strong association between seasonal changes and climatic factors (temperature, humidity, etc.).

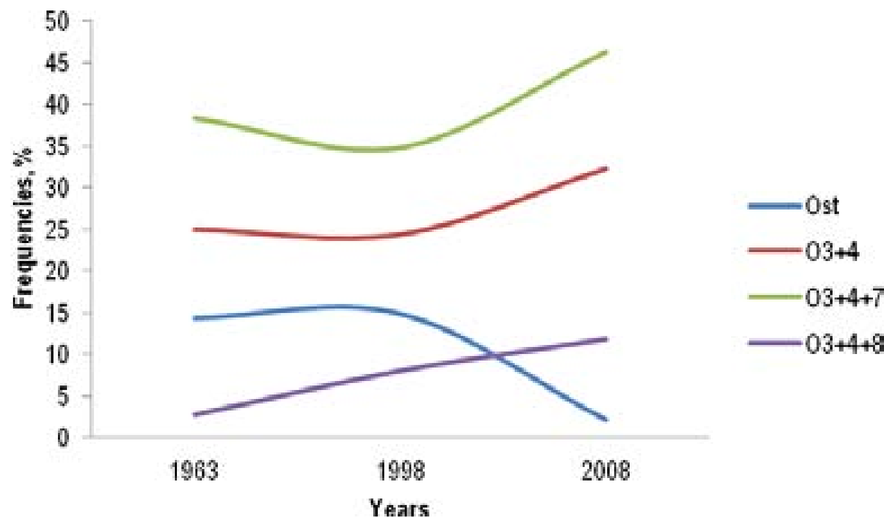


Figure 4.5: Temporal variation in gene arrangement frequencies in population of València. The graph is based on the data published in PREVOSTI 1966, SOLÉ *et al.* 2002 and our observations.

Perelló (Spain). The number of gene arrangements found was 10 in the population of Perelló (Table 4.1). The pattern of inversion frequencies in this area differed from those of three other populations: the most frequent arrangement in Perelló was O_{3+4} (43.44%), while O_{3+4+7} (22.13%) and O_{3+4+8} (16.39%) were considerably less frequent. The frequency of O_{ST} was slightly higher in this population than for the populations from Màlaga and València getting 7.38%. The rest of the arrangements were registered at very low frequencies between 0.82 and 4.92%. There were no time-series data for this location.

Montpellier (France). The populations of France and the Netherlands were sampled in August, 2009. Seven gene arrangements within the 148 examined lines were detected in the population of Montpellier. As it is shown in Table 4.1 the frequency of $O_{\underline{3+4}+7}$ decreases with latitude and it was almost four times lower than in Barcelona, but increased considerably over time getting 1.35% in 2009 vs. 0.9% thirty years ago. The gene arrangements $O_{\underline{3+4+8}}$, $O_{\underline{3+4}}$ and O_{ST} were detected at relatively high frequencies, 36.49%, 34.46% and 23.65% respectively. The frequencies of the rest of the arrangements were scarce. Significant differences between two more recent samples were found in this population (χ^2 = 14:55, P = 0.024; f. d. = 6). They were due to an excess of the arrangement $O_{\underline{3+4+8}}$ in 2009, which was the most frequent in the last sample, and the drop of frequency of the O_{ST}, which was the most frequent in the capture of 1998. The graph on Figure 4.6 shows a clear decrease of the frequency of O_{ST} arrangement over time. Its frequency has dropped from 61.1% in 1972 (PREVOSTI *et al.* 1984) to 23.65% in the present study. On the contrary, the frequencies of gene arrangements $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ went up in this period of time rising from 15.1% to 34.46% in the case of $O_{\underline{3+4}}$ and from 20.9 to 36.49% for $O_{\underline{3+4+8}}$. The explanation in this case could be the long-term change in polymorphism, which was observed in this species (BALANYÀ *et al.* 2004; 2006) where there were a lot of changes in the frequency of chromosomal inversions as it could be predicted by response to climate change at chromosomal level. Thus, the "warm" arrangement $O_{\underline{3+4+8}}$ is increasing while the "cold" O_{ST} is decreasing in its frequency in all European populations (SOLÉ *et al.* 2002; BALANYÀ *et al.* 2004). It cannot be discarded that the differences could be explained by the seasonality of polymorphism as the samplings were made in different months. The new collection was taken on the 1st of September and the former was carried out in late October, this may explain the higher frequency of the warm arrangement in the sample taken two months earlier.

Dijon (France). Only four gene arrangements were detected in Dijon in the present study. Table 4.1 indicates that the O_{ST} chromosomal arrangement was the most abundant (52.35%), more than doubling the frequency of each of the remaining gene arrangements: namely, $O_{\underline{3+4}}$ at 24.16% and $O_{\underline{3+4+8}}$ at 22.82%. The $O_{\underline{3+4}+7}$ arrangement was not present in this locality. There is no time-series data available for this population.

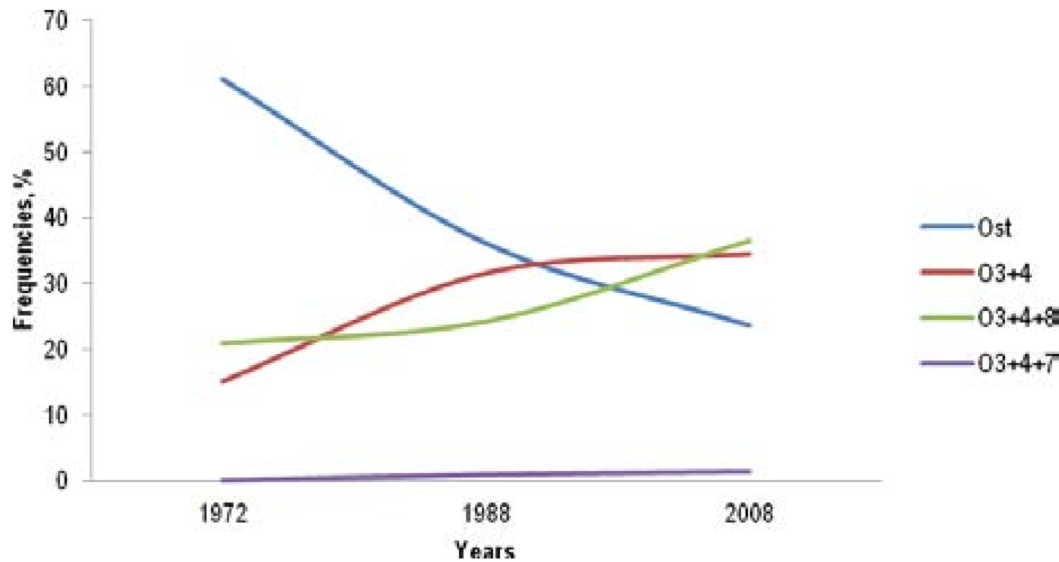


Figure 4.6: Temporal variation in gene arrangement frequencies in population of Montpellier. The graph is based on the data published in PREVOSTI *et al.* 1984, SOLÉ *et al.* 2002 and our observations.

Groningen (The Netherlands). The population of Groningen had a similar pattern of chromosomal inversion polymorphism as Dijon with O_{ST} gene arrangement getting the highest frequency, 54.61%, and $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements achieving similar frequency estimations equal to 21.71 and 21.05% correspondingly (Table 4.1). The graph on Figure 4.7 present similar temporal trend as it had been seen in the population of Montpellier with the frequencies of $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ increasing while O_{ST} decreasing over time, although the differences between the two more recent collections were not significant.

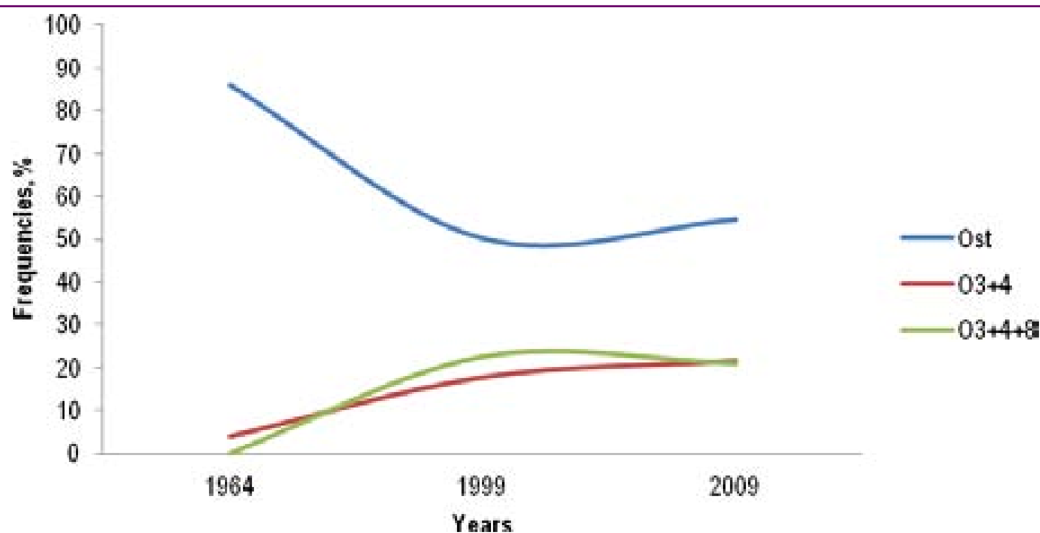


Figure 4.7: Temporal variation in gene arrangement frequencies in population of Groningen. The graph is based on the data published in KRIMBAS 1964, BALANYÀ *et al.* 2004 and our observations.

4.2. Physical localization of candidate genes

We selected thirty genes that were identified in a previous work (Laayouni *et al.* 2007) as candidate genes for thermal adaptation in *Drosophila subobscura* and localized them by *in situ* hybridization to study their locations relative to inversions. A physical mapping of 14 of those candidate genes, which responded differentially to thermal adaptation in the laboratory, was already conducted in a previous study, all localizing in the O chromosome of *D. subobscura* (Laayouni *et al.* 2007). We started our research by checking their cytological localization as well as that of 16 additional candidate gene probes whose orthologous sequences mapped in the same chromosomal element in *D. melanogaster*. The hybridization signals for nine out of the 14 localizations were inconsistent with the previously reported localization, probably due to a previous labeling mismatch. Nine of the 30 candidate genes localized in *Segment I*, and 21 in *Segment II*. Their exact positions on the chromosome *O* are shown on Figure 4.8 and Supplementary Table S1.

Besides the characterization of inversion content, the genes localized in the present work as well as in previous studies (Aguadé 1988a; Arboleda 2008; Cuenca *et al.* 1998; Ibnsouda *et al.* 1993; Laayouni *et al.* 2007; Mestres *et al.* 2004; Moltó *et al.* 1992; Munté *et al.* 2005; Rozas and Aguadé 1993; Sánchez-Gracia and Rozas 2011; Segarra, Ribot and Aguadé 1996) altogether with microsatellite loci (Santos *et al.* 2010) will serve as orientation markers for the assembling of the *Drosophila subobscura* genome currently ongoing by Barcelona Subobscura Initiative. In total 66 markers have been localized by *in situ* hybridization on the *O* chromosome as the result of different studies including this one. The information on the genes such as their full names, cytological localization and genomic position of the orthologous sequences in the genomes of *D. melanogaster* and *D. pseudoobscura* are included in Supplementary table S1.

The physical localization of genomic regions and their comparison to other related species can give insight on the speciation process, on the genomic reorganization during the evolution of a lineage, or the gene conservation through time. *Drosophila* is a very interesting genus for studies addressing genome evolution because there are 12 different species that have been already fully sequenced (*Drosophila* 12 Genomes Consortium 2007). The ancestral karyotype within the *Drosophila* genus consists of one dot and five acrocentric chromosome pairs, the same karyotype presented by *D. subobscura*. It is well known that the gene content of these six different elements is highly conserved, although there is extensive gene reshuffling within elements, referred to as

Muller's elements (MÜLLER 1940; *DROSOPHILA* 12 GENOMES CONSORTIUM 2007). Unfortunately, there is still a great gap of information on most *Drosophila* species, but comparative studies with available genomic sequences can be a very useful evolutionary tool for screening homologies and evaluating chromosomal synteny.

Orthologous sequences of 52 gene regions were searched in the genome of *D. melanogaster* (ADAMS *et al.* 2000) and *D. pseudoobscura* (RICHARDS *et al.* 2005) using Flybase (http://www.flybase.org). To reconstruct the synteny map, the approximate physical localizations of gene regions for *D. pseudoobscura* were inferred from their genomic locations available in Flybase server assuming that all cytological bands contain the same DNA content (SCHÄFER *et al.* 2010) and the length of chromosome *2* is equal to 30.8 Mb (RICHARDS *et al.* 2005). Assuming this length and the number of cytological bands from the TAN´s map (1935, 1937) updated by KASTRITSIS and CRUMPACKER (1966) and SCHAEFFER *et al.* (2008), we have obtained a rough average estimation of the sequence length per band. Several regions whose physical localizations were detected previously and published by SCHAEFFER *et al.* (2008) were used as landmarks (Supplementary table S1).

When comparing the physical positions of the 52 gene regions mapping in the *O* chromosome of *D. subobscura* with the homologous chromosomes of the other two species: *3R* of *D. melanogaste*r and *2* of *D. pseudoobscura* (Figure 4.9) we found a poor synteny in the order of markers but high conservation among chromosomal elements. The same pattern was found in the study of SANTOS *et al.* 2010 based on fluorescent *in situ* hybridization (FISH) of *D. subobscura* microsatellite loci in comparison to their genomic locations in these other two species. Overall, the comparisons of the locations of the 66 *D. subobscura* genomic markers from the *O* chromosome in the genomes of *D. pseudoobscura* and *D. melanogaster* showed high conservation on the same chromosome element, as previously observed with other gene markers and chromosomal elements (SEGARRA and AGUADÉ 1992; SEGARRA, RIBOT, AGUADÉ 1996; PAPACEIT, AGUADÉ and SEGARRA 2006). However, great internal shuffling by paracentric inversions had to occur inside a chromosome during *Drosophila* species divergence to explain the great lack of conservation in markers position within (GONZÁLEZ, RANZ and RUIZ 2002; PAPACEIT, AGUADÉ and SEGARRA 2006; *DROSOPHILA* 12 GENOMES CONSORTIUM 2007; BHUTKAR *et al.* 2008). Though highly reorganized, the wide conservation of homologous chromosomal content between species corroborates the existence of

low inter-arm translocations during the evolution of the *Drosophila* genus. In conclusion, despite the low number of chromosomal translocations between *Drosophila* species, the high number of chromosomal inversions among *Drosophila* lineages precludes using even closely related species such as *D. pseudoobscura* as reference to join contigs to properly assemble the *D. subobscur*a genome. Thus caution is needed when using other *Drosophila* species as a reference in genome projects.

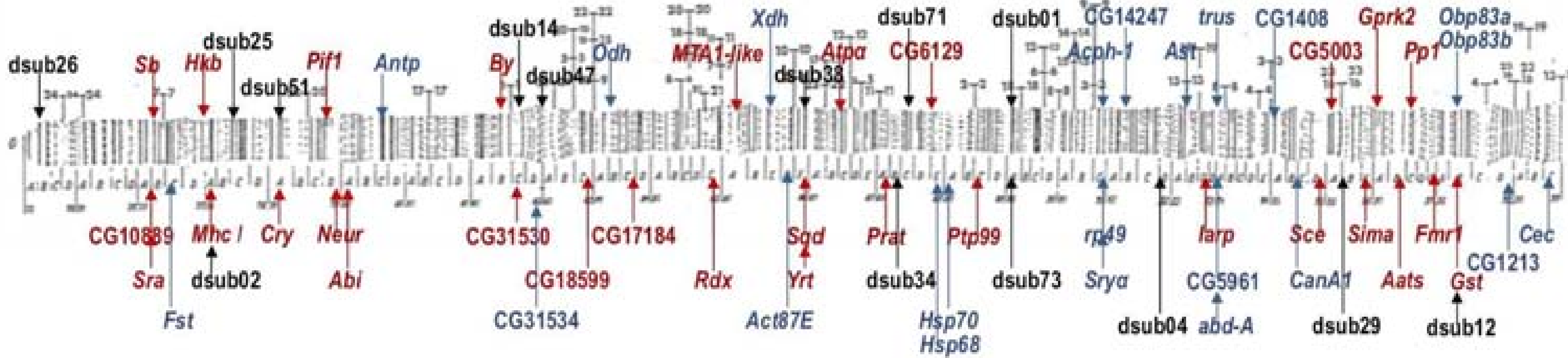


Figure 4.8: Diagram of *O* chromosome with physical localizations of candidate and non-candidate genes and microsatellites made by *in situ* hybridization. Red arrows and names correspond to genes localized in this study; blue arrows and names correspond to the genes with localizations published earlier in different articles (see text for references), black names and arrows correspond to microsatellites localizations published in SANTOS *et al*. (2010).

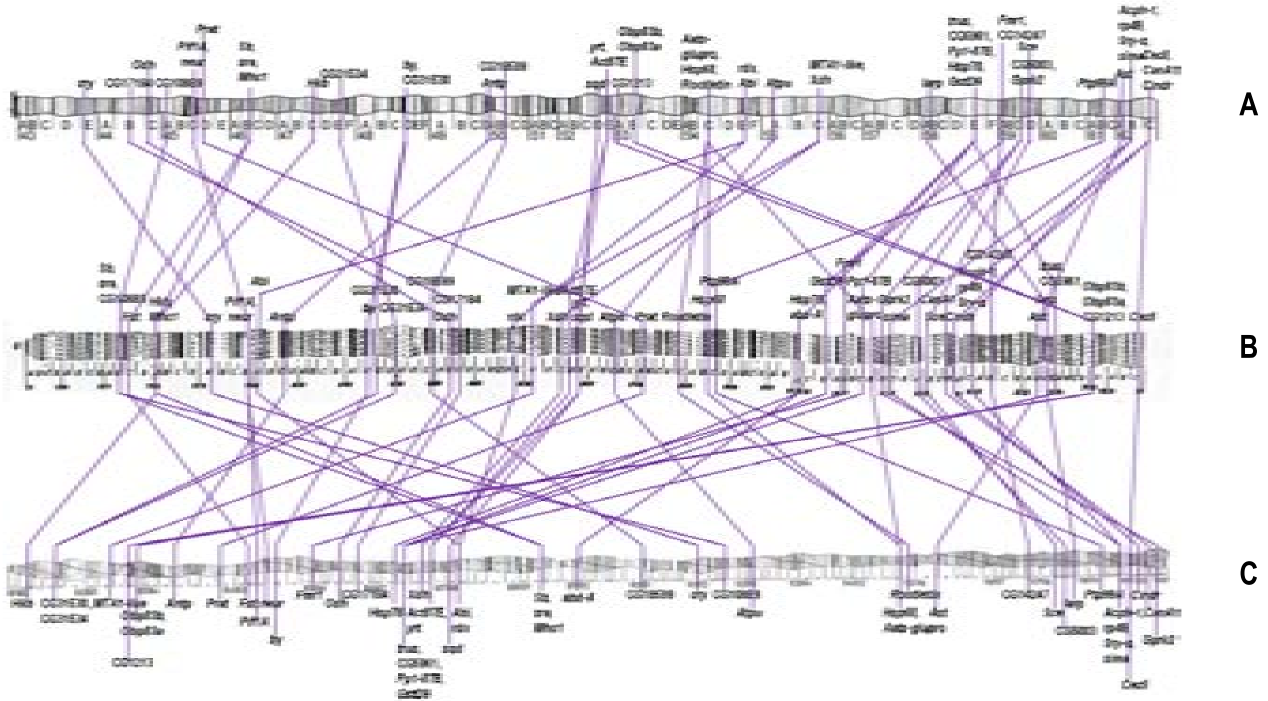


Figure 4.9: Comparison of the physical locations of the 52 gene regions across the homologous chromosomes of *D. pseudoobscura* (A), *D. subobscura* (B) and *D. melanogaster* (C). Each chromosome is oriented so that its centromere is pointing up. The length of each chromosome is the same to facilitate across-species comparison and does not reflect size differences between chromosomes.

4.3. GENETIC DIVERSITY AND DIVERGENCE OF THE GENE REGIONS

All six sequenced genes were localized inside of inversions in *Segment I* of the *O* chromosome (see Figure 3.3), which made them suitable to analyze the genetic content of the most frequent arrangements O_{ST}, O_{3+4} and O_{3+4+8} between the two latitudinally distributed populations (Barcelona and Málaga) separated by more than 800Km linear distance. Overall 60 individual lines were used with 38 to 53 lines being sequenced per each genomic region (Table 4.2). Twenty-five lines sequenced for all gene regions were used to form a concatenated dataset.

Table 4.2: Number of lines for different gene regions, populations and chromosomal arrangements.

Gene region	Populations	Inversions		
		O_{ST}	O_{3+4}	O_{3+4+8}
Acph-1	BCN	9	10	10
	MLG	7	10	5
Ast	BCN	9	7	9
	MLG	4	9	3
larp	BCN	9	11	10
	MLG	6	9	5
reg_*larp*	BCN	8	10	10
	MLG	3	5	2
CG5961	BCN	9	10	9
	MLG	4	9	3
trus	BCN	9	10	9
	MLG	4	9	3
Fmr1	BCN	8	10	9
	MLG	5	9	2
reg_*Fmr1*	BCN	9	10	10
	MLG	7	10	7
Concat.	BCN	7	7	6
	MLG	2	3	0

The multiple alignment of the eight genetic regions in the 25 isogenic lines of *D. subobscura* consisted in 12.067 sites after excluding sites with alignment gaps. The length of each region as well as their molecular functions/biological process and primers used in sequencing work are shown in Table 4.3.

The full names of the regions and their locations on the *O* chromosome are detailed in Supplementary table S1. A total of 683 nucleotide polymorphic sites (301 singletons), which correspond to at least 699 mutations, were detected: 197 in coding regions (149 synonymous and 48 nonsynonymous) and 484 in noncoding regions (see haplotypes in Supplementary Table S2).

Gene region	Primers (5'-3')	Size, bp	Molecular Function/Biological process
Acph-1	F-TCCTATGGTCAACGCCTATCG R-GTTTTTCATTACCAAATGCAC	1865	Acid phosphatase activity
Ast	F-CCACGAGATAATAGGCGGAAA R-GTCCGCAGCCCTTCAACTTTG	1840	Hormone activity/ Neuropeptide signaling pathway
larp	F-CCAATTCGCACTCGATTGAT R-CGTGGTATGCGATTAGTCGG	1915	La type RNA-binding/mitochondrion inheritance; mitotic chromosome condensation; centrosome separation; spindle assembly involved in male meiosis; syncytial blastoderm mitotic cell cycle; male meiosis; spindle assembly; autophagic cell death; salivary gland cell autophagic cell death
reg_*larp*	F-CTACACTTGGCCGACTCCTC R-CGATGATAGGCAGATGGCTT	1555	*Ca. 4 kb upstream larp* Regulatory region of *larp*
CG5961/	F-CTGTCTGCAAAGGCTTCTATG	598	Unknown
trus	R-ACCCAGCACTTGGACAATCG	1384	Unknown
Fmr1	F-CCATTCACCAGACCTTCCTT R-ACAGCCAAGTCGTTCTACCA	1972	Protein binding; protein self-association; mRNA binding; RNA binding/ biological regulation; cellular component organization or biogenesis; multicellular organism reproduction; regulation of developmental process; neuron differentiation; rhythmic process; learning or memory; synaptic transmission; locomotory behavior; neuron projection development
reg_*Fmr1*	F-GGGCAGCCGTAAGTTAACAA R-GAGCCGAACTTCCACCAATA	850	*Ca. 1 kb upstream Fmr1* Regulatory region of *Fmr1*
Total		12067	

Table 4.3: **Characteristics of the sequenced regions.**

4.3.1. GENETIC DIFFERENTIATION BETWEEN POPULATIONS WITHIN CHROMOSOMAL ARRANGEMENTS

One of the aims of the present work was to study the genetic differentiation between two distant populations at the molecular level. As it was described above in the previous chapter the European populations differ significantly in their chromosomal arrangement polymorphism. For instance, the chromosomes with O_{ST} arrangement represent the 27% of the Barcelona population, but its frequency decreases to 5% in the population of Málaga, the same pattern was observed with the arrangement $O_{\underline{3+4+8}}$ with 7% frequency in Málaga and 26% in Barcelona. These differences could be due to differential selection to local environments despite high levels of gene flow (PASCUAL *et al.* 2001; PEGUEROLES *et al.* 2013). The genetic differentiation between populations and gene arrangements was estimated with F_{ST}, which is the average number of differences in the sequences of a population in relation to the average number between two populations (HUDSON, SLATKIN and MADDISON 1992). When comparing populations for the same arrangement, F_{ST} values were small in all cases and negative in 31% of the comparisons (Table 4.4). Negative F_{ST} values imply that the mean number of differences within population is higher than between populations. No differentiation was observed between populations with the exception of the regulatory region of *Fmr1* for $O_{\underline{3+4}}$ arrangement, whose F_{ST} was small (0.049) but significant (P=0.003; Table 4.4). Thus, despite we found differences in chromosomal arrangement frequencies between the two populations, at the genetic level, there were no significant differences between them for the same

arrangement in 23 out of the 24 comparisons encompassing the different gene regions and arrangements studied. Moreover, the levels of differentiation obtained were independent from the gene location inside the inversions suggesting high levels of gene flow between localities or a very recent origin of the arrangements.

Contrasting results had been previously obtained in different studies. Results of the first electrophoretic studies at the *Lap*, *Pept-1* and *Acph* loci (located in *Segment I* of the *O* chromosome) did not detect genetic differentiation among populations within a given chromosomal arrangement (PINSKER and SPERLICH 1981; LARRUGA and PINSKER 1984; PINSKER and BÖHM 1989). However, PREVOSTI *et al.* (1983) showed that the frequency of allele 0.40 of the *Pept-1* locus within the O_{3+4} chromosomal arrangement was negatively correlated with latitude. In addition, FONTDEVILA *et al.* (1983) and RODRIGUEZ-TRELLES (1993) detected seasonal changes repeated over a period of fifteen years for *Lap* and *Pept-I* within the O_{ST} arrangement. Moreover, analysis of disequilibrium between *Lap* and *Pept-1* for a set of European populations of *D. subobscura* showed that the observed deviations from random association are inconsistent with sampling error (ZAPATA and ALVAREZ 1992).

Table 4.4: Genetic differentiation between populations calculated as F_{ST}.

	BCN *vs.* MLG		
	O_{ST}	O_{3+4}	O_{3+4+8}
Acph-1	0.071	-0.020	-0.063
Ast	0.021	-0.010	0.236
larp	0.043	0.079	-0.016
reg_*larp*	-0.053	0.072	0.072
CG5961	-0.048	-0.071	0.292
trus	0.137	-0.022	0.000
Fmr1	0.020	0.001	-0.169
reg_*Fmr1*	0.004	**0.049***	0.130

P-value of *S*nn (HUDSON 2000): *, P<0.05.

No significant genetic differentiation was detected between populations using restriction length polymorphism in O_{ST} and O_{3+4} arrangements (ROZAS and AGUADÉ 1990; ROZAS *et al.* 1995) also suggesting high levels of gene flow. Using lethal genes as markers in Palearctic as well as colonizer populations extensive gene flow was also reported (ZIVANOVIC, ARENAS and MESTRES 2007; MESTRES *et al.* 2004). Similarly, using microsatellite loci no differentiation was detected between European populations for chromosomal arrangements of the *A, J* and *U* chromosomes

(SIMOES *et al.* 2012). Finally using both candidate and non-candidate genes to thermal adaptation no genetic differentiation was detected within arrangements of distantly related populations in the two sides of the Mediterranean Sea (ARAÚZ *et al.* 2011; PEGUEROLES *et al*. 2013). No significant DNA sequence differentiation was found between El Pedroso (Spain) and Bizerte (Tunisia) for the $O_{[3+4]}$ group (including $O_{\underline{3+4}+7}$ and $O_{\underline{3+4+8}}$), despite in these group different overlapping inversions were included (SÁNCHEZ-GRACIA and ROZAS 2011). In the same way, no significant genetic differentiation was observed among North American populations of *D. pseudoobscura* using genes located in the inverted regions (SCHAEFFER *et al.* 2003). However, in the studies on *D. melanogaster*, using microsatellite markers, which were not necessarily related to the inversions, significant differentiation was observed between African populations, where this species is native (CARACRISTI and SCHLÖTTERER 2003; POOL and AQUADRO 2006). Therefore, the low differentiation between *D. subobscura* populations may depend on their isolation and might be affected by their high ability for active dispersion (SERRA, PEGUEROLES and MESTRES 1987) or by migration facilitated by passive transport associated to human activities (PASCUAL *et al.* 2007). Nonehtless, genetic differentiation within arrangement between introduced (American) and ancestral (European) populations as observed for the *Odh* gene (MESTRES *et al*. 2004; GÓMEZ-BALDÓ *et al*. 2008; ARAÚZ *et al*. 2011) would be due not solely to isolation but mostly to the founder effect. In European populations, gene flow seems to be extensive due to the high similarity of the gene content of the same arrangement across a large gradient and thus the clinal frequencies of some inversions are likely to be maintained by strong selection (PREVOSTI *et al*. 1988; BALANYÀ *et al*. 2006).

In summary, despite using some candidate genes for thermal adaptation (which are suspected to be selected in the chromosome inversions where they are located) we observed no differences between populations within arrangement, except in the regulatory region of *Fmr1* for $O_{\underline{3+4}}$, despite their significant differences in chromosomal frequencies. Moreover, in the previous study no fitness differences were observed when comparing heterokaryotypes carrying both chromosomes belonging to the same or to different populations (PEGUEROLES *et al*. 2010a). Thus, we conclude that the adaptive value of inversions can be maintained regardless the lack of genetic differentiation within arrangements from different populations. These results do not agree with the expectation of the coadaptation model understood as it was formulated by Th. Dobzhansky, that predicts genetic differentiation between populations (DOBZHANSKY 1950) and apparently does not apply neither to *D. pseudoobscura* (SCHAEFFER *et al.* 2003), but do support the local adaptation

hypothesis of KIRKPATRICK and BARTON (2006).

4.3.2. NUCLEOTIDE DIVERSITY

Nucleotide variation estimates are shown in Table 4.5 and on Figure 4.10. Genetic variability were obtained for each chromosomal arrangement pooling data of the different populations to increase sample size given the lack of genetic differentiation between localities within arrangements. Almost in all cases the number of haplotypes detected approached the number of sequences analyzed except in the case of *larp* in $O_{\underline{3+4+8}}$ and *CG5961* in O_{ST} and $O_{\underline{3+4}}$ arrangements.

For each gene region Θ_{sil} was somewhat higher than π_{sil} (Table 4.4) but the difference between them was only significant for $O_{\underline{3+4}}$ arrangement (Z=3.1; *P*=0.002), indicating significant excess of low frequency alleles in this arrangement, confirmed further with Tajima´s neutrality test (see below). These data are also supported by previous results on restriction length polymorphism in *rp49* gene of ROZAS and AGUADÉ (1990), where Hudson's estimate of heterozygosity per nucleotide (Θ) was higher than π in the same three chromosomal arrangements due to the high number of rare sites, which may indicate some deviation from neutrality.

The estimations of silent nucleotide diversity and number of segregating sites were similar between the three chromosomal arrangements (Table 4.6). The absence of differences in nucleotide diversity is in contradiction with previous data, where the estimates of nucleotide diversity were higher in $O_{\underline{3+4}}$ than in O_{ST} (ROZAS and AGUADÉ 1994; MUNTÉ *et al.* 2005 but see PEGUEROLES *et al.* 2013). For instance differences between O_{ST} and $O_{\underline{3+4}}$ arrangements for the number of singletons, silent and nonsynonymous polymorphism had been reported for the *Acph-1* gene (NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a), while no differences in variability had been previously reported between these two arrangements for the coding region of *Fmr1* (PEGUEROLES *et al.* 2013). However, the difference in π_{sil} was nearly significant for the pair of arrangements $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ (Table 4.6) with higher silent variability in $O_{\underline{3+4+8}}$ arrangement. The higher diversity in $O_{\underline{3+4+8}}$ chromosomal arrangement was mainly due to the CG5961 gene region (Figure 4.10), which includes only large exon and presents a reduced number of haplotypes in the other two chromosomal arrangements. The highest variability was detected in *Acph-1*, which was already characterized as highly variable in previous studies

(MUNTÉ *et al.* 2005; NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a; NAVARRO-SABATÉ, AGUADÉ and SEGARRA 2003).

Table 4.5: Nucleotide variation and divergence for each candidate gene region and arrangement.

Gene		*Acph-1*	*Ast*	*larp*	*reg_larp*	CG5961	*trus*	*Fmr-1*	*reg_Fmr1*	Concatenated data
Size, bp		1865	1840	1915	1555	598	1384	1972	850	12067
O_{ST}	*n*	16	13	15	11	13	13	13	16	9
	h	16	13	15	11	7	13	13	15	9
	S	95	45	35	105	8	23	78	28	284
	singletons	38	30	24	75	7	12	55	13	185
	π	0.013	0.006	0.004	0.016	0.005	0.005	0.009	0.007	0.007
	π_{sil}	0.021	0.007	0.009	0.016	0.009	0.011	0.011	0.007	0.010
	Θ_{sil}	0.025	0.009	0.012	0.022	0.015	0.013	0.016	0.009	0.012
	K_{sil}	0.285	0.12	0.126	0.09	0.166	0.250	0.219	0.227	0.171
O_{3+4}	*n*	20	16	20	15	19	19	19	20	10
	h	20	16	17	15	9	19	19	18	10
	S	93	77	36	57	7	30	75	31	282
	singletons	39	44	25	29	4	22	37	16	173
	π	0.012	0.010	0.003	0.009	0.003	0.004	0.008	0.008	0.007
	π_{sil}	0.020	0.011	0.007	0.009	0.010	0.005	0.010	0.008	0.010
	Θ_{sil}	0.022	0.014	0.012	0.012	0.012	0.011	0.014	0.010	0.011
	K_{sil}	0.284	0.12	0.123	0.09	0.168	0.243	0.222	0.225	0.169
O_{3+4+8}	*n*	15	12	15	10	12	12	11	17	6
	h	15	12	11	10	10	10	11	17	6
	S	82	59	26	83	15	31	44	43	225
	singletons	27	33	18	29	4	28	21	31	119
	π	0.013	0.008	0.003	0.016	0.007	0.004	0.007	0.009	0.008
	π_{sil}	0.022	0.010	0.006	0.016	0.028	0.008	0.009	0.009	0.012
	Θ_{sil}	0.022	0.012	0.007	0.018	n.a.	0.014	0.010	0.015	0.012
	K_{sil}	0.279	0.119	0.124	0.084	0.168	0.246	0.222	0.228	0.169

Figure 4.10: Genetic variation per silent site for whole genomic sequences and each gene arrangement.

The regulatory region of *larp* was also highly variable in O_{ST} and $O_{\underline{3+4+8}}$ gene arrangements probably contributing to the nearly significant levels of differentiation when comparing $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements. Our results are in contradiction with previous studies where diversity was higher for $O_{\underline{3+4}}$ and lower in $O_{\underline{3+4+8}}$ (ROZAS and AGUADÉ 1990).

Table 4.6: **Comparison of the nucleotide diversity between gene arrangement pairs, pooling the two populations, with the Wilcoxon Matched Pairs Test.**

Arrangement pair	*S*		π_{sil}		Θ_{sil}		π	
	Z	*P*	*Z*	*P*	*Z*	*P*	*Z*	*P*
O_{ST}/O_{3+4}	0.711	0.477	0.457	0.647	0.175	0.861	0.188	0.851
O_{ST}/O_{3+4+8}	1.007	0.314	0.350	0.726	0.738	0.460	0.982	0.326
O_{3+4}/O_{3+4+8}	0.178	0.859	1.896	0.058	0.031	0.975	0.909	0.364

4.3.3. GENETIC DIFFERENTIATION BETWEEN CHROMOSOMAL ARRANGEMENTS

When comparing chromosomal arrangements, F_{ST} values were large and *S*nn values were almost always significantly different from zero (Table 4.7). Between $O_{ST}/O_{\underline{3+4}}$ and $O_{ST}/O_{\underline{3+4+8}}$ arrangements F_{ST} values ranged between 0.298-0.818 and 0.314-0.766 respectively, and were similar for each gene region, with the smallest value corresponding to the gene *larp* and the highest to the *Fmr1*. The higher values of genetic differentiation in all comparisons including O_{ST} chromosomal arrangement *vs.* comparisons between $O_{\underline{3+4}+X}$ chromosomes are in agreement with previous observations of ROZAS *et al.* (1999) based on *rp49* gene. Nonetheless, although lower, significant genetic differentiation was observed between $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ in all gene regions with the exception of *trus* gene. At the same time there were no fixed mutations between $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements in any gene region and the numbers of shared polymorphisms were larger than for the other two pairs of gene arrangements, except in gene *larp* and its regulatory region. Despite the lack of fixed differences and the presence of shared polymorphisms between these two arrangements, genetic differentiation prevails because of the relatively high number of almost fixed differences between them.

As shown in Table 4.7, and Figure 4.11, the highest estimates of genetic differentiation between the O_{ST} and $O_{\underline{3+4}}$ arrangements were detected in the genes CG5961, *Fmr1* and *trus.* However differentiation was not evenly distributed within each gene fragment since it was especially high in

the coding region of *trus* and in the intronic region of *Fmr1* (Supplementary Table S3). The estimation of the F_{ST} value for *Fmr1* in a previous work (PEGUEROLES *et al.* 2013) was much lower (0.497) most likely because the chosen region in that study included more coding than noncoding positions. The F_{ST} values for the regions CG5961 and *trus* were also high in the paper of MUNTÉ *et al*. (2005), where they were analyzed together as region P22, although differentiation was slightly lower (0.66) than in our study. The genes *Acph-1* and *Ast* also had similar values to ours estimations of F_{ST} between the O_{ST} and $O_{\underline{3+4}}$ arrangements equal to 0.43 and 0.53 correspondingly (MUNTÉ *et al*. 2005). The *larp* gene showed the lowest values of F_{ST}. The regulatory regions of *larp* and *Fmr1* showed similar genetic differentiation between these two chromosomal arrangements, higher than its coding region for *larp* and lower than its coding region for *Fmr1*.

Table 4.7: Genetic differentiation between chromosomal arrangements calculated as F_{ST}.

	O_{ST} vs. $O_{\underline{3+4}}$	O_{ST} vs. $O_{\underline{3+4+8}}$	$O_{\underline{3+4}}$ vs. $O_{\underline{3+4+8}}$
Acph-1	**0.412*** **	**0.408*** **	**0.161*** **
Fixed	0	0	0
Shared	44	34	52
Ast	**0.482*** **	**0.533*** **	**0.145***
Fixed	**3**	0	0
Shared	18	13	37
larp	**0.298*** **	**0.314*** **	**0.076** **
Fixed	0	**1**	0
Shared	7	3	4
reg_*larp*	**0.494*** **	**0.433*** **	**0.520*** **
Fixed	0	0	0
Shared	13	44	19
CG5961	**0.818*** **	**0.623*** **	**0.041***
Fixed	**6**	0	0
Shared	0	1	4
trus	**0.707*** **	**0.652*** **	0.068
Fixed	**8**	**1**	0
Shared	0	9	9
Fmr1	**0.763*** **	**0.766*** **	**0.051***
Fixed	**18**	**19**	0
Shared	6	8	29
reg_*Fmr1*	**0.435*** **	**0.382*** **	**0.021** **
Fixed	**1**	0	0
Shared	2	6	11
Concatenated data[a]	**0.638*** **	**0.612*** **	**0.261*** **

P-value of *S*nn (Hudson, 2000): *, 0.01<P<0.05; **, 0.001<P<0.01; ***, P<0.001. *Fixed* and *Shared* correspond to the numbers of fixed and shared differences between samples. [a]The concatenated data set includes all eight regions and only 25 lines in total (see materials and methods).

The divergence estimations between O_{ST} and $O_{\underline{3+4+8}}$ were more homogeneous, but the same gene regions demonstrated the highest and the lowest values. Genetic differentiation was strong in each region as well as in the concatenated data set in coding and noncoding regions of the genes in these two pairs of gene arrangements with one exception in coding region of *Fmr1* where the

differentiation between O_{ST} and $O_{\underline{3+4+8}}$ was non-significant (Supplementary Table S3). Low differentiation was observed between $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ for most gene regions except for the regulatory region of *larp* where it was even higher than between pairwise comparisons including O_{ST} chromosomal arrangement (Figure 4.11). The low F_{ST} values between these arrangements are easily comprehensible taking into consideration that they differ by only one big inversion (O_8) since inversions O_3 and O_4 are shared by them which allows higher levels of gene flux through recombination between $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements.

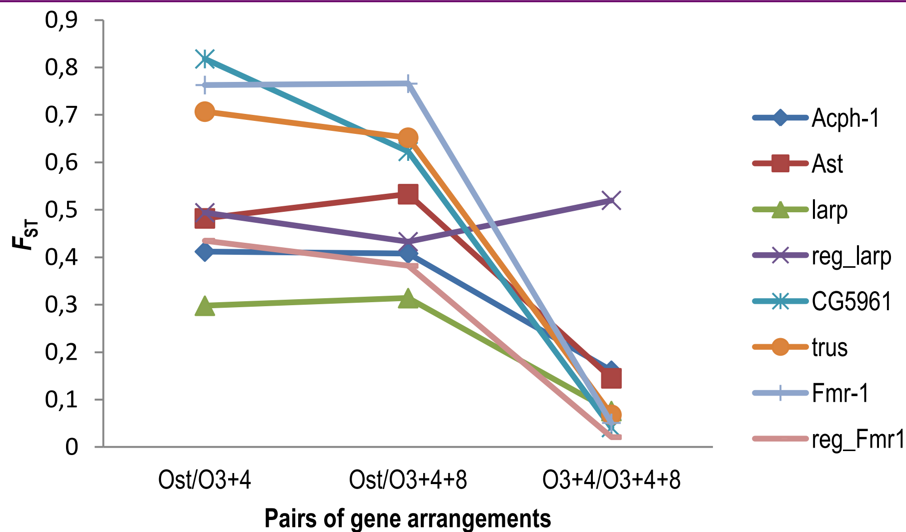


Figure 4.11: **Genetic differentiation between gene arrangements: F_{ST}.**

In the present study, F_{ST} values obtained for arrangements differing by a single inversion ($O_{\underline{3+4}}$ *vs.* $O_{\underline{3+4+8}}$) were significant but smaller than compared to overlapped inversions (O_{ST} *vs.* $O_{\underline{3+4}+X}$, where X means another inversion) as it was found in previous works (ROZAS *et al.* 1999; MUNTÉ *et al.* 2005; PEGUEROLES *et al.* 2013). Similarly, genetic differentiation between A_2 and A_{ST} arrangements (also differing by a single inversion) was significant, but F_{ST} values were also smaller than when comparing with overlapped inversions (NÓBREGA *et al.* 2008). Finally, one of the genes located inside *SI* region (*rp49*) was also sequenced in $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+23}}$ arrangements, and again F_{ST} values obtained comparing these arrangements with the $O_{\underline{3+4}}$ (also differing only by a single inversion) were significant but lower than when comparing overlapping inversions (ROZAS *et al.* 1999). The same pattern was observed in *D. pseudoobscura*, since higher levels of genetic differentiation were found when comparing arrangements differing by overlapped inversions (SCHAEFFER *et al.* 2003). However, the rule of higher differentiation involving overlapping inversions

seems not to be accomplished when comparing $O_{\underline{3+4+8}}$ and $O_{\underline{3+4+23}}$ arrangements (O_8 and O_{23} overlapped) (Rozas *et al.* 1999) or $O_{\underline{3+4}+1}$ and $O_{\underline{3+4}+7}$ (O_1 and O_7 also overlapped) (Pegueroles 2010). This lack of high differentiation could be explained by recombination mediated by other arrangements differing by a single non-overlapped inversion. Most of the inversions described in the *O* chromosome of *D. subobscura* form complex arrangements with $O_{\underline{3+4}}$ arrangement (Araúz *et al.* 2009), which presents high frequency in southern European populations, the area where most of the analyzed samples come from (Krimbas 1992; Krimbas 1993). Thus, $O_{\underline{3+4+8}}$, $O_{\underline{3+4+23}}$, $O_{\underline{3+4}+1}$ and $O_{\underline{3+4}+7}$ chromosomes could recombine with one of the several arrangements differing only by one inversion (Pegueroles *et al.* 2010b). On the contrary, during meiosis, heterokaryotypes carrying the O_{ST} arrangement and another arrangement in general will form two inversion loops in region *SI*. Thus, recombination rate in heterokaryotypes for the O_{ST} arrangement in that region should be highly reduced (which is confirmed by our study elucidating significantly lower levels of recombination rate for the pairs of arrangements compared involving O_{ST}; see below). Consequently, O_{ST} could more easily maintain a cluster of adaptive genes conferring a selective advantage and explain the significant latitudinal clines in Europe and America (Prevosti *et al.* 1988; Balanyà *et al.* 2003).

The distributions of nucleotide diversity (π) and genetic differences between each pair of arrangements across each gene region were analyzed by the sliding window approach. On Figure 4.12 some examples of such distributions are shown for the two candidate genes to thermal adaptation *larp* and *Fmr1* and their regulatory regions and one non-candidate gene *trus*.

The graphs show that there are highly divergent zones in the genes *Fmr1* and *trus* in arrangement pairs O_{ST}-$O_{\underline{3+4}}$ and O_{ST}-$O_{\underline{3+4+8}}$ while there is no differentiation between $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$. For the gene *Fmr1* the main difference was detected in intron region around 750-1000 bp (D_{XY}≈0.35). In *trus* great differences between arrangements were found in its exons around 1-700 bp and 1000-1100 bp, many corresponding to nonsynonymous substitutions (see Supplementary Tables S2), but the values of D_{XY} were not so high ranging around 0.012-0.042. As for the gene *larp* the arrangements did not show high divergence, although the largest levels of D_{XY} were detected around 500-650 bp, 1250-1400 bp and 1750-2000 in O_{ST}-$O_{\underline{3+4}}$ and O_{ST}-$O_{\underline{3+4+8}}$. In the case of regulatory region of *Fmr1,* there were divergent positions between chromosomal arrangement pairs O_{ST}-$O_{\underline{3+4}}$ and O_{ST}-$O_{\underline{3+4+8}}$ (although the differences were not very high and the level of D_{XY} around

0.015 – 0.04), but not between $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ as in the rest of the gene regions. The situation was somewhat different for the regulatory region of *larp*, which has shown great divergence in all pairs of chromosomal arrangements (as seen in Figure 4.12) around nucleotide positions 900-1130.

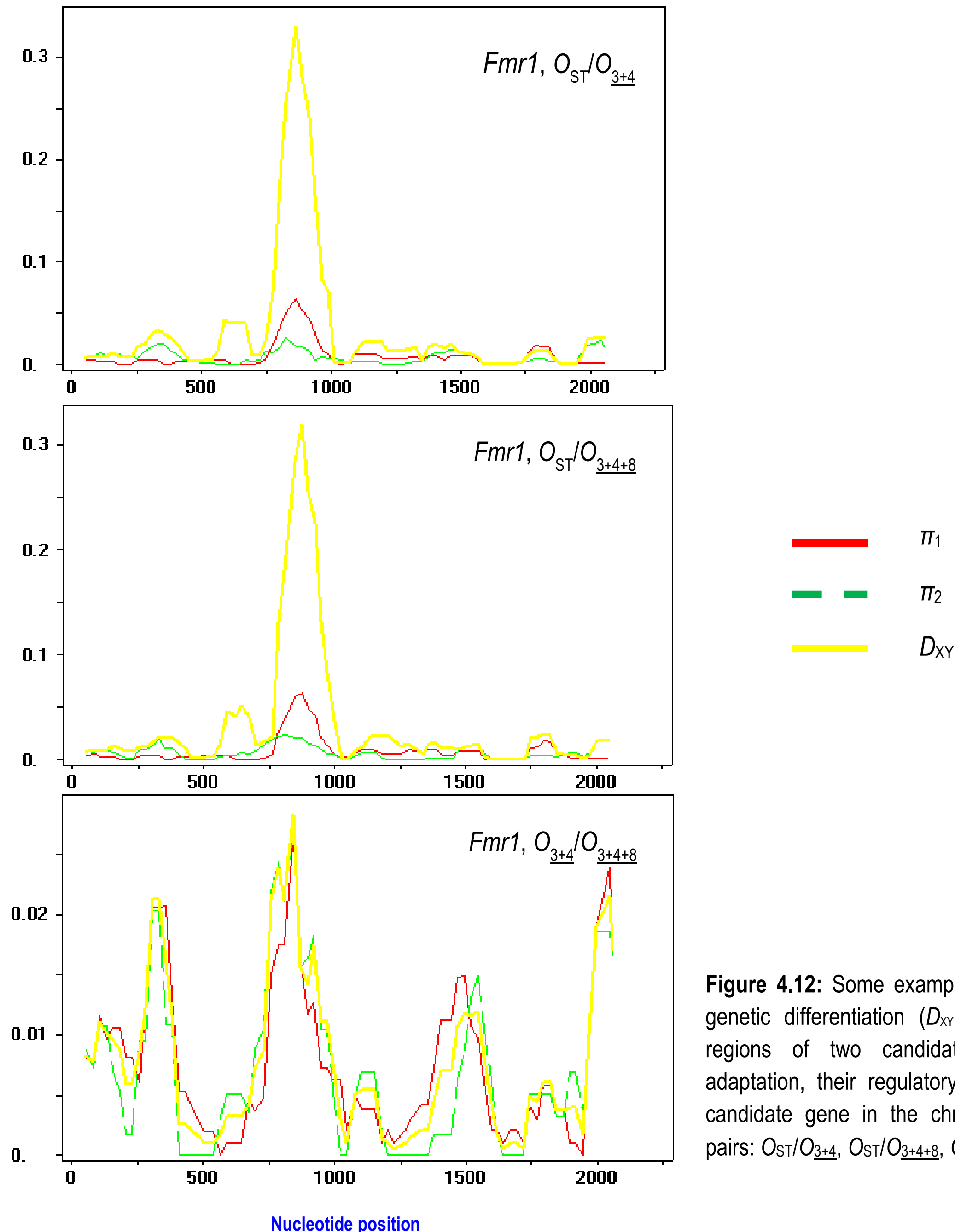


Figure 4.12: Some examples of variability (π) and genetic differentiation (D_{XY}) along the sequenced regions of two candidate genes for thermal adaptation, their regulatory regions and one non-candidate gene in the chromosomal arrangement pairs: $O_{ST}/O_{\underline{3+4}}$, $O_{ST}/O_{\underline{3+4+8}}$, $O_{\underline{3+4}}/O_{\underline{3+4+8}}$.

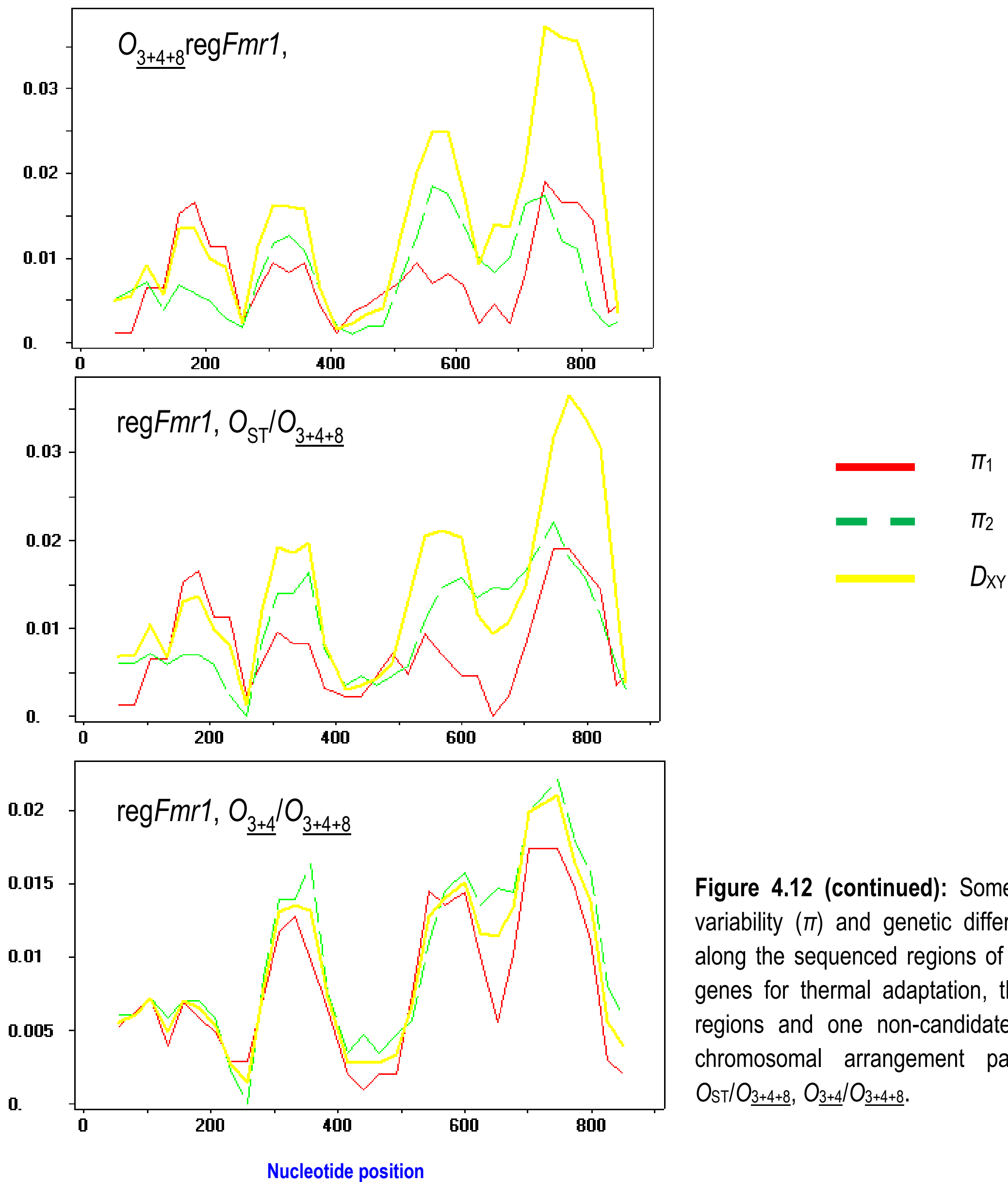
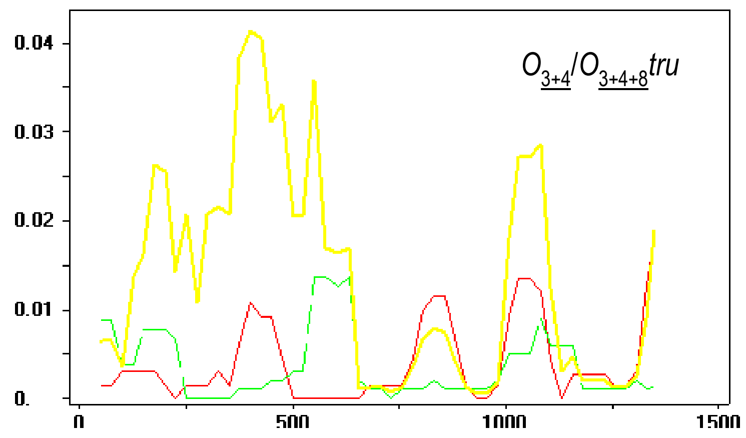
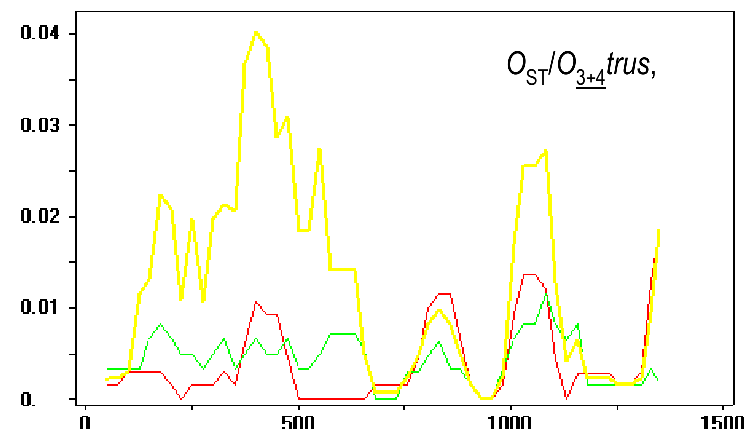
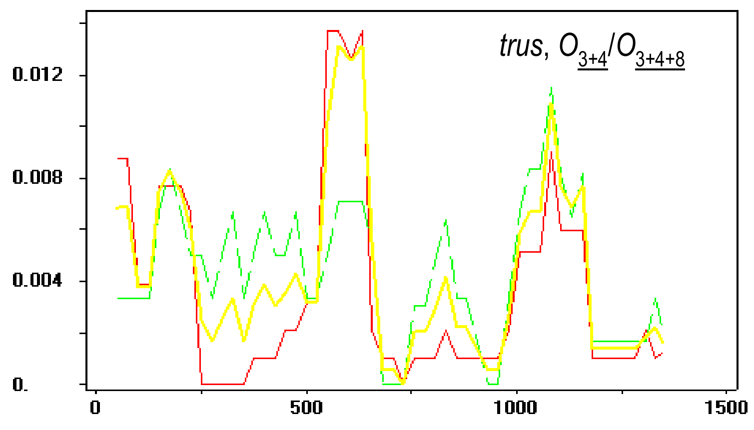


Figure 4.12 (continued): Some examples of variability (π) and genetic differentiation (D_{XY}) along the sequenced regions of two candidate genes for thermal adaptation, their regulatory regions and one non-candidate gene in the chromosomal arrangement pairs: $O_{ST}/\underline{O_{3+4}}$, $O_{ST}/\underline{O_{3+4+8}}$, $\underline{O_{3+4}}/\underline{O_{3+4+8}}$.





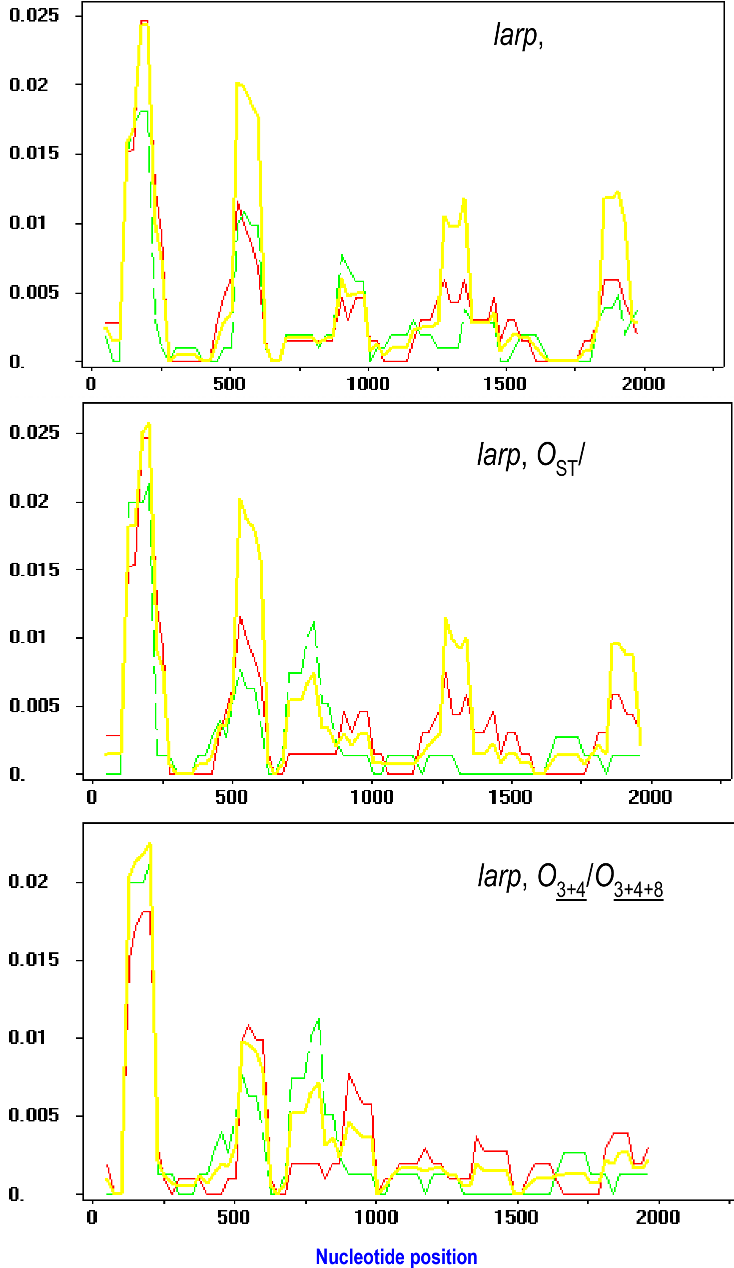


Nucleotide position

π_1

π_2

D_{XY}



π_1

π_2

D_{XY}

Figure 4.12 (continued): Some examples of variability (π) and genetic differentiation (D_{XY}) along the sequenced regions of two candidate genes for thermal adaptation, their regulatory regions and one non-candidate gene in the chromosomal arrangement pairs: $O_{ST}/O_{\underline{3+4}}$, $O_{ST}/O_{\underline{3+4+8}}$, $O_{\underline{3+4}}/O_{\underline{3+4+8}}$.

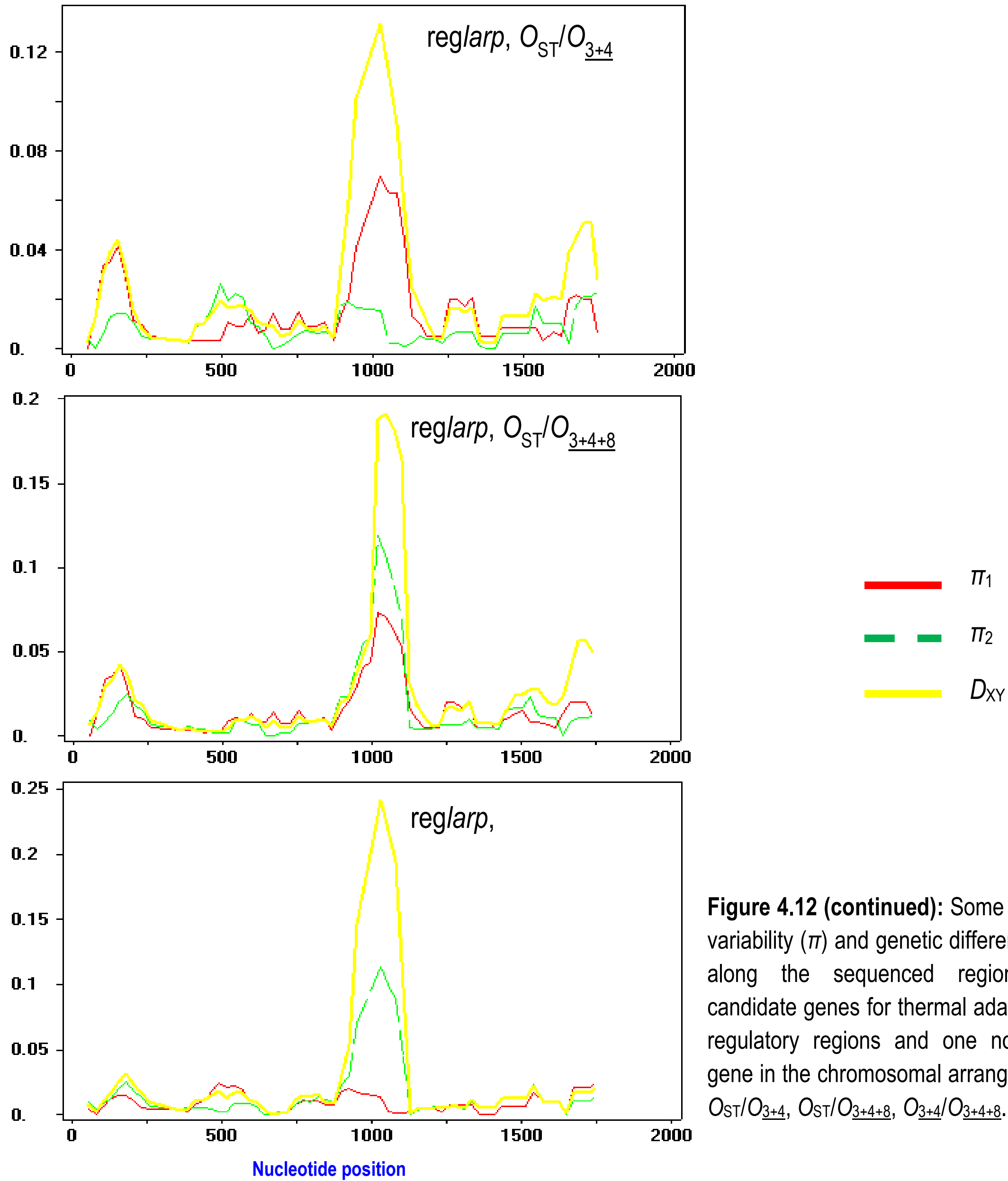


Figure 4.12 (continued): Some examples of variability (π) and genetic differentiation (D_{XY}) along the sequenced regions of two candidate genes for thermal adaptation, their regulatory regions and one non-candidate gene in the chromosomal arrangement pairs: $O_{ST}/O_{\underline{3+4}}$, $O_{ST}/O_{\underline{3+4+8}}$, $O_{\underline{3+4}}/O_{\underline{3+4+8}}$.

Higher levels of variability usually accompanied the high divergent values in both chromosomal arrangements under comparison in all gene regions, although there were positions in each region with much difference between π and D_{XY} values in arrangement pairs O_{ST}-$O_{\underline{3+4}}$ and O_{ST}-$O_{\underline{3+4+8}}$, which could indicate the action of selection in these positions. For the rest of the genes similar patterns were observed with higher levels of D_{XY} in O_{ST}-$O_{\underline{3+4}}$ and O_{ST}-$O_{\underline{3+4+8}}$ pairs than for the arrangement pair $O_{\underline{3+4}}$-$O_{\underline{3+4+8}}$.

D_{XY} is an estimate of the average number of nucleotide substitutions per site between arrangements. However, the levels of nucleotide diversity influence its value, and they can be affected by factors such as neutral mutation rate and recombination. On the other hand, F_{ST} estimates do not seem to be influenced by nucleotide diversity levels, being more suitable to compare levels of differentiation between different genes (PEGUEROLES *et al.* 2013). That was the reason why we choose the latter estimate to measure overall differentiation between populations and gene arrangements.

We failed to detect significant correlation between levels of silent nucleotide diversity in a gene region and its physical distance to the nearest inversion break point (Table 4.8, Figure 4.13). No correlation was neither found between F_{ST} values for each gene region and its genetic distance to the nearest inversion break point for the three pairwise chromosomal arrangements comparisons (Table 4.8, Figure 4.14). The non-significant correlation coefficients for both genetic diversity and divergence estimates could suggest the selection as the main factor maintaining genetic differentiation between inversions independently from the position of the genes.

Table 4.8: Correlation coefficients between π_{sil} and F_{ST} values and the distance of the eight regions to the nearest inversion break point.

	Arrangement	Correlation coefficient, β	*P*
π_{sil}	O_{ST}	-0.650	0.081
	O_{3+4}	-0.608	0.110
	O_{3+4+8}	-0.569	0.141
F_{ST}	O_{ST}/O_{3+4}	-0.054	0.899
	O_{ST}/O_{3+4+8}	-0.007	0.987
	O_{3+4}/O_{3+4+8}	-0.173	0.683

Overall, regions close to break points did not show any reduction in nucleotide diversity, as it could be expected under neutral model of inversions evolution, and the variability was even higher for some genes situated near the break points. In fact, as was mentioned before, *Acph-1* shows the highest π_{sil} and Θ_{sil} values, in all gene arrangements, despite its tight linkage to the proximal break point of the O_3 inversion. We calculated the distance to the nearest break point of O_3, O_4 or O_8 inversion for each gene assuming that the average size of a chromosomal band is equal to 82 Kb

as it was calculated dividing the whole length of the O chromosome (≈31 Mb) by the number of bands (PEGUEROLES *et al.* 2010a).

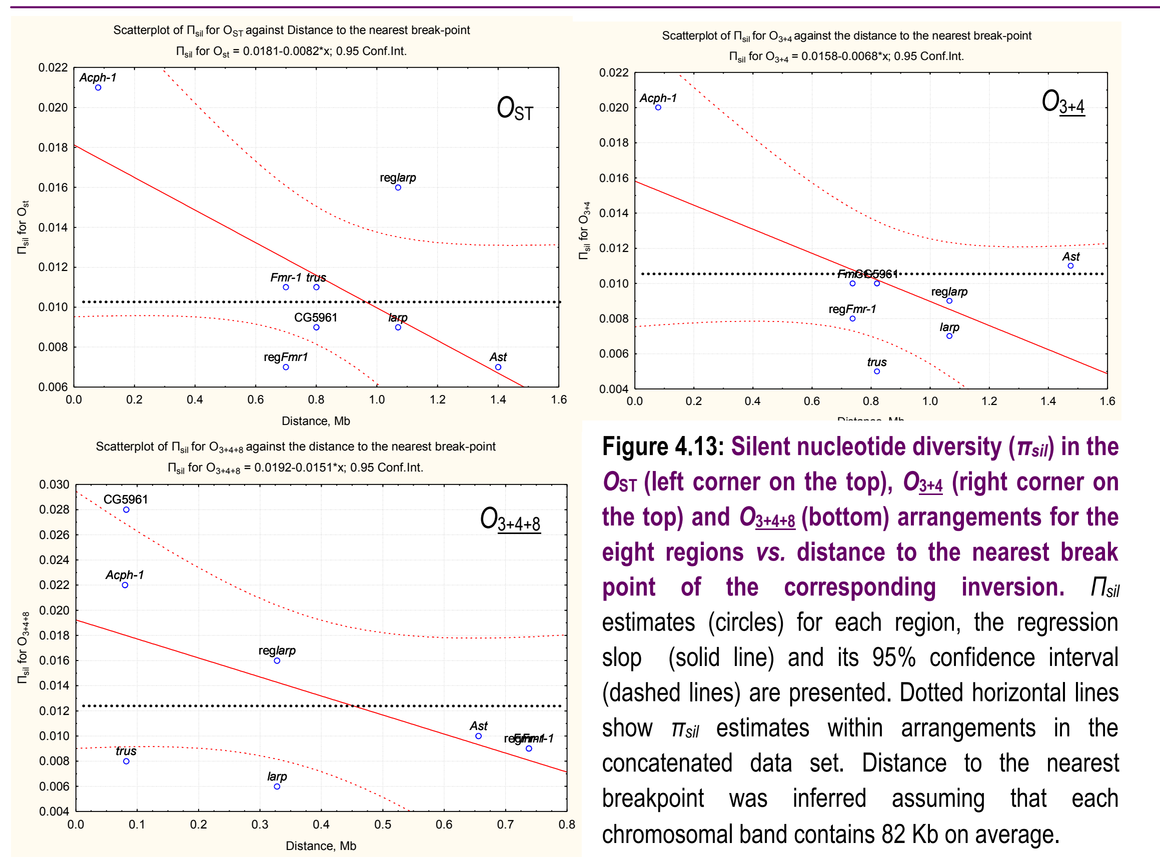


Figure 4.13: Silent nucleotide diversity (π_{sil}) in the O_{ST} (left corner on the top), O_{3+4} (right corner on the top) and O_{3+4+8} (bottom) arrangements for the eight regions *vs.* distance to the nearest break point of the corresponding inversion. Π_{sil} estimates (circles) for each region, the regression slop (solid line) and its 95% confidence interval (dashed lines) are presented. Dotted horizontal lines show π_{sil} estimates within arrangements in the concatenated data set. Distance to the nearest breakpoint was inferred assuming that each chromosomal band contains 82 Kb on average.

The recovery of variability is not expected to be uniform across the inverted region, since recombination is expected to be reduced near break points due to chromosomal mispairing (NAVARRO *et al.* 1997). Experimental studies showed that the variability is lower in nucleotide markers separated from the break point by a distance of about 1 Mb or less (ANDOLFATTO, DEPAULIS and NAVARRO 2001). Thus, the location of genes inside the inverted regions can be a key factor to take into account when studying levels of nucleotide diversity. We have detected significant differences in the silent nucleotide levels between gene regions, despite no relation with their location inside the inversion was detected. Moreover, we have not detected inferior variability in our genes located fairly close to breakpoints. Only for *larp* the reduction in variability was observed, although in this case it could be due to a selective process rather than its location with respect to the break point of the inversion. Studies in *D. subobscura* show that in general there is no decrease in variability associated with the proximity of the break points (MUNTÉ *et al.* 2005; NÓBREGA *et al.*

2008). For example, in seven genes located within 1Mb from the break points of the inversion O_{ST} and O_{3+4} the reduction in variability was not observed, and even the *Acph1* gene, which is located within 0.2 Mb, presented greater variability than the rest (MUNTÉ *et al.* 2005). The high values of observed variability and divergence in the gene *Acph1* (MUNTÉ *et al.* 2005 and this study) suggest that it could present an abnormally high rate of mutation. In another study in *D. subobscura*, but in this case in the A_2 and A_{ST} inversions, the nearest region to an inversion breakpoint showed the lowest level of variation, but not when considering it relative to divergence, to account for the mutation rate (NOBREGA *et al.* 2008). However, in other species such as *D. buzzatti* the expected reduction in diversity close to breakpoints was observed (LAAYOUNI *et al.* 2003). Therefore, the experimental data do not always show this reduction in variability, which appears to be strongly influenced by the species. Factors such as gene flow, mutation and age of the inversion can largely influence the results.

Figure 4.14: Genetic differentiation (F_{ST}) between arrangements O_{ST} and O_{3+4} (left corner on the top), O_{ST} and O_{3+4+8} (right corner on the top), O_{3+4} and O_{3+4+8} (bottom) vs. distance to the nearest break point of the corresponding inversion. F_{ST} estimates (circles) for each region, the regression slop (solid line) and its 95% confidence interval (dashed lines) are presented. Dotted horizontal lines show F_{ST} estimates within arrangements in the concatenated data set. Distance to the nearest breakpoint was inferred assuming that each chromosomal band contains 82 Kb on average.

Furthermore, no high F_{ST} values were found for genes *Acph-1*, *trus* and CG5961 despite being located very close to O_3 (*Acph-1*) and O_8 (*trus* and CG5961) break points (Figure 4.14). Nevertheless, *rp49* and *P236* genes presented higher levels of genetic differentiation when located

near inversion breakpoints (ROZAS and AGUADÉ 1994; NÓBREGA *et al.* 2008) than when located in a more central position (ROZAS *et al.* 1999) as the regulatory region of *larp* in our study. Thus, genetic differentiation close to inversion break points can also be eroded through time at a gene specific rate depending on selection acting upon this gene.

All in all, the differentiation between arrangements could be due to either recent origin of an inversion, recombination reduction mediated by inversion or/and selection binding together coadapted genes captured within inversions. Many observations support the adaptive value of inversions in *D. subobscura* such as the latitudinal clines for some chromosomal arrangements (PREVOSTI *et al.* 1988). In particular, O_{ST} and $O_{\underline{3+4}}$ present opposite latitudinal clines, with O_{ST} being more frequent in northern Europe and $O_{\underline{3+4}}$ being more frequent in the south (SOLÉ *et al.* 2002; BALANYÀ *et al.* 2004). Besides, genetic differentiation could be also maintained by lower rate of exchange inside the inverted regions. This effect could be reinforced by the presence of overlapped inversions as in the $O_{\underline{3+4}}$ arrangement, which reduces in a greater extent genetic exchange in comparison to single inversions (NÓBREGA *et al.* 2008).

4.3.4. LINKAGE DISEQUILIBRIUM

Genetic differentiation between arrangements may also be detected by the analysis of linkage disequilibrium between nucleotide polymorphic sites and the type of chromosomal arrangement. Indeed, the presence of fixed or almost fixed variants between arrangements at a particular site should cause a strong association between the variants at this site and the type of chromosomal arrangement. This analysis was performed for all informative polymorphic sites and the chromosomal arrangements. The extent of linkage disequilibrium was estimated by the percentage of comparisons that were significant after applying the Fisher test for each gene region and concatenated dataset and by the average R^2 value.

For each population, gene region, chromosomal arrangement and pairs of chromosomal arrangements, the percentage of pairwise comparisons between positions that are significant before and after applying the Bonferroni correction and the parameter of global disequilibrium *ZnS* for comparison of the three arrangements were calculated. In addition, to see whether there are associations between genes, the linkage disequilibrium for the eight concatenated regions of the

two populations was also analyzed separately. Furthermore, we calculated the total linkage disequilibrium per gene arrangement grouping the two populations for each region as well as for the eight concatenated genes. All the data on linkage disequilibrium can be find in the Supplemental Table S4.

No informative sites were found in several gene regions in population of Málaga: in the regulatory region of *larp* and CG5961 for O_{ST} and $O_{\underline{3+4+8}}$ arrangements, and in *Ast*, *trus* and *Fmr1* for $O_{\underline{3+4+8}}$ arrangement due to the scarce number of lines. So, as there were not enough individuals in the Málaga population, we grouped the two populations to increase the sample and to compare directly the two arrangements and thus, some associations that were non-significant previously because they may not have enough weight, became significant.

If recombination in heterokaryotypes is reduced, as in our study (see next section below), and knowing that recombination in homokaryotypes is free, we would expect lower levels of LD within arrangement than when grouping sequences from two arrangements. This was the case considering each gene region separately as well as combining them into concatenated dataset.

After Bonferroni correction the percentages of pairwise comparisons were equal to zero in most of the genes when each arrangement was considered separately, with few exceptions: *larp* in all chromosomal arrangements, and its regulatory region, *trus* and *Fmr1* for the $O_{\underline{3+4}}$ arrangement. The highest values of *ZnS* and percentage of LD significant by Fisher were found for the regulatory region of *larp* in $O_{\underline{3+4+8}}$ arrangement, although after Bonferroni correction there was no significant associations. As for when including pairs of chromosomal arrangements, the percentage of LD equal to zero was found only for CG5961 in $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ arrangement pair, while it was the highest for this gene in $O_{ST}/O_{\underline{3+4}}$ together with the *ZnS* value. These values were also high in the genes *tru*s and *Fmr1* combining the same pair of chromosomal arrangements. These three gene regions had also the highest values of linkage disequilibrium in the arrangement pair $O_{ST}/O_{\underline{3+4+8}}$. The lower *ZnS* values when grouping $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements for all genes suggest that recombination between them would not be negligible. The percentages of significant associations by Fisher increased considerably when chromosomal arrangements were grouped by pairs (31% in $O_{ST}/O_{\underline{3+4}}$, 29.3% in $O_{ST}/O_{\underline{3+4+8}}$ and 7% in $O_{\underline{3+4}}/O_{\underline{3+4+8}}$). These rather high percentages contrasting with that of arrangements analyzed separately reflects that recombination is restricted in heterokaryotypes. Similar result was reported by MUNTÉ *et al.* (2005) for the chromosomal arrangement pair $O_{ST}/O_{\underline{3+4}}$

with significant LD in concatenated data set with 385 informative sites equal to 28.8% of the pairwise comparisons. This percentage dropped to ~5% when each chromosomal arrangement was analyzed separately. Global estimates of LD, measured as *ZnS* were also higher combining chromosomal arrangements than within them: 0.2698, 0.3216, 0.1431 in the total sample of $O_{ST}/O_{\underline{3+4}}$, $O_{ST}/O_{\underline{3+4+8}}$, $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ and 0.1360, 0.1203, 0.2947 for O_{ST}, $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements respectively. Recombination in homokaryotypes would explain the lower percentage of significant LD within arrangement. The higher *ZnS* values in $O_{\underline{3+4+8}}$ from Barcelona could be due to the low frequency of this arrangement, increasing the frequency of heterokaryotypes and the presence of recombinants between arrangements (see below).

Therefore, the maintenance of these chromosomal arrangements in the population contributes to an overall reduction of recombination in the studied genes, as shown by the high percentages of significant associations between the alternative variants of informative sites and the chromosomal arrangement in each gene region as well as in concatenated dataset.

To visualize these associations, all the values of the R^2 statistic (Hill and Robertson 1968), as it is less dependent on allele frequencies, were represented graphically for each pair of chromosomal arrangements, against the distance in base pairs considering that each amplified gene fragment is consequential (Figure 4.15). The order of the genes in each comparison follows their location in corresponding arrangement (the order in O_{ST} for $O_{ST}/O_{\underline{3+4}}$, the order in $O_{\underline{3+4}}$ for $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ and the order in $O_{\underline{3+4+8}}$ for $O_{ST}/O_{\underline{3+4+8}}$; Figure 3.3). Thus, the actual distance between nucleotides in different genes in linkage disequilibrium was highly underestimated. The associations are more or less homogeneously distributed along the concatenated sequence independently on the distances between nucleotides in the arrangement pairs $O_{ST}/O_{\underline{3+4}}$ and $O_{ST}/O_{\underline{3+4+8}}$, despite some of them involved genes separated by 7 Mb on the *O* chromosome. In $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ most of the associations correspond to positions situated very close to each other, i. e., different nucleotides of the same gene region, and the majority of these associations were within regulatory region of *larp*.

In the previous work of Navarro-Sabaté, Aguadé and Segarra (2003) the analysis of linkage disequilibrium between nucleotide polymorphic sites and gene arrangements included the same arrangements from a Spanish population as our study, but it was restricted only to the gene *Acph-1*. The rational of that analysis was the same as we used here and as stated above that high levels of linkage disequilibrium in joint samples of two arrangements (relative to linkage disequilibrium

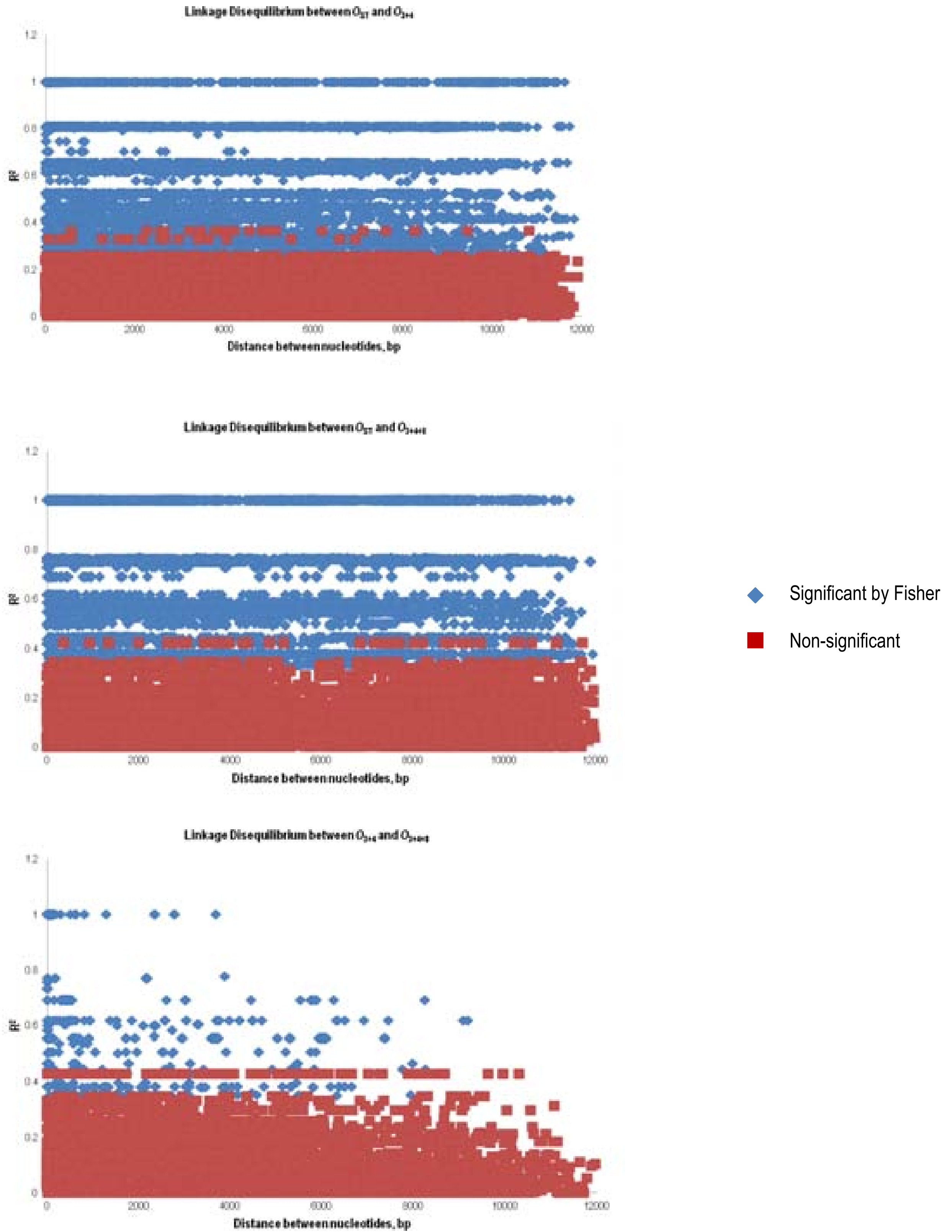


Figure 4.15: Linkage disequilibrium between pairs of informative positions, assessed by R^2 and identified by color according to its significance by Fisher's exact test, pairwise grouping the three chromosomal arrangements.

within each arrangement) indicate a strong genetic differentiation between them and thus a strong reduction of recombination in heterokaryotypes. As in the present study the highest genetic differentiation was found in the samples including the O_{ST} lines and the lowest differentiation in the sample combining $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ lines (Navarro-Sabaté, Aguadé and Segarra 2003).

4.3.5. Gene flux

Gene flux, defined as the probability of allele exchange during meiosis in heterokaryotypic females including both crossover and gene conversion, is an important factor decreasing genetic differentiation between inversions (Navarro *et al.* 1997). Since the unique origin of most inversions has been widely accepted (Powell 1997), during the first stages genetic variability within the inversion should be low and genetic differentiation among arrangements should be high. Through time, variability can be recovered by two non-excluding mechanisms, mutation and recombination, and the latter, which includes crossovers and gene conversion, seems to occur at a higher rate than mutation (Schaeffer and Anderson 2005). However, selective pressure can maintain genetic differentiation among arrangements (Hoffmann, Sgro and Weeks 2004). Therefore, the observation of strong genetic differentiation for genes located across an inverted region, as it is in the present work, was interpreted in two possibilities: the lack of double crossovers or selection action against recombinant individuals (Munté *et al.* 2005; Schaeffer and Anderson 2005). Non-selective factors, such as the length and the age of inversions, can also contribute to the maintenance of the genetic differentiation among arrangements.

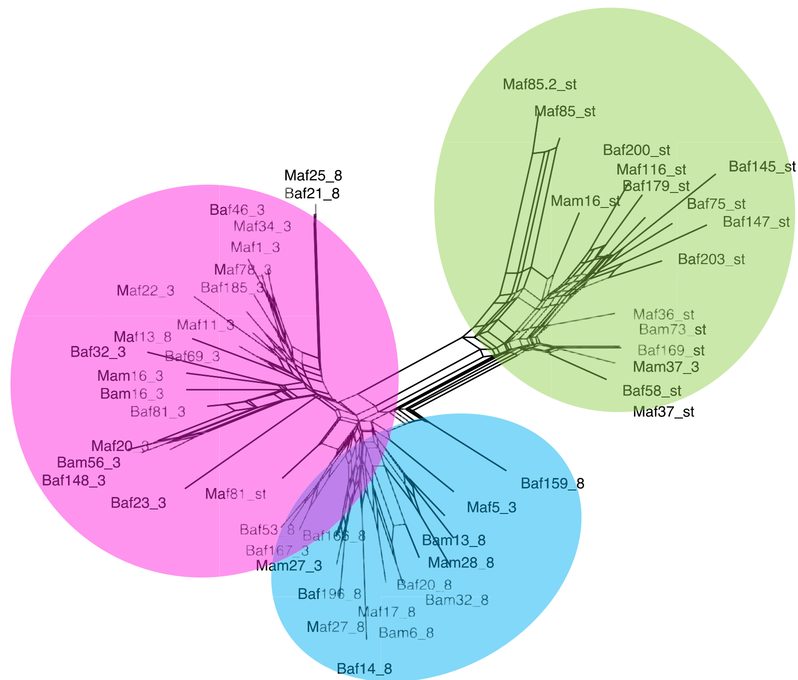
The level of recombination inside the inversion depends on the size of the inversion. According to Navarro *et al.* (1997), the minimum length for double crossover inside an inversion is 20 cM. Thus, the longer the inversion, the higher would be the probability of double crossovers occurrence. Taking into account that the approximate length of the O_{ST} inversion relative to $O_{\underline{3+4}}$ is more than 40 cM (Pegueroles *et al.* 2010a) it is much higher than the theoretical minimum length and therefore could often give double overlaps. The study of eight genes in the *SI* region, located within the O_3 inversion, revealed high levels of genetic differentiation between the $O_{\underline{3+4}}$ and the O_{ST} arrangements, as well as lower genetic variability in the latter (Munté *et al.* 2005). Our data do not confirm the later result, demonstrating the absence of significant difference in the nucleotide diversity between the two arrangements, which were also very well differentiated in our study.

Nucleotide diversity for all arrangements in the concatenated data set was similar and equal to 0.007-0.008. They estimated the genetic length of O_3 inversion in 27.4 cM, considering that the genetic length of the *O* chromosome of *D. subobscura* is 228.3 cM (LOUKAS, KRIMBAS and VERGINI 1979). However, with the newer estimation of its genetic length (190.7 cM; PEGUEROLES *et al*. 2010a), the length of O_3 inversion was estimated as 22.9 cM, being very close to the minimum size necessary to allow double crossovers formation. Furthermore, it is worth pointing out that O_3 inversion although ancestral is never found alone in natural populations, and it is always encountered forming a complex of two overlapped inversions. Following PEGUEROLES *et al.* (2010a) we calculated the length of O_4 and O_8 inversions, considering the average length of a chromosomal band equal to 82 Kb and weighted mean recombination rate in chromosome *O* of *D. subobscura* of 7cM/Mb. Thus we found the length of O_4 and O_8 inversions approximately equal to 53 and 55 cM respectively, being quite similar and higher than the minimum genetic length required for the formation of even number of crossovers. Thus, both genetic differentiation and variability could be influenced by the presence of overlapped inversions.

Recombination networks were calculated using all sequences separately for each gene region and the concatenated data set (Figure 4.16). Evolutionary relationships are usually represented using phylogenetic trees, based on a model of evolution dominated by mutations and speciation events. More realistic models must also account for gene genesis, loss and duplication events, hybridization, horizontal gene transfer or recombination. One of the programs that offer such phylogenetic network reconstruction is Splits Tree (HUSON and BRYANT 2006), which combines both phylogenetic tree and networks.

For all genomic regions individuals carrying O_{ST} arrangement were well distinguished from the rest of the lines, which is consistent with the strong genetic differentiation detected between arrangements, although there were few recombinant individuals of another arrangement clustering with them in almost each case. Individuals with $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements in all networks, except of the reg_*larp* and *Acph-1*, clustered together, showing great number of recombination events between them, consistently with the previous studies on *rp49* region for $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements (ROZAS and AGUADÉ 1993; ROZAS *et al.* 1999).

Acph-1



Ast

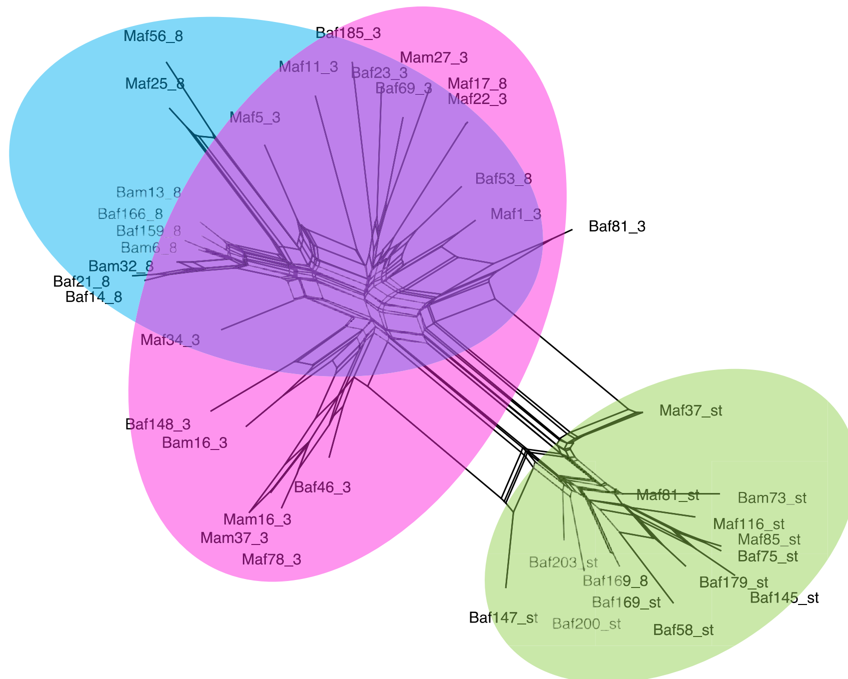
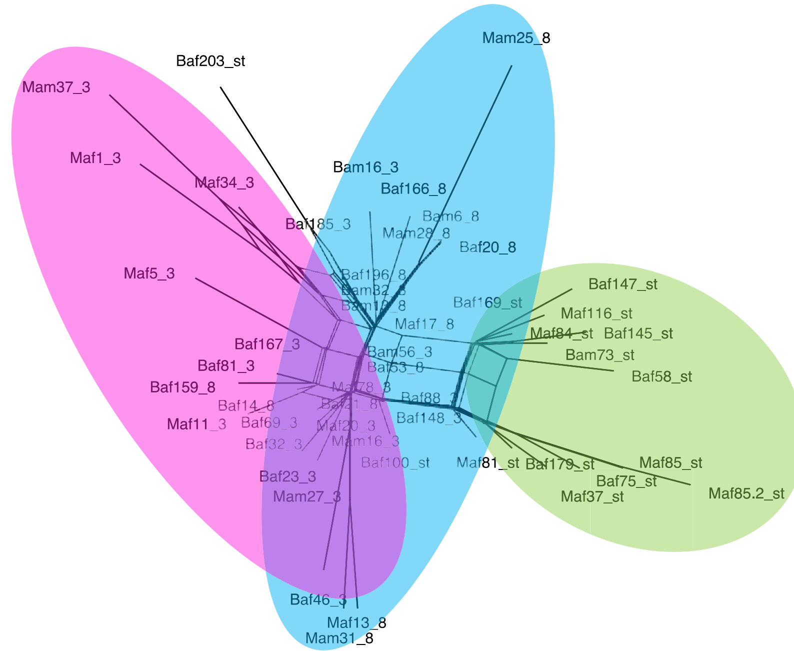


Figure 4.16: Recombination networks for the genes *Acph-1* and *Ast*. The green area represents the cluster of O_{ST} arrangement; the blue area corresponds to O_{3+4+8} and the pink to O_{3+4}. The names of individual lines from population of Barcelona start with "Ba" and from Málaga with "Ma". The lines that have "_st" at the end of the name carried O_{ST} arrangement; the lines with "_3" correspond to O_{3+4} and "_8" to O_{3+4+8}.

larp

0.0010



reg_*larp*

0.0010

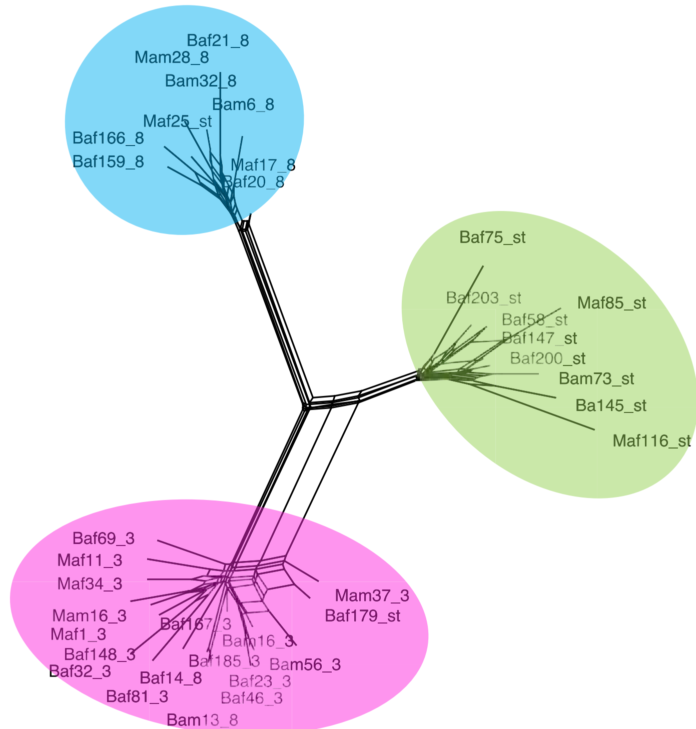
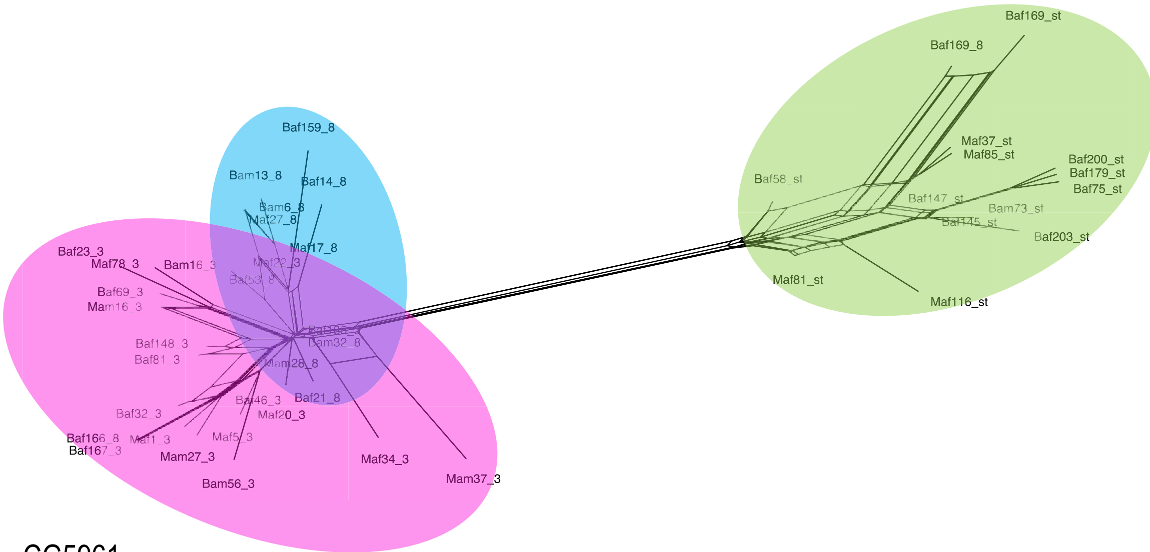


Figure 4.16 (continued): Recombination networks for the gene *larp* and its regulatory region. The green area represents the cluster of O_{ST} arrangement; the blue area corresponds to $O_{\underline{3+4+8}}$ and the pink to $O_{\underline{3+4}}$. The names of individual lines from population of Barcelona start with "Ba" and from Málaga with "Ma". The lines that have "_st" at the end of the name carried O_{ST} arrangement; the lines with "_3" correspond to $O_{\underline{3+4}}$ and "_8" to $O_{\underline{3+4+8}}$.

trus

0.0010



CG5961

0.0010

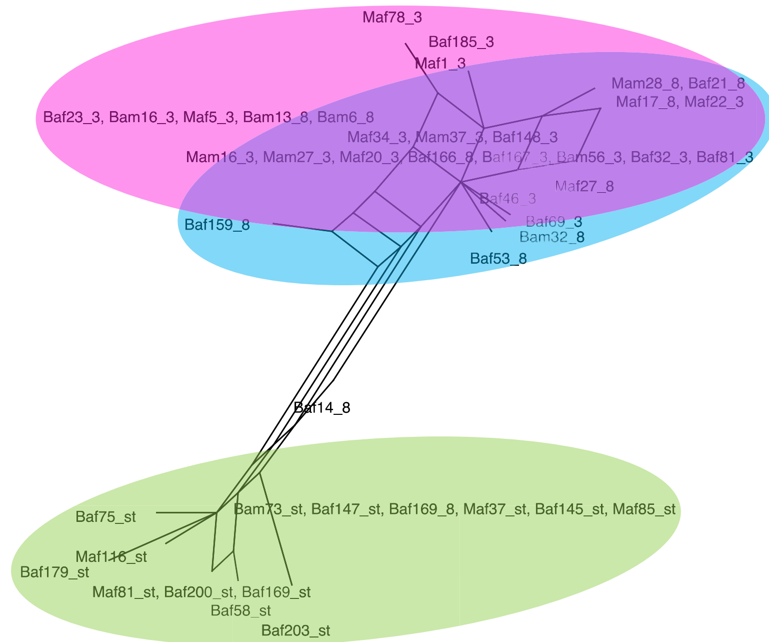
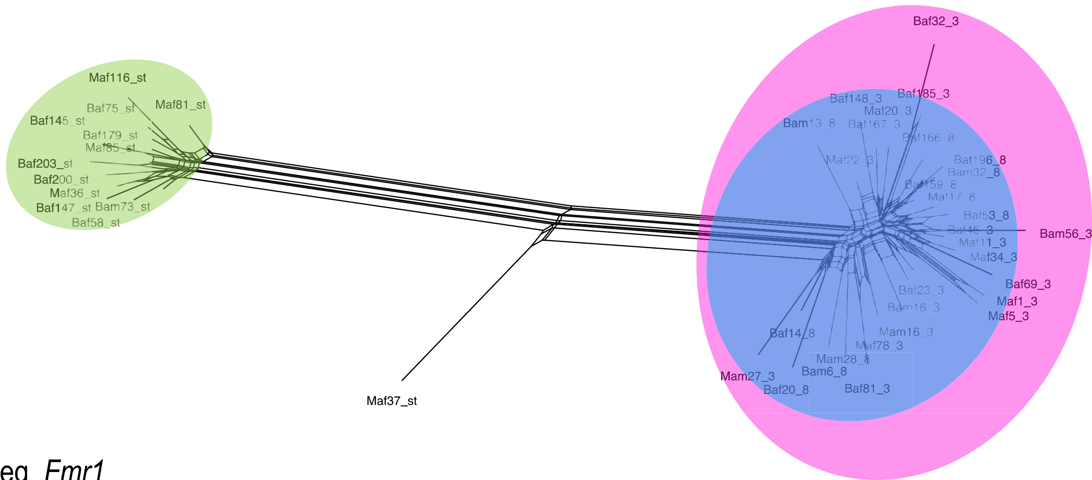


Figure 4.16 (continued): Recombination networks for the genes *trus* and CG5961. The green area represents the cluster of O_{ST} arrangement; the blue area corresponds to $O_{\underline{3+4+8}}$ and the pink to $O_{\underline{3+4}}$. The names of individual lines from population of Barcelona start with "Ba" and from Málaga with "Ma". The lines that have "_st" at the end of the name carried O_{ST} arrangement; the lines with "_3" correspond to $O_{\underline{3+4}}$ and "_8" to $O_{\underline{3+4+8}}$.

Fmr1

0.0010



reg_*Fmr1*

0.0010

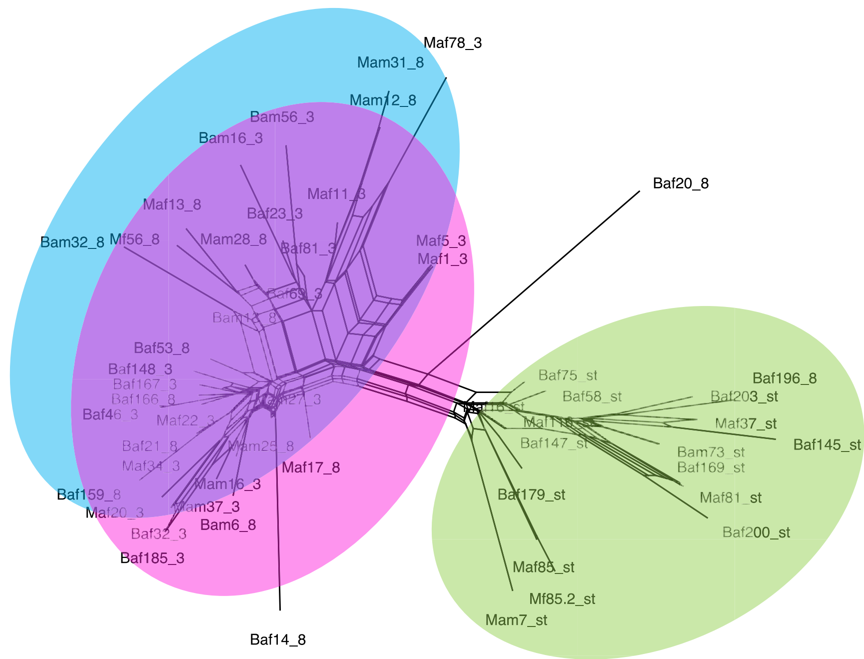


Figure 4.16 (continued): Recombination networks for the gene *Fmr1* and its regulatory region. The green area represents the cluster of O_{ST} arrangement; the blue area corresponds to $O_{\underline{3+4+8}}$ and the pink to $O_{\underline{3+4}}$. The names of individual lines from population of Barcelona start with “Ba” and from Málaga with “Ma”. The lines that have “_st” at the end of the name carried O_{ST} arrangement; the lines with “_3” correspond to $O_{\underline{3+4}}$ and “_8” to $O_{\underline{3+4+8}}$.

Concatenated dataset

0.0010

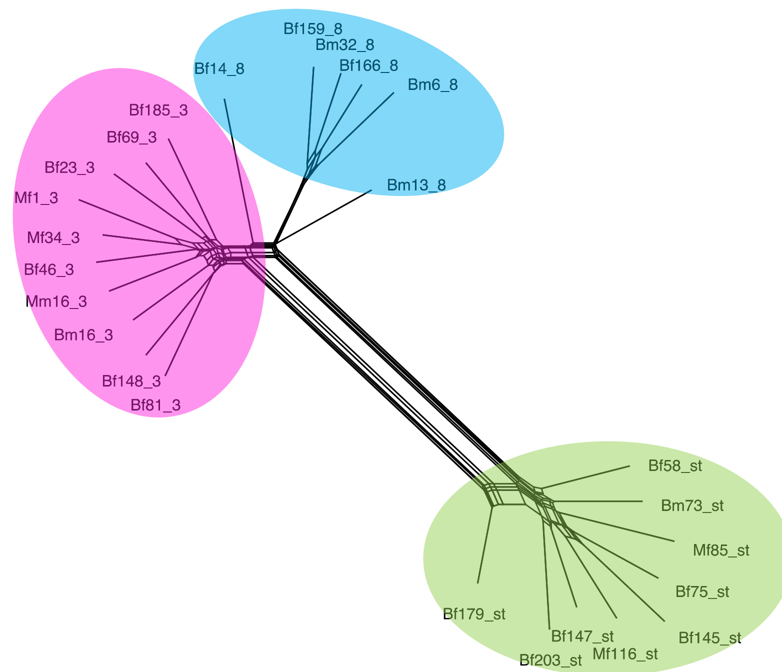


Figure 4.16 (continued): **Recombination network for the concatenated dataset.** The green area represents the cluster of O_{ST} arrangement; the blue area corresponds to $O_{\underline{3+4+8}}$ and the pink to $O_{\underline{3+4}}$. The names of individual lines from population of Barcelona start with "Ba" and from Málaga with "Ma". The lines that have "_st" at the end of the name carried O_{ST} arrangement; the lines with "_3" correspond to $O_{\underline{3+4}}$ and "_8" to $O_{\underline{3+4+8}}$.

However, in the regulatory region of *larp* and gene *Acph-1* the individuals of all three arrangements were very well distinguished as well as in the network of concatenated dataset, where all individuals are well grouped in separate cluster according to their arrangement. For most genes the reticulated network showing connections between arrangements and lines suggest high levels of recombination among them. Under the assumption that inversions are monophyletic, the topology of the trees would indicate that gene flux is higher among $O_{\underline{3+4}+X}$ sequences than between $O_{\underline{3+4}+X}$ and O_{ST} sequences. None of the network trees was able to distinguish between the two populations, whose sequences were randomly clustered in all genomic regions according to their chromosomal arrangements.

Gene conversion tracts (GCT) between arrangements were estimated with the methodology of BETRÀN *et al*. (1997) implemented in DnaSP software (Table 4.9) and detected fragments are identified in the Supplemental table S2 corresponding to the polymorphic sites in each gene region. The method depends mainly on two parameters: *Ψ*, that is a probability of a site being informative

for a gene conversion tract between two gene arrangements and ϕ, the parameter of the geometric distribution, from which the average true tract length can be estimated. The algorithm detects the observed number of GCT and its observed length. The Ψ parameter is related to the extent of genetic differentiation among sequences. So, the more different are the analyzed sequences, the higher the value of Ψ and, therefore, more accurate are the number and size of GCT detected. The obtained values of Ψ in our study ranged between 0.0004 and 0.028, with larger estimates for the regulatory region of *larp* and gene *Fmr1*, which are the genomic regions presented higher genetic differentiation. On average the values of Ψ in our study were similar to those reported in BETRÀN *et al.* (1997) and ROZAS *et al.* (1999) for the gene *rp49* comparing the same gene arrangements, where they ranged from 10^{-3} to 10^{-2}. The values of Ψ were higher between O_{ST} and $O_{\underline{3+4}+X}$ sequences than among $O_{\underline{3+4}+X}$ lines (Table 4.9) in agreement with values of Ψ previous studies for different genes (ROZAS *et al.* 1999, PEGUEROLES *et al.* 2010b, CALABRIA 2012). The estimations from other species such as *D. pseudoobscura* (values ranging from 0.002 to 0.059; SCHAEFFER and ANDERSON 2005) and D. *buzzatti* (values ranging from 0.004 to 0.015; LAAYOUNI *et al.* 2003) were comparable to those obtained by us. The length of the tracts was quite variable from 2 to 1543 bp. The largest tracts were found in the genes *Ast* and *Acph-1* and regulatory region of *larp*. Although no distinction can be made between gene conversion and double crossover events if the tract includes the outermost informative nucleotides, genetic exchange by double crossover is more plausible explanation since it is expected to affect longer chromosomal regions than gene conversion, which are expected to be small (i.e. 122 bp tract length on average in *D. subobscura*; BETRÀN *et al.* 1997). So, very large GCTs (around or more than 500bp) were considered as products of double crossovers. Therefore, 24 out of 78 recombination events found by us were considered as a result of even number of crossovers (Table 4.8). So, as it was established, gene conversion is the dominant force at intragenic level that breaks the association between the positions of different inversions (ANDOLFATTO and NORDBORG 1998), as it was found in the other species in studies conducted in the *rosy* locus of *D. melanogaster* (CHOVNIC 1973), *rp49* locus of *D. subobscura* (ROZAS and AGUADÉ 1994) and locus of *amylase* in *D. pseudoobscur*a (POPADIC, POPADIC and ANDERSON 1995) among others.

As expected, some amount of recombination was detected for all analyzed genes, although not for all arrangements (Figure 4.17: a, b, c). In total, 60% (36) of recombinant individuals were detected with tracts that might be explained by gene conversion and double crossovers. The population, in

which more recombination events were found, was Barcelona (70%; 22 individuals), but in Málaga they were also abundant (48.3%; 14 individuals). The number of recombinant lines was the highest for the $O_{\underline{3+4}}$ arrangement (15; 75%), followed by $O_{\underline{3+4+8}}$ (12; 63.2%) and O_{ST} (9; 42.9%). There were more recombination events between gene arrangements $O_{ST}/O_{\underline{3+4}}$ (29) and $O_{ST}/O_{\underline{3+4+8}}$ (26) than between O_{3+4} and $O_{\underline{3+4+8}}$ (21) (Figure 4.17: a, b and c). The percentages of recombinant lines per gene region, arrangement and population are shown in the Figure 4.17 (d).

Theoretical studies point out that recombination will be more efficiently inhibited near inversion breakpoints, where gene flux would be primarily due to gene conversion, whereas in the central part of the inversion both gene conversion and double crossovers could contribute to gene flux (NAVARRO *et al.* 1997). Thus, recombination reduction mediated by inversions could play an important role in the maintenance of genetic differentiation between O_{ST} and $O_{\underline{3+4}+X}$ arrangements.

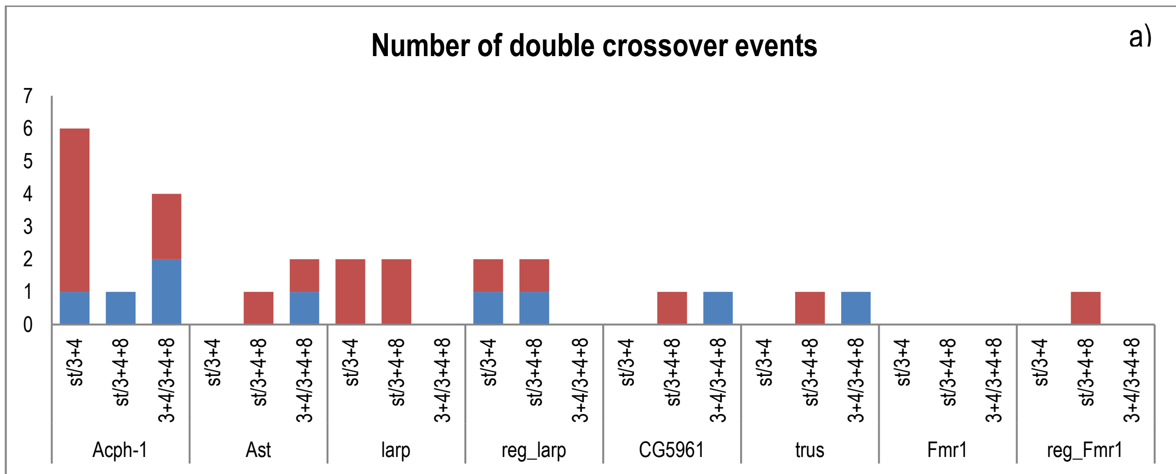
Larger numbers of recombination events (29 in total) and recombinant lines were found in the gene *Acph-1* especially in the arrangement O_{3+4}, where the proportion of recombinants in populations of Málaga and Barcelona reached 44% and 60% respectively (Figure 4.17 and Table 4.9), which is consistent with high variability and strong genetic divergence between gene arrangements in this gene, but contradict the hypothesis of NAVARRO *et al.* (1997) that the shorter the distance to the nearest inversion break point, less recombination events will be found given its linkage to the O_3 inversion break point. The smallest numbers of recombination events were found for the genes CG5961 (2) and *trus* (4), which are located close to the O_8 inversion break point. The coefficients of correlation between numbers of double crossovers and gene conversion tracts, separately for each kind of event and altogether, and the distance to the nearest inversion breakpoint were calculated (Table 4.10) and regression graphs for each pair of arrangements are represented in the Figure 4.18.

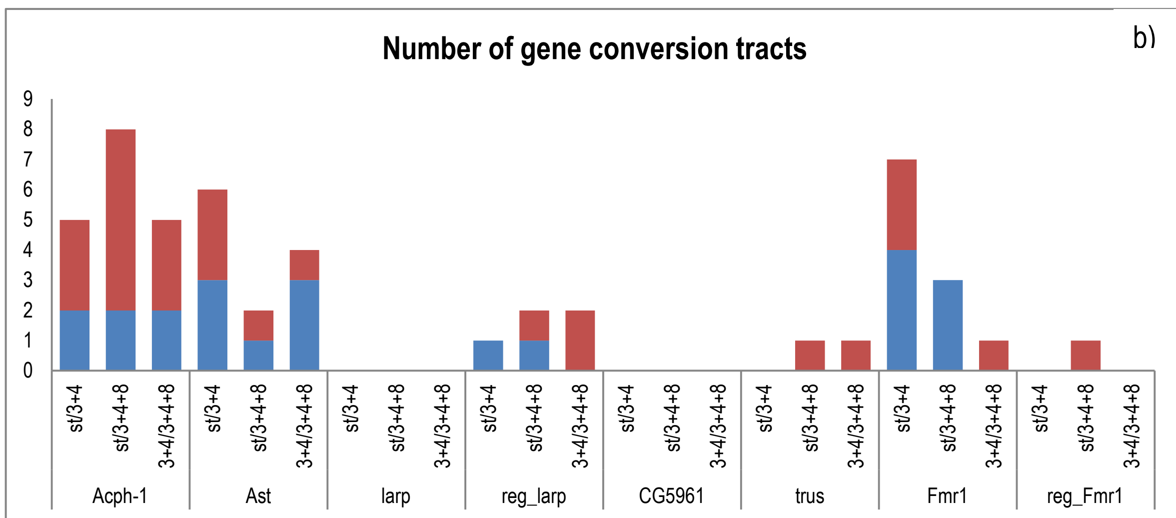
In all cases the correlations were non-significant and negative for the arrangements $O_{ST}/O_{\underline{3+4}}$ and $O_{ST}/O_{\underline{3+4+8}}$, while for the $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ all of them were positive due to the most central location of *Acph-1* when comparing these two last chromosomal arrangements. After excluding the gene *Acph-1* from the analysis, the coefficients continued being non-significant, but switched their sign in all cases, except in $O_{ST}/O_{\underline{3+4+8}}$ for recombination events, so that the number of double crossover and GC events for $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ showed negative correlation with the distance, while the correlation for $O_{ST}/O_{\underline{3+4}}$ and $O_{ST}/O_{\underline{3+4+8}}$ changed to be positive.

Table 4.9: Gene conversion tracts detected between each pair of gene arrangements per each population.

Region	Between arrangements	Name of the line	Population	Gene arrangement	ψ	Informative sites	Location	Length (bp)	Double crossover?
Acph-1	O_{ST}/O_{3+4}	Baf200	Barcelona	O_{ST}	0.012	54	135-620	486	+
		Maf81	Málaga	O_{ST}	0.012	54	70-1438	1369	+
		Maf85.2	Málaga	O_{ST}	0.012	54	70-698	629	+
		Bam56	Barcelona	O_{3+4}	0.012	54	758-782	25	
		Bam16	Barcelona	O_{3+4}	0.012	54	70-111	42	
		Baf23	Barcelona	O_{3+4}	0.012	54	884-1438	555	+
		Baf148	Barcelona	O_{3+4}	0.012	54	758-782	25	
		Baf167	Barcelona	O_{3+4}	0.012	54	601-1220	620	+
		Mam27	Málaga	O_{3+4}	0.012	54	601-1220	620	+
		Maf5	Málaga	O_{3+4}	0.012	54	601-620	20	
		Maf20	Málaga	O_{3+4}	0.012	54	758-782	25	
	O_{ST}/O_{3+4+8}	Baf20	Barcelona	O_{3+4+8}	0.011	50	111-135	25	
		Baf20	Barcelona	O_{3+4+8}	0.011	50	479-485	7	
		Baf14	Barcelona	O_{3+4+8}	0.011	50	743-809	67	
		Baf53	Barcelona	O_{3+4+8}	0.011	50	479-485	7	
		Baf159	Barcelona	O_{3+4+8}	0.011	50	111-485	375	
		Baf159	Barcelona	O_{3+4+8}	0.011	50	1262-1265	4	
		Maf81	Málaga	O_{ST}	0.011	50	70-1438	1369	+
		Maf85.2	Málaga	O_{ST}	0.011	50	70-135	66	
		Maf13	Málaga	O_{3+4+8}	0.011	50	797-809	13	
		Maf13	Málaga	O_{3+4+8}	0.011	50	986-1438	453	+
	O_{3+4}/O_{3+4+8}	Baf32	Barcelona	O_{3+4}	0.004	29	25-26	2	
		Baf167	Barcelona	O_{3+4}	0.004	29	593-1220	628	+
		Baf14	Barcelona	O_{3+4+8}	0.004	29	797-809	13	
		Baf21	Barcelona	O_{3+4+8}	0.004	29	593-620	28	
		Mam27	Málaga	O_{3+4}	0.004	29	355-1220	866	+
		Maf5	Málaga	O_{3+4}	0.004	29	593-620	28	
		Maf13	Málaga	O_{3+4+8}	0.004	29	593-1711	1119	+
		Maf25.1	Málaga	O_{3+4+8}	0.004	29	593-620	28	
Ast	O_{ST}/O_{3+4}	Baf147	Barcelona	O_{ST}	0.009	40	661-1081	419	
		Baf69	Barcelona	O_{3+4}	0.009	40	679-682	4	
		Baf81	Barcelona	O_{3+4}	0.009	40	582-682	101	
		Maf37	Málaga	O_{ST}	0.009	40	781-1081	299	
		Mam27	Málaga	O_{3+4}	0.009	40	679-682	4	
		Maf34	Málaga	O_{3+4}	0.009	40	1522-1583	62	
	O_{ST}/O_{3+4+8}	Baf147	Barcelona	O_{ST}	0.009	28	775-781	7	
		Baf196	Barcelona	O_{3+4+8}	0.009	28	39-1583	1534	+
		Maf37	Málaga	O_{ST}	0.009	28	781-923	143	
	O_{3+4}/O_{3+4+8}	Baf196	Barcelona	O_{3+4+8}	0.002	19	841-923	83	
		Baf53	Barcelona	O_{3+4+8}	0.002	19	808-1344	521	+
		Maf5	Málaga	O_{3+4}	0.002	19	882-900	19	
		Maf34	Málaga	O_{3+4}	0.002	19	495-900	406	
		Maf17	Málaga	O_{3+4+8}	0.002	19	895-1081	187	
larp	O_{ST}/O_{3+4}	Baf100	Barcelona	O_{ST}	0.002	7	1301-1898	580	+
		Baf203	Barcelona	O_{ST}	0.002	7	1301-1898	577	+
	O_{ST}/O_{3+4+8}	Baf100	Barcelona	O_{ST}	0.002	9	1301-1898	580	+
		Baf203	Barcelona	O_{ST}	0.002	9	1301-1898	577	+
reg_larp	O_{ST}/O_{3+4}	Baf179	Barcelona	O_{ST}	0.015	42	922-1511	582	+
		Baf69	Barcelona	O_{3+4}	0.015	42	1694-1697	4	
		Maf25	Málaga	O_{ST}	0.015	42	1078-1726	633	+
		Mam37	Málaga	O_{3+4}	0.015	42	1293-1726	419	
	O_{ST}/O_{3+4+8}	Baf179	Barcelona	O_{ST}	0.020	46	922-1511	582	+
		Bam13	Barcelona	O_{3+4+8}	0.020	46	1041-1074	34	
		Baf14	Barcelona	O_{3+4+8}	0.020	46	1041-1074	34	
		Maf25	Málaga	O_{ST}	0.020	46	193-1726	1426	+
	O_{3+4}/O_{3+4+8}	Bam13	Barcelona	O_{3+4+8}	0.020	40	790-1074	280	
		Baf14	Barcelona	O_{3+4+8}	0.020	40	929-1074	145	
CG5961	O_{ST}/O_{3+4+8}	Baf14	Barcelona	O_{3+4+8}	0.013	11	58-142	85	
		Baf196	Barcelona	O_{3+4+8}	0.013	11	34-299	266	
		Baf159	Barcelona	O_{3+4+8}	0.013	11	298-299	2	
	O_{3+4}/O_{3+4+8}	Maf22	Málaga	O_{3+4}	0.001	9	32-406	375	
trus	O_{ST}/O_{3+4+8}	Baf196	Barcelona	O_{3+4+8}	0.009	21	173-335	163	
		Baf196	Barcelona	O_{3+4+8}	0.009	21	425-1372	942	+
	O_{3+4}/O_{3+4+8}	Baf166	Barcelona	O_{3+4+8}	0.001	4	28-190	163	
		Maf22	Málaga	O_{3+4}	0.001	4	608-1121	514	+
Fmr1	O_{ST}/O_{3+4}	Baf32	Barcelona	O_{3+4}	0.028	70	2033-2039	7	
		Baf81	Barcelona	O_{3+4}	0.028	70	286-289	4	
		Baf185	Barcelona	O_{3+4}	0.028	70	2033-2039	7	
		Maf37	Málaga	O_{ST}	0.028	70	810-831	22	
		Maf37	Málaga	O_{ST}	0.028	70	843-851	7	
		Maf37	Málaga	O_{ST}	0.028	70	856-910	55	
		Mam27	Málaga	O_{3+4}	0.028	70	2033-2039	7	
	O_{ST}/O_{3+4+8}	Maf37	Málaga	O_{ST}	0.028	71	810-831	22	
		Maf37	Málaga	O_{ST}	0.028	71	843-910	66	
		Mam28	Málaga	O_{3+4+8}	0.028	71	851-852	2	
	O_{3+4}/O_{3+4+8}	Baf81	Barcelona	O_{3+4}	0.0004	5	286-289	4	
reg_Fmr1	O_{ST}/O_{3+4+8}	Baf20	Barcelona	O_{3+4+8}	0.006	15	545-773	228	
		Baf196	Barcelona	O_{3+4+8}	0.006	15	189-811	620	+

‘Gene arrangement’ indicates the arrangement of an individual where GCT was found. *Ψ* is a probability for the position being informative for a gene conversion event. ‘+’ in the column ‘Double crossover?’ indicates that a fragment is too large (around 500 bp or more) to be GCT and could be considered the product of a double crossover.





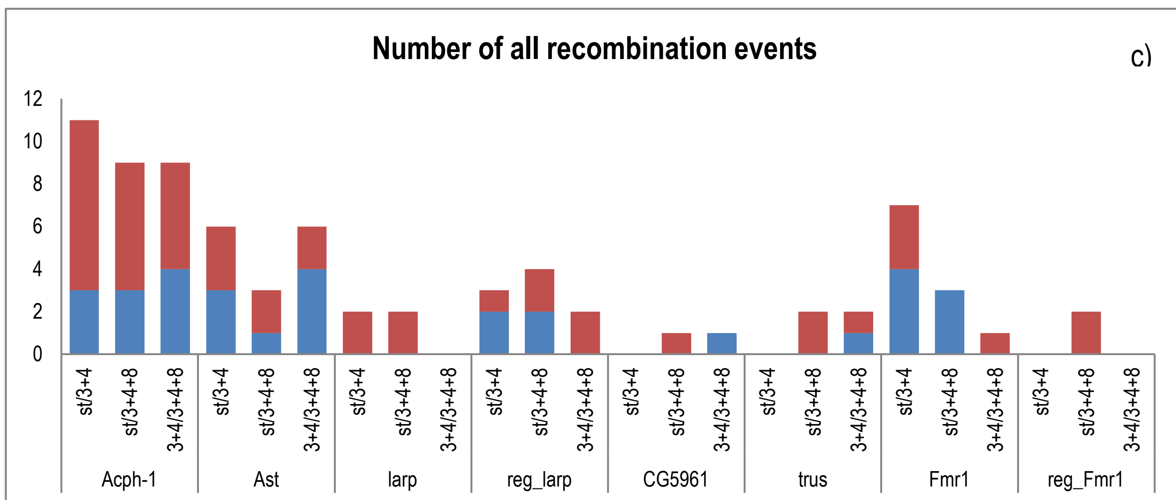


Figure 4.17: **The number of double crossover products (a), gene conversion events (b) and both types of events (c) in each gene region, pair of gene arrangements and population.** Blue color corresponds to the population of Barcelona and red to the population of Málaga.

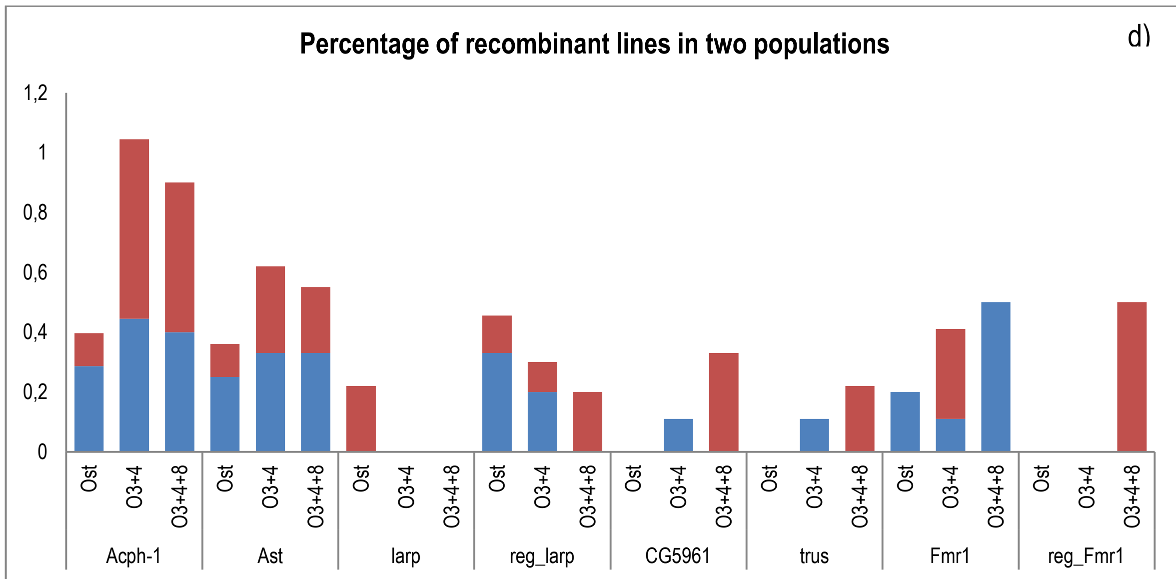
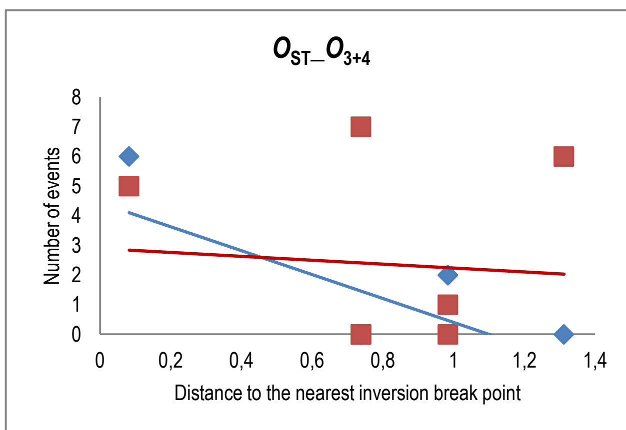
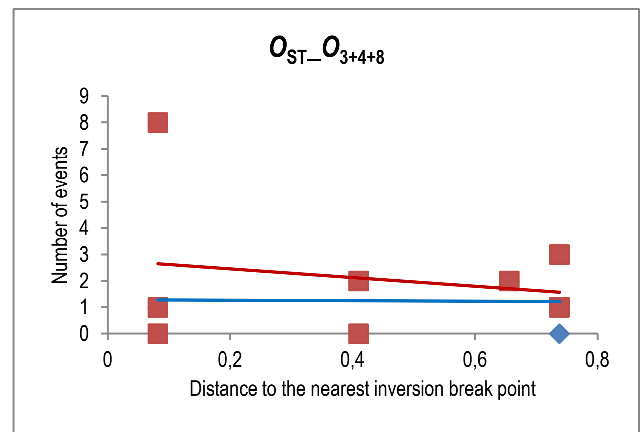


Figure 4.17 (Continued): d) the percentage of recombinant lines (presented both with double crossovers and GCT) per population, gene arrangement and gene region. Blue color corresponds to the population of Barcelona and red to the population of Málaga.





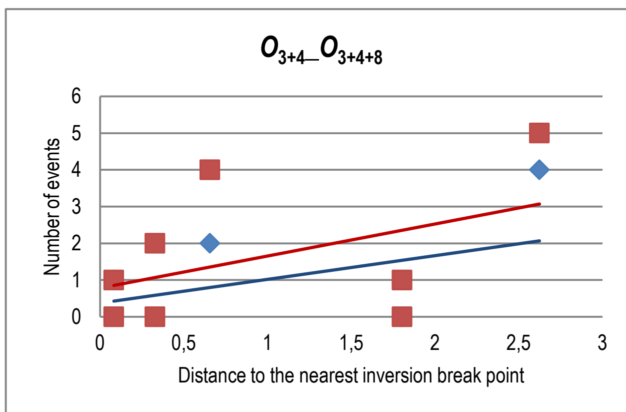


Figure 4.18: Correlation between number of recombination events and distance to the nearest inversion break point for each gene region and pair of gene arrangements. Red squares and tendency lines correspond to gene conversion tracts; blue diamonds and tendency lines correspond to double crossover products.

Table 4.10: Correlation coefficients between recombination and gene conversion events per pair of gene arrangements and the distance of gene regions to the nearest inversion break point.

		All gene regions		*Acph-1* excluded	
		β	P	β	P
Double crossover events	O_{ST}/O_{3+4}	-0.666	0.071	0.292	0.525
	O_{ST}/O_{3+4+8}	-0.042	0.920	-0.118	0.800
	O_{3+4}/O_{3+4+8}	0.443	0.272	-0.398	0.377
Gene conversion tracts	O_{ST}/O_{3+4}	-0.075	0.860	0.381	0.399
	O_{ST}/O_{3+4+8}	-0.184	0.663	0.613	0.143
	O_{3+4}/O_{3+4+8}	0.440	0.275	-0.096	0.837
All events	O_{ST}/O_{3+4}	-0.407	0.317	0.501	0.252
	O_{ST}/O_{3+4+8}	-0.202	0.631	0.522	0.229
	O_{3+4}/O_{3+4+8}	0.459	0.252	-0.221	0.635

It has been previously suggested that recombination should be higher in central parts of the inversion because genes located in this region could be influenced by double crossovers and gene conversion, while genes located close to inversions breakpoints should be mainly influenced by gene conversion (NAVARRO *et al.* 1997).

Nevertheless double recombinants have been observed close to the break points when analyzing the offspring of heterokaryotypic females indicating that double crossovers close to the break point can occur although in very low frequency (PEGUEROLES *et al.* 2010a). Since recombination between heterokaryotypes is reduced inside inverted regions (NAVARRO *et al.* 1997; STUMP *et al.* 2007; PEGUEROLES *et al*. 2010a), higher diversity levels are expected outside inverted regions than inside them.

In the region where genes can freely recombine between both arrangements (*SII* region), estimates of π_{sil} ranged from 0.004 to 0.024 for the O_{ST} arrangement and from 0.006 to 0.023 for $O_{\underline{3+4}}$ combining data of CALABRIA (2012) and PEGUEROLES *et al.* (2013) and after excluding genes such as *Yrt* that can act as an outlier due to its high nucleotide variability. The π_{sil} values for 11 genes located inside the inverted region between the two arrangements (*SI* region), combining results of a previous study (MUNTÉ *et al.* 2005) and the present work, ranged from 0.004 to 0.018 for the O_{ST} arrangement and from 0.005 to 0.013 for the $O_{\underline{3+4}}$ arrangement, after excluding the *Acph1* gene, which also could act as a mutation outlier. The ranges of nucleotide variability mostly overlap for genes located inside and outside the inverted regions, and non-significant differences were detected when performing a Mann-Whitney U Test (O_{ST}: P = 0.142, $O_{\underline{3+4}}$: P = 0.457), although the highest values are found outside the inversions. Reduced diversity within inversions has been found in the *D. persimilis* lineage, indicative of a recent fixation process despite inversions may

have arisen long ago (McGAUGH and NOOR 2012). Our results suggest that despite the studied inversions are maintained by selection (PREVOSTI *et al.* 1988; BALANYÀ *et al.* 2006); recombination could also contribute in some extent to recover variability inside reasonably old polymorphic inversions.

In all studies on the nuclear genes that have been made on the *SI*, it has been observed that the genetic differentiation between arrangements O_{ST} and $O_{\underline{3+4}}$ is high and significant (ROZAS and AGUADÉ 1994; NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a; MUNTÉ *et al.* 2005; PEGUEROLES *et al.* 2013 and the present study). The high levels of genetic differentiation detected between these arrangements may be influenced by the inhibition of recombination due to the complex of two overlapping inversions. The analyses made in nuclear genes located inside of medium-sized inversion, also show significant genetic differentiation (ROZAS *et al.* 1999; NOBREGA *et al.* 2008; the present study). However, the values of genetic differentiation are lower in the simple than in the overlapping inversions (NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a; ROZAS et al. 1999; MUNTÉ *et al.* 2005, PEGUEROLES *et al.* 2010b; the present study).

Our results confirm the previous studies and show that genetic differentiation between O_{ST} and $O_{\underline{3+4}+X}$ chromosomes is strong and extends homogeneously all over the inversion. Therefore, genetic exchange between arrangements has been strongly suppressed even in the central part of the inversion loop. The strong differentiation detected might be explained either by the absence of double crossovers in the O_3 inversion loop or by the elimination of double crossover products by natural selection.

Most of the inversions described in *D. subobscura* form complexes with the arrangement $O_{\underline{3+4}}$, which is more common in populations of southern Europe (KRIMBAS 1993; SOLÉ *et al.* 2002; ARAÚZ *et al.* 2009). Therefore, in natural populations it is expected to find a large number of individuals who are heterozygous for $O_{\underline{3+4}}$ and other arrangement (such as $O_{\underline{3+4+8}}$, $O_{\underline{3+4+23}}$, $O_{\underline{3+4}+1}$ and $O_{\underline{3+4}+7}$), so that the two chromosomes of individuals differ only by one inversion and may have higher frequency of recombination. In contrast, the gene flow between O_{ST} and other arrangements is expected to be very low, because their presence in heterozygous individuals imply the formation of two inversion loops at minimum, excepting when the heterozygote is formed by simple inversions such as O_6, O_5 and O_{11}, which are usually found at low frequency and restricted to northern

Europe, or with new arrangements formed by recombination, which are also found in very low frequency.

4.3.6. TESTS FOR NEUTRALITY AND FOOTPRINTS OF SELECTION

After excluding recombinant individuals Tajima's *D* test presented non-significant values for all genes, populations and arrangements when the populations were considered separately. The values were negative for the majority of genes in both populations (Supplementary Table S5a, b). A similar result was obtained with Fu and Li's *D* statistic with mostly negative non-significant values in the majority of cases with several exceptions in the population of Barcelona: significant negative values corresponded to coding positions of *larp* in $O_{\underline{3+4}}$ arrangement in both populations, and CG5961 in O_{ST} in the population of Barcelona. The positive significant values of this test were detected in the exonic part of *trus* in $O_{\underline{3+4+8}}$ in the population of Málaga, but this result could be altered by small number of sequenced lines. After pooling the two populations together significant negative values of Tajima's *D* were obtained for the same genes: coding sequences of *larp* in all arrangements, due to synonymous substitutions in $O_{\underline{3+4}}$ and nonsynonymous substitutions in O_{ST} and $O_{\underline{3+4+8}}$ arrangements and intronic part of *trus* in $O_{\underline{3+4}}$ (Table 4.11). Almost the same picture was observed with Fu and Li's *D* statistic: significant negative values were obtained for the coding regions of *larp* in all arrangements, whole sequence of *trus* in $O_{\underline{3+4}}$ and CG5961 in O_{ST}. The significant Fu and Li´s *D* value for gene *trus* in $O_{\underline{3+4}}$ was already reported by MUNTÉ *et al.* (2005) for El Pedroso population (northwestern Spain) although in that study it was included in the gene region P154, which combines two different genes (*trus* and CG5961) and they didn´t distinguish between them. However in that study they failed to detect significant *D* for the same region in O_{ST}, which we have detected for gene CG5961.

Negative *D* statistic values correspond to an excess of low frequency polymorphism, which could be due to either demography or selection. To further elucidate this issue, the pairwise nucleotide difference distribution, or mismatch distribution (SLATKIN and HUDSON 1991; ROGERS and HARPENDING 1992) and the R_2 statistic (RAMOS-ONSINS and ROZAS 2002), which is a test for recent population expansion, was estimated because it is more suitable for small sample sizes presenting recombination. R_2 estimates ranged between 0.061 and 0.266 and its significance was assessed considering the rate of recombination (ρ) estimated for each gene using the composite likelihood

method of HUDSON (2001) since rejection of the null hypothesis of constant population size depends on the level of recombination implemented. It was found that the significant deviation from neutrality due to an excess of low frequency polymorphism for the genes *larp* and *trus* in $O_{\underline{3+4}}$ arrangement in Barcelona and pooled populations and for the regulatory region of *Fmr1* in $O_{\underline{3+4+8}}$ in pooled populations could be due to population expansion as the R_2 values in these cases are significant. In the rest of the cases with significant deviation from neutral equilibrium the null hypothesis of constant size could not be rejected.

Figure 4.19 shows two situations: when R_2 value is significant and distribution of pairwise differences coincides with expectation of population growth, giving as an example gene *Fmr1* in $O_{\underline{3+4}}$ arrangement, and when this value is non-significant and two curves do not coincide as in the case of O_{ST} and $O_{\underline{3+4+8}}$ arrangements for the same gene. The same gene in two arrangements O_{ST} and O_{3+4} was studied by PEGUEROLES *et al.* (2013) in populations of Barcelona and Mt. Parnes with similar non-significant negative values of Tajima's *D*, Fu and Li's *D* in pooled populations and significant R_2 for $O_{\underline{3+4}}$ arrangement. So, the constant population size model was rejected for all chromosomal arrangements.

Table 4.11: Neutrality tests and test of population expansion for the eight regions in pooled populations using Ramos-Onsins and Rozas' R_2. Significant values are in bold. The significance of R_2 was calculated by coalescent simulations with estimated levels (ρ) of recombination.

		N	Tajima's *D*						Fu and Li's *D*			R_2	*Rho* (ρ)	*P*
			Whole sequence	Exons	Introns	Synonymous	Nonsynonymous	Silent	Whole sequence	Introns	Exons			
Acph-1	O_{ST}	13	-0,609	-0,369	-1,053	-0,061	-1,551	-0,476	-0,534	-0,184	-0,677	**0.106**	71.137	**0.043**
	O_{3+4}	17	-0,539	-0,259	-1,452	-0,065	-1,246	-0,447	-0,427	-0,231	-0,528	**0.105**	>100	**0.032**
	O_{3+4+8}	12	-0,506	-0,548	-0,499	-0,239	-1,713	-0,354	-0,047	-0,203	0,165	0.124	>100	0.242
Ast	O_{ST}	11	-0,734	-0,234	-0,992	0,086	-1,129	-0,688	-1,080	-0,985	-1,124	**0.078**	52.609	**0.000**
	O_{3+4}	14	-0,780	0,425	-0,799	0,869	-1,155	-0,755	-1,202	-1,440	0,620	**0.086**	76.562	**0.001**
	O_{3+4+8}	11	-0,686	-0,445	-0,835	0,238	-1,430	-0,614	-0,678	-0,622	-0,657	0.108	12.529	0.063
larp	O_{ST}	11	-1,344	**-1,933***	0,836	-1,673	**-1,965***	-0,867	-2,053	-0,114	**-2,559****	**0.062**	17.785	**0.001**
	O_{3+4}	20	**-1,810***	**-2,078***	-0,782	**-2,125***	-1,747	-1,711	**-2,862****	-0,019	**-3,570****	**0.061**	6.575	**0.001**
	O_{3+4+8}	15	-1,310	**-1,893***	1,014	-1,660	**-1,845***	-0,702	-1,809	0,446	-2,144	**0.090**	1.236	**0.023**
reg_larp	O_{ST}	9	-0,523	-	-0,523	-	-	-0,523	-1,323	-1,323	-	**0.104**	91.015	**0.016**
	O_{3+4}	13	-1,053	-	-1,053	-	-	-1,053	-0,976	-0,976	-	**0.094**	>100	**0.009**
	O_{3+4+8}	8	-0,773	-	-0,773	-	-	-0,773	-0,575	-0,575	-	0.130	>100	0.236
trus	O_{ST}	13	-0,643	-0,967	-0,656	-0,861	-1,149	-0,568	-0,963	-0,668	-0,910	0.104	27.420	0.058
	O_{3+4}	18	-1,498	-1,269	**-1,956***	-1,660	-0,481	**-1,893***	**-2,249***	-1,816	-2,053	**0.058**	5.195	**0.002**
	O_{3+4+8}	10	-0,894	-0,432	-1,712	0,124	-1,401	-0,619	-0,924	-1,864	-0,596	0.150	0.016	0.455
CG5961	O_{ST}	13	-1,658	-1,658	-	-1,486	-1,468	-1,486	**-2,903****	-	**-2,903****	0.119	10.005	0.070
	O_{3+4}	18	-0,665	-0,665	-	-0,376	-1,165	-0,376	-0,954	-	-0,954	0.099	>100	0.065
	O_{3+4+8}	10	-0,364	-0,364	-	0,025	-1,401	0,025	0,247	-	0,247	0.131	12.834	0.237
Fmr1	O_{ST}	12	-0,467	-0,279	-0,512	-0,279	n.a.	-0,467	-0,676	-0,941	-0,275	0.158	78.107	0.864
	O_{3+4}	19	-1,186	-1,064	-1,129	-0,949	-1,165	-1,167	-1,442	-0,930	-1,931	**0.080**	>1009	**0.000**
	O_{3+4+8}	11	-0,312	-0,746	-0,264	-0,746	n.a.	-0,312	-0,172	-0,161	-0,150	0.122	59.251	0.156
reg_Fmr1	O_{ST}	16	-0,934	-	-0,934	-	-	-0,934	-0,711	-0,711	-	0.096	9.925	0.056
	O_{3+4}	20	-0,881	-	-0,881	-	-	-0,881	-1,720	-1,720	-	**0.090**	18.796	**0.049**
	O_{3+4+8}	15	-1,091	-	-1,091	-	-	-1,091	-2,070	-2,070	-	**0.071**	1.517	**0.003**

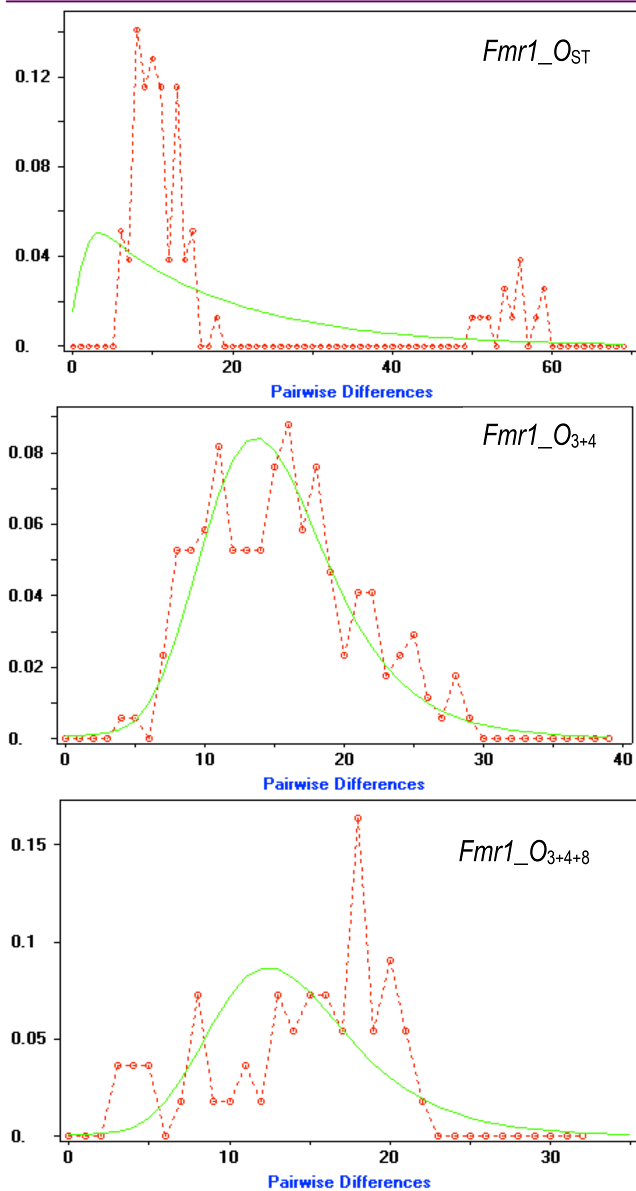


Figure 4.19: Graphical representation of the distribution of the pairwise differences under expectation of population expansion for three arrangements (also called *mismatch distribution*). Examples were chosen pooling two populations for the gene *Fmr1* that has significant values of R_2 in O_{3+4} arrangement and non-significant in O_{ST} and O_{3+4+8}.

The trend toward an excess of rare alleles detected by the Tajima's and Fu and Li´s tests in Barcelona and Málaga populations, significant in several cases, is similar to that observed for genes located inside the inverted regions in the previous studies (ROZAS *et al.* 1999; MUNTÉ *et al.* 2005, PEGUEROLES *et al.* 2013) and would reflect the selective sweep (MUNTÉ *et al.* 2005) that drove these arrangements to their equilibrium frequencies, although the situation of population expansion should be taken into account. The hypothesis of a recent population expansion after a bottleneck was further supported by the R_2 statistic. However, the positive although non-significant values of Tajima's D in the case of *Acph-1* and *trus* and non-significant R_2 values in most of the

cases, when two populations were considered separately in our study led to discard the hypothesis of an expansion of the whole species (ROZAS *et al.* 1999), which was also confirmed in the previous study on duplicated *Acp70A* gene, which is located outside of inverted regions and showed positive but non-significant Tajima's *D* values (CIRERA and AGUADÉ 1998).

The analysis on the five genes not associated with inversions in populations of Barcelona and Mt. Parnes (PEGUEROLES *et al.* 2013) suggested that *D. subobscura* has passed through a population expansion, although a positive *D* value for the gene *Atpα* in Mt. Parnes was detected. The number of significant R_2 estimates is almost the same in two populations (seven for Barcelona and six for Málaga). When considering expansion within arrangements in pooled populations, the null hypothesis of constant size could be rejected for all arrangements, as $O_{\underline{3+4+8}}$ demonstrates significant R_2 values for two, O_{ST} for four and $O_{\underline{3+4}}$ for seven gene regions out of eight. The inferred expansion in the $O_{\underline{3+4}}$ arrangement could be the result of its increase in frequency due to selection after the Pleistocene glaciation since this arrangement is considered to be warm adapted (BALANYÀ *et al*. 2004). Interestingly, this arrangement shows increased frequency worldwide matching recent global warming (BALANYÀ *et al*. 2006).

To determine the type of selection acting on each gene region grouped by arrangement, the McDonald and Kreitman test (MKT) was carried out, which compares the amount of variation within a species, or chromosomal arrangement as in our work, to the divergence between species, we used *D. pseudoobscura* for comparison, at two types of sites, one of which is putatively neutral and used as the reference to detect selection in another type of sites. As the test was initially described (MCDONALD and KREITMAN 1991), these sites were synonymous (putatively neutral) and non-synonymous in a coding region. However, the test for selection can potentially be extended to any two types of sites, provided that one of them is assumed to evolve neutrally and that both types of sites are linked in the genome sharing the same evolutionary history (EGEA, CASILLAS and BARBADILLA 2008).

MKT is less sensitive to the demographical effects than other similar selection tests (NIELSEN 2001). Under neutral model, any demographic factor affecting variability would be expected to act equally on neutral and non-neutral differences. However, when some of the nonsynonymous mutations are slightly deleterious, then P_n/P_s itself is sensitive to demographic effects. For example, in case of increase in the population size during the expansion process slightly

deleterious mutations could become fixed by genetic drift and P_n/P_s could become smaller than D_n/D_s since slightly deleterious mutations contribute disproportionately to divergence comparing with polymorphism (EYRE-WALKER 2002). It is important to note when interpreting results and distinguishing between different types of selection that the hypothesis of positive selection predicts more fixed replacement differences than predicted by the neutral mutation-random drift hypothesis since these are the changes that may have fitness effects. The hypothesis of balancing selection makes a different prediction: a preponderance of replacement polymorphisms maintained by selection. Negative selection is much more frequent than positive selection, therefore, in many cases the positive selection is masked and difficult to detect. The power of MKT to detect adaptive selection increases when excluding low-frequency polymorphisms (ANDOLFATTO 2005; CHARLESWORTH and EYRE-WALKER 2008). However, this approach also makes the test more sensitive to demographic effects (EYRE-WALKER 2002), so in these cases it is advisable to know the history of the population in order to ensure consistency in the effective population size. Another assumption of the test that must be taken into account is that it considers the selective constraints are constant over time. If there has been a recent change in selective constraints, either by a relaxation of selection or increased selective constraint, the proportions D_n/D_s and P_n/P_s would not have been expected to be equal. In addition, it is also assumed that the sites being compared are closely linked and the null hypothesis can be wrongly rejected in cases of intermediate levels of recombination (ANDOLFATTO 2008).

Both tests for coding and noncoding regions were performed pooling the two populations to increase the statistical power (Tables 4.12 and 4.13) and excluding the recombinant lines. It was observed that α in coding regions had positive values in eight out of eighteen cases (six genes within three inversions) and ranged between -3.849 and 1. The significant selection was detected in the coding portion of *Acph-1* for $O_{\underline{3+4}}$ arrangement ($O_{\underline{3+4}}$: α=0.612, P=0.019; Table 4.12) and in all arrangements it was positive due to an excess of nonsynonymous fixed changes (D_n), which is different from the results of NAVARRO-SABATÉ, AGUADÉ and SEGARRA (1999a; 2003) who found an excess of nonsynonymous polymorphism (P_n) in the same arrangements in populations of Spain and Tunisia, using the sequence of *D. guanche* as reference.

Table 4.12: McDonald and Kreitman test for coding regions in pooled populations.

		N	Polymorphism		Divergence		*NI*	*α*	X^2	*P*
			P_s	P_n	D_s	D_n				
Acph1	O_{ST}	13	38	11	133.87	53.87	0.719	0.280	0.761	0.382
	O_{3+4}	17	49	8	125.39	52.79	0.387	**0.612**	5.478	**0.019**
	O_{3+4+8}	13	42	8	128.73	52.79	0.464	0.533	3.465	0.062
Ast	O_{ST}	13	5	1	34.01	5.06	1.343	-0.343	0.061	0.804
	O_{3+4}	15	5	1	29.31	3.02	1.939	-0.939	0.288	0.591
	O_{3+4+8}	11	4	2	29.31	3.02	4.849	-3.849	3.557	0.109
Iarp	O_{ST}	13	10	8	115.22	85.63	1.076	-0.076	0.022	0.881
	O_{3+4}	20	14	13	118.09	85.63	1.280	-0.280	0.364	0.546
	O_{3+4+8}	15	9	12	115.22	85.63	1.794	-0.794	1.624	0.202
CG5961	O_{ST}	13	7	2	65.07	11.18	1.662	-0.662	0.351	0.553
	O_{3+4}	18	5	1	72.70	11.18	1.300	-0.300	0.053	0.817
	O_{3+4+8}	10	8	2	70.77	11.18	1.582	0.582	0.293	0.588
trus	O_{ST}	13	12	1	102.26	22.40	0.380	0.619	0.881	0.347
	O_{3+4}	18	13	8	98.73	20.33	2.987	**-1.987**	4.886	**0.027**
	O_{3+4+8}	10	5	2	102.25	20.33	2.011	-1.011	0.666	0.414
Fmr1	O_{ST}	13	5	0	19.80	2.00	0.000	1.000	0.496	0.480
	O_{3+4}	19	12	1	17.40	2.00	0.723	0.276	0.064	0.799
	O_{3+4+8}	11	9	0	18.60	2.00	0.000	1.000	0.939	0.332

N, sample size; P_s, neutral polymorphic sites; P_n, non-neutral polymorphism; D_s, neutral divergence; D_n, non-neutral divergence; *NI*, neutrality index; *α*, proportion of adaptive substitutions. The significant values of *α* with $p<0.05$ and nearly significant are in bold. The estimates were computed with the divergence corrected by JUKES and CANTOR (1969).

Table 4.13: McDonald and Kreitman test for noncoding regions in pooled populations.

		N	Polymorphism		Divergence		*NI*	*α*	X^2	*P*
			P_s	P_n	D_s	D_n				
Acph1	O_{ST}	13	20	15	74.13	96.99	0.573	0.426	2.237	0.134
	O_{3+4}	19	26	24	67.85	90.01	0.695	0.304	1.246	0.264
	O_{3+4+8}	13	23	21	67.82	94.18	0.657	0.342	1.520	0.217
Ast	O_{ST}	11	5	26	15.20	153.74	0.514	0.485	1.466	0.225
	O_{3+4}	16	7	62	11.23	150.28	0.662	0.337	0.674	0.411
	O_{3+4+8}	11	4	34	12.52	153.74	0.692	0.307	0.373	0.541
Iarp	O_{ST}	15	10	6	66.19	31.79	1.249	-0.249	0.158	0.690
	O_{3+4}	20	11	8	67.51	30.40	1.614	-0.614	0.881	0.347
	O_{3+4+8}	15	6	5	67.50	31.79	1.769	-0.769	0.803	0.369
regIarp	O_{ST}	9	7	50	66.19	397.86	1.188	-0.188	0.165	0.684
	O_{3+4}	13	11	56	67.51	384.50	0.893	0.106	0.099	0.751
	O_{3+4+8}	8	6	35	67.50	405.75	0.970	0.029	0.004	0.948
trus	O_{ST}	13	8	6	63.62	47.22	1.010	-0.010	0.000	0.985
	O_{3+4}	18	8	5	60.17	48.50	0.775	0.224	0.179	0.672
	O_{3+4+8}	11	4	2	60.21	47.22	0.637	0.362	0.261	0.609
Fmr1	O_{ST}	13	3	66	14.36	331.42	0.953	0.046	0.005	0.941
	O_{3+4}	19	8	58	13.14	353.80	0.269	**0.730**	8.778	**0.003**
	O_{3+4+8}	11	6	35	13.14	358.05	0.214	**0.785**	10.256	**0.001**
regFmr1	O_{ST}	16	3	22	14.36	54.26	1.940	-0.940	0.966	0.325
	O_{3+4}	20	8	29	13.14	52.10	0.914	0.085	0.031	0.859
	O_{3+4+8}	15	6	25	13.14	52.17	1.049	-0.049	0.007	0.929

N, sample size; P_s, neutral polymorphic sites; P_n, non-neutral polymorphism; D_s, neutral divergence; D_n, non-neutral divergence; *NI*, neutrality index; *α*, proportion of adaptive substitutions. The significant values of *α* with $p<0.05$ are in bold. The estimates were computed with the divergence corrected by JUKES and CANTOR (1969).

Significant purifying selection was detected in the exonic part of the gene *trus* (α=-1.987, P=0.027) in $O_{\underline{3+4}}$ arrangement (see table 4.12). When each population was considered separately (Supplementary Table S6a), similar results were observed in the two populations with a few exceptions (significant positive selection in *Acph-1* gene for $O_{\underline{3+4+8}}$ arrangement in the population of Barcelona and strong negative selection in the gene *Ast* for the same arrangement in the population of Málaga was also found). As for the noncoding regions, including introns of all genes (but CG5961, for which only exon was sequenced) and regulatory regions of two candidate genes, the prevalence of positive values of α (14 positive *vs*. 7 negative values) was observed in the pooled dataset (Table 4.13) although significant deviation from neutrality was only found for the gene *Fmr1* in the arrangements $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ (α=0.730, P=0.003 and α=0.785, P=0.001 accordingly).

When considering each population separately, the population of Barcelona gave similar result with significant deviation from neutral mutation-random drift in *Fmr1* gene for the same arrangements (Supplementary Table S6b). In Málaga three cases of significant positive selection were detected corresponding to the *Acph-1* gene in O_{ST} arrangement (α=0.719, P=0.008), to the regulatory region of *larp* (α=0.779, P=0.015) in $O_{\underline{3+4+8}}$ and to the gene *Fmr1* for arrangement $O_{\underline{3+4}}$ (α =0.741, P=0.008). Such discrepancy in the results when the two populations were combined could be attributed to smaller sample size in the Málaga population, especially for the $O_{\underline{3+4+8}}$ arrangement.

4.3.5. PATTERNS OF SELECTION

For an advantageous inversion, the action of directional selection would rapidly drive the new arrangement to its equilibrium frequency. As a result of this rapid increase, all regions included in the new arrangement would be completely depleted of variation even when the inversion had reached a relatively high frequency. Indeed, inverted chromosomes would initially be monomorphic for the particular haplotype captured by the inversion, which would include not only members of the coadapted gene complex but also neutral variants. The establishment of an inversion can thus be envisaged as a partial hitchhiking or selective sweep (MAYNARD-SMITH and HAIGH 1974) that would lead to an initial genetic differentiation of inverted and non-inverted chromosomes. Moreover, new mutations arising independently in the different arrangements would contribute to their further differentiation. Genetic exchange between chromosomal arrangements, either by gene conversion

or by double crossover, could, however, erode any genetic differentiation. Most important, it could break down the coadapted gene complexes putatively underlying the selective advantage of inversions.

In the absence of selection, genetic differentiation would decay according to the rate of genetic exchange among arrangements. The gene conversion rate would be uniformly distributed along the inversion loop, whereas the contribution of double crossovers to genetic exchange would be considerably higher in the central part of the inversion loop (NAVARRO *et al.* 1997). Under this scenario (*i.e.*, in which genetic exchange increases with physical distance to inversion break points), genetic differentiation among arrangements would be weaker in the central part of the loop than near the break points (NAVARRO, BARBADILLA and RUIZ 2000). In contrast, if selection were maintaining coadapted gene complexes, it would counteract the homogenizing effect of genetic exchange on members of the complex. The differential action of selection would cause different levels of genetic differentiation along the inversion, but no relationship would be expected between the level of differentiation and the physical distance to break points. Analysis of nucleotide variation along an inversion can thus inform us about the role played by natural selection in the establishment and maintenance of chromosomal polymorphism.

Several observations lead to the conclusion that the inversions of *D. subobscura* are adaptive, but perhaps the most important is the existence of latitudinal clines for the inversion frequency in the same direction in European as well as in the American populations (PREVOSTI *et al.* 1988; BALANYÀ *et al.* 2003). In particular the arrangements O_{ST} and $O_{\underline{3+4}}$, which were studied more extensively in *D. subobscura*, presented opposite distributions: the first is more common in northern Europe and demonstrates a positive correlation with latitude, while the arrangement $O_{\underline{3+4}}$ shows a negative correlation (KRIMBAS 1993; MESTRES *et al.* 1994; SOLÉ *et al*. 2002; BALANYÀ *et al.* 2004).

The establishment and maintenance of inversion polymorphism in natural populations of *Drosophila* has been explained by a superior fitness of heterokaryotypes (DOBZHANSKY 1970). The pattern of variation detected in the present multilocus study is consistent with the action of natural selection in the establishment of O_{ST}, $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ chromosomal arrangements. The general trend toward an excess of low frequency variants in the derived arrangements and the significant neutrality tests would reflect the partial hitchhiking or selective sweep that drove these arrangements to their equilibrium frequencies.

After the partial selective sweep associated with the establishment of a new inversion, a strong depletion of variation is expected around the breakpoints and also in very close-by regions (Andolfatto, Depaulis and Navarro 2001). Indeed, new variation in these regions can be introduced only by mutation, as gene conversion would be suppressed due to mechanical problems in synapses. Although some of the regions studied here are rather close to the breakpoints, none of them exhibits a reduction in variation. Indeed, estimates of π_{sil} in these regions are similar to, although slightly lower than, the value estimated for the *Acp70A* region in *D. subobscura* (π_{sil} = 0.016; Cirera and Aguadé 1998), which is located in a chromosomal region not affected by inversions. These results, and the detection of gene conversion tracts in most of the regions studied, indicate that their distance to the nearest break point is high enough for gene conversion to have contributed to the recovery of variation.

Because of the mutational process and the suppression of recombination in inversion heterokaryotypes, some genetic differentiation between arrangements is expected. However, differentiation may be counter-balanced by genetic exchange between arrangements. It has been proposed that the level of genetic exchange should be higher for markers located in central positions of the inversion loop than for markers located near the breakpoints (Rozas and Aguadé 1994; Navarro *et al.* 1997; Navarro, Barbadilla and Ruiz 2000; Andolfatto, Depaulis and Navarro 2001). Indeed, near breakpoints, genetic exchange should be possible only by gene conversion, whereas in the central region of the inversion loop not only gene conversion but also double crossover can contribute to the genetic exchange. The genetic exchange is considered to be eventually absent at the inversion break points themselves where chromosomes cannot establish synapsis properly (Wesley and Eanes 1994; Andolfatto, Wall and Kreitman 1999; Cáceres *et al.* 1999; Cáceres, Puig and Ruiz 2001). Empirical data obtained for markers located at different positions relative to the breakpoints in *D. melanogaster* support the proposed differential genetic exchange along the inversion (Hasson and Eanes 1996). However, double crossovers were empirically detected in regions close to breakpoints in *D. subobscura* (Pegueroles *et al.* 2010b)

The multilocus analysis reported here clearly indicates and confirms the previous results that genetic differentiation between O_{ST} and $O_{\underline{3+4}+X}$ is strong and extends all over the inversion. There is no evidence for the higher genetic exchange between arrangements expected in the central part of

the inversion loop in the presence of gene conversion and double crossovers (NAVARRO *et al.* 1997). The rather homogeneous distribution of genetic exchange detected across the inversion would indicate, therefore, that selection has acted against the recombinant chromosomes.

The occurrence over evolutionary time of double crossovers inside an inversion loop may be contingent on its length and age. Considering the empirical values of interference in *Drosophila*, NAVARRO *et al.* (1997) suggested that double crossover is unlikely only in short inversions (<20 cM). The estimated length of the O_3 and O_4 (≈23 and 53 cM respectively) would thus *a priori* support that double crossovers could be reduced but probable in these inversions. In addition, the time elapsed since the origin of $O_{\underline{3+4}}$ and O_{ST} (0.25–1.06 MYA; ROZAS and AGUADÉ 1994; PEGUEROLES *et al.* 2013) is long enough for double crossovers to have broken the initial associations, at least in the central part of the inversion loop. Double crossovers also have not been effective in eroding the genetic differentiation in the central part of the ≈55-cM-long inversion that differentiates the $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements (ROZAS *et al.* 1999; NAVARRO-SABATÉ, AGUADÉ and SEGARRA 2003) as it was confirmed in the present study.

Accepting the occurrence of double crossovers, selection acting against the products of genetic exchange between chromosomal arrangements, and more specifically against double-crossover products, would be the most plausible explanation for the strong genetic differentiation detected in the eight regions studied. Indeed, epistatic fitness interactions among genes within the inversion would result in the lower fitness of those among-arrangement recombinants that affected the coadapted complex. Sets of coadapted linked genes would be broken more likely by double crossover than by gene conversion, as the lengths of the segments affected by gene conversion are much shorter (HILLIKER *et al.* 1994; BETRÀN *et al.* 1997). Consequently, selection would have acted mostly against double-crossover products.

The eight gene regions studied, whether or not identified as candidates for thermal adaptation, exhibited a strong genetic differentiation. The significantly high genetic differentiation between the three arrangements by almost all gene regions despite the large size of some inversions might be indicating the possible existence of selection patterns for all of them. For epistatic selection to explain this result, the regions need not be the targets of selection themselves, but they should be tightly linked to coadapted genes. Our observation would imply a rather high number of target genes or, alternatively, fewer genes with stronger effects. Indeed, the high level of interlocus LD

detected in the total sample of $O_{ST}/O_{\underline{3+4}}$ and $O_{ST}/O_{\underline{3+4+8}}$ chromosomes (Figure 4.14) indicates that the regions linked to each arrangement have followed independent evolutionary histories. These does not seems surprising since heterokaryotypes for these two combinations form two and three overlapping small inversion loops respectively, which may reduce recombination. However the proportion of double crossovers identified in relation to gene conversion tracts comparing these two combinations is similar to that comparing $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ chromosomes (0.2-0.5) with *Acph-1* showing the larger number of recombinant lines irrespective of being more closely located to the inversion breakpoint (when comparing to O_{ST}) or more centrally located in the inversion loop suggesting a clear role of selection maintaining the strong genetic differentiation between arrangements.

Prevalence of genetic differentiation despite the presence of genetic exchange between arrangements is also reflected in the recombination networks inferred from nucleotide variation of the genes. The clustering of lines into two main groups (O_{ST} and $O^*_{\underline{3+4}}$) in the majority of gene regions is consistent with restricted genetic exchange between these chromosomal classes. In contrast, $O^*_{\underline{3+4}}$ lines do not completely cluster according to their gene arrangement, which is consistent with a higher level of genetic exchange within $O^*_{\underline{3+4}}$ than between O_{ST} and $O^*_{\underline{3+4}}$. However, genetic exchange within $O^*_{\underline{3+4}}$ has not been high enough to erase completely the original chromosomal relationships among lines, as reflected in the partial clustering of lines by gene arrangement (Figure 4.16) and the significant genetic differentiation between these two arrangements for almost all gene regions (Table 4.7). A similar result was reported earlier, according to genealogies at the *rp49* (ROZAS *et al.* 1999) and the *Acph-1* gene regions (NAVARRO-SABATÉ, AGUADÉ and SEGARRA 2003), which are tightly linked and centrally located in the O_8 inversion.

The absence of significant correlations between the position of the gene regions inside an inversion and their values of nucleotide diversity and F_{ST} could be interpret as the selective force acting on them, including the effect of hitchhiking, rather than the positional effects. Furthermore, no significant differences in the pattern obtained for Tajima's and Fu and Li's (with outgroup) tests were found. In nearly all comparisons values were negative, suggesting an excess of polymorphisms at low frequency. In most of the cases values were not significant, with the exception of the exons from *larp* in all arrangements and *trus* in $O_{\underline{3+4}}$ arrangement in the tests of

Tajima and Fu and Li and also CG5961 in O_{ST} in the Fu and Li´s test alone. The deviation patterns in these genes could be associated to the selective sweep as well as purifying selection.

However, we cannot discard that extreme values were assessed as significant as a consequence of the bias to negative values. The deviation from neutrality in these genes as a result of the demographical changes cannot be discarded either. In a previous study carried out with $O_{\underline{3+4}}$ and O_{ST} arrangements, this pattern was interpreted as an expansion of Barcelona populations (PEGUEROLES *et al.* 2013), which is actually could be the case taking into account the significant values of R_2, detecting population expansion in our study.

MKT is less sensitive to the demographical effects than other similar selection tests, however, the test assumes that population size does not change over time, which is probably not the case in the studied chromosomal arrangements and could influence our results.

The genes *larp* and *Fmr1* were selected from a list of candidate genes to be involved in thermal adaptation (LAAYOUNI *et al.* 2007), whereas the rest of the genes were chosen from the work of MUNTÉ *et al.* (2005) as good candidates to detect differentiation between two populations, if any, and between closely related gene arrangements $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$. When grouping arrangements, significant pairwise comparisons for LD were obtained for genes separated by a high physical distance (7 Mb). These LD observations could be influenced by the maintenance of coadapted gene complexes (HOFFMANN and RIESEBERG 2008). The observation of linkage disequilibrium between non-overlapping inversions, as in the $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ studied arrangements, could be consequence of the presence of epistatic interactions between them (KRIMBAS 1993; SPERLICH and FEUERBACH-MRAVLAG 1974).

The gene *larp* encodes the *Drosophila* orthologue of La-related protein containing a domain exhibiting an outstanding homology with La type RNA-binding protein and plays the essential role in multiple processes in *Drosophila* mitotic cell division and male meiosis (ICHIHARA *et al.* 2007; BLAGDEN *et al.* 2009), that could be related with maintaining male fertility in altered thermal conditions as it is well known that *Drosophila* males became sterile in high temperature regimes.

The significant deviation from the neutrality as an evidence of positive selection was detected in the intronic positions of the gene *Fmr1* for $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ gene arrangements by McDonald and

Kreitman test. The gene *Fmr1,* whose some of the biological functions is protein and RNA binding, takes part in such processes as synaptic transmission, locomotory behavior and biological regulation (WAN *et al.* 2000; Table 4.3), which could be important in terms of adaptation to temperature especially on the behavioral level, where the chromosome *O* seems to play a substantial role (REGO *et al.* 2010; DOLGOVA *et al.* 2010).

The higher proportion of nonsynonymous fixed differences in the exonic part of a gene as a proof of positive selection was only found for the gene *Acph-1* in $O_{\underline{3+4}}$ arrangement in each population as well as in the joined population and in $O_{\underline{3+4+8}}$ in the population of Barcelona. Acid phosphatase-1 (ACPH-1) is a lysosomal enzyme that was first described in the genus *Drosophila* by MACINTYRE (1966). Immunological studies in this genus have shown that ACPH-1 is rapidly evolving in comparison to other enzymatic proteins such as superoxide dismutase (SOD) and 6-phosphogluconate dehydrogenase (GPDH) (COLLIER, MACINTYRE and FEATHERSTON 1990). *Acph-1* is located inside and rather close to one of the ends of the inversion loop formed in $O_{ST}/O_{\underline{3+4}}$ and $O_{ST}/O_{\underline{3+4+8}}$ heterokaryotypes. This is not the case for the $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ heterokaryotypes, in which *Acph-1* maps in a central position of the inversion loop (Figure 4.14). Consequently, genetic differentiation at *Acph-1* is expected to be higher between O_{ST} relative to both $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ than between the two latter arrangements and actually it is: the higher values of F_{ST} between O_{ST} relative to $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ than between the two latter arrangements, data on linkage disequilibrium between nucleotide polymorphic sites and the type of arrangement (see Supplementary Table S4) are consistent with this expectation. This gene presented high levels of nucleotide variation and recombination, despite being the closest gene to the nearest break point. Contrary to our results for the MKT, previous studies detected an excess of nonsynonymous polymorphism in *Acph-1* using *D. guanche* as a reference species. This excess resulted in extraordinary high neutrality index for O_{ST} (*NI*=9.01) (NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a), as well as for $O_{\underline{3+4+8}}$ (*NI*=3.80) and $O_{\underline{3+4}}$ (*NI*=1.74) (NAVARRO-SABATÉ, AGUADÉ and SEGARRA 2003), which differed considerably from our results, where values of *NI* were smaller than 1, ranging between 0.39 and 0.72, for these three arrangements, resulting into positive values of *α*, which is evidence of an excess of nonsynonymous fixed changes and thus of the positive selection and fast fixation of advantageous alleles. The discrepancy could arise by the use of different outgroups differing in their divergence time (RUSSO, TAKEZAKI and NEI 1995; GAO *et al.* 2007).

In addition to the significant deviation from the neutrality demonstrated by Tajima´s D and F and Li's tests the significant purifying selection was detected in exons of *trus* for O_{3+4} arrangement in both populations considered separately as well as pooled together due to an excess of nonsynonymous polymorphism. Different selective factors have been proposed to explain the excess of nonsynonymous polymorphism. First, balancing selection maintaining protein variation or, alternatively, positive selection fluctuating over time or space might contribute to an excess of amino acid variation (GILLESPIE 1991). Second, mild negative selection acting on replacement mutations can cause these mutations to be maintained in the population at low frequency, but rarely to become fixed (OHTA 1992). Third, a relaxation of selection acting in the past might maintain previously non-tolerated nonsynonymous polymorphism (TAKAHATA 1993). The excess of nonsynonymous polymorphism detected at *trus* can be likely explained according to the nearly neutral model of molecular evolution (OHTA 1992), and assuming that most replacement mutations in these genes are slightly deleterious. Apart from the variants responsible for the polymorphism, almost all singleton nonsynonymous variants detected in O_{3+4} arrangement of *trus* are derived and absent from the other arrangements. Therefore, it can be inferred that these variants appeared by mutation and were kept at low frequency within each arrangement by weak negative selection; i.e. nonsynonymous mutations are slightly deleterious. The pattern of variation in O_{3+4} arrangement of *trus* has the same signature as previously reported for most genes of mitochondrial genome (mtDNA) in *D. melanogaster* and other organisms (BALLARD and KREITMAN 1994; NACHMAN, BOYER and AQUADRO 1994; RAND, DORFSMAN and KANN 1994; NACHMAN *et al.* 1996; RAND and KANN 1996; HASEGAWA, CAO and YANG 1998; KENNEDY and NACHMAN 1998; WISE, SRAML and EASTEAL 1998). This signature has been mainly explained according to the nearly neutral model and assuming that nonsynonymous mutations are weakly selected. As discussed by WEINREICH and RAND (2000), two characteristics of the mitochondrial genome that may contribute to the detected pattern of variation are the lack of recombination and its lower effective size relative to the nuclear genome. Therefore, it is tempting to argue that these characteristics may also contribute to the pattern of variation found at *trus* in O_{3+4} arrangement. The lack of recombination in mtDNA might have the same reason as the reduction in recombination at this gene in heterokaryotypes. In addition, a reduction in recombination can be viewed as a reduction in effective size as a consequence of the Hill-Robertson effect (HILL and ROBERTSON 1966). In any way, a small effective size results in a longer

persistence of slightly deleterious mutations in the population (i.e., negative selection against slightly deleterious mutations is relaxed).

This excess of nonsynonymous polymorphism should be also expected in other genes associated with the different arrangements. Our data on the rest of the genes and previous results for *rp49* (Rozas *et al.* 1999) do not support this expectation, because in these genes no nonsynonymous polymorphism was detected in any of the studied gene arrangements. However, it has to be considered that the distribution of the selective coefficient acting against nonsynonymous mutations may differ among genes depending on the functional constraints of the encoded proteins. Constraints on the protein that coded by *trus* would likely be relatively high, as indicated by the rather low nonsynonymous divergence at corresponding gene (data not shown). The high functional constraints on the protein could have greatly contributed to the detected results. The effect of chromosomal polymorphism on nonsynonymous variation would thus vary with the level of constraints of the encoded protein, and such an effect might be negligible in genes subjected to strong purifying selection, like for example *rp49*.

All in all, there were no significant differences between variability, gene arrangement differentiation, levels of LD or recombination, as well as selection, between candidate and non-candidate genes for thermal adaptation, all these parameters seem to depend on particular evolutionary history of a gene, its situation inside of inversion, the length of the sequenced region, etc. more than on its relation to adaptation to temperature.

Non-synonymous changes differentiating O_{ST} from $O^*_{\underline{3+4}}$ was found in the genes *trus* (see Supplementary Tables S2, nucleotides 173* and 218*) and *Ast* (nucleotide 1522*, Supplementary Tables S2). However, none of the two candidate genes involved in thermal adaptation show nonsynonymous changes differentiating arrangements. Selective pressure could focus on regulatory regions instead of coding regions (Torgerson *et al.* 2009). In the present study the footprint of selection was detected in regulatory region of *larp* for $O_{\underline{3+4+8}}$ arrangement in the population of Málaga, but this result could be affected by the small size of the sample.

In summary, significant genetic differentiation has been detected between chromosomal arrangements, although certain levels of recombination were also found, but there was a lack of genetic differentiation between populations for the same inversion. Taking into consideration ours

and previous results on different genes from different populations, the theory of coadaptation understood as Dobzhansky it formulated, implicating the differentiation of genetic content of inversions between different populations as well as epistatic interactions between loci inside the inversion (DOBZHANSKY 1950), does not seem to apply in *D. subobscura*, but the most appropriate hypothesis for explanation of the maintenance of inversions in populations would be the local adaptation hypothesis (KIRKPATRICK and BARTON 2006). In addition, the finding of a possible epistatic interaction between genes would not refute the hypothesis of local adaptation, as the authors of this theory suggest that, although it is not essential, epistatic interactions would help to maintain the genetic content of inversions. Additionally, under the coadaptation hypothesis, inversion heterozygotes are adaptively superior to the corresponding homozygotes only if the chromosome pair has been drawn from the same or neighboring localities (DOBZHANSKY and EPLING 1948). This was not confirmed by the recent experiment of PEGUEROLES *et al.* (2010b), since heterokaryotype descendants were always more frequent regardless of the origin of the lines used: they come from very distant populations, such as Greece, Spain and Sweden. Thus, different arrangements, though they come from different populations, would harbor the same adaptive alleles (captured in the inversion) granting adaptation success to a given environment.

4.3.6. AGE OF CHROMOSOMAL ARRANGEMENTS

Molecular data have been used to date evolutionary events in chromosomal arrangements at the population level under the assumption of constant substitution rate (NEI 1987; KIMURA 1983) and considering their likely unique origin (POWELL 1997 and references therein). These estimates can be inferred from the time of coalescence of the sequences, taking into account that variability accumulated in the sequences is proportional to its origin. However, it is known that allele genealogies and therefore the time of coalescence are affected by selection (HUDSON 1990). This effect of selection extends to linked neutral markers. For timing purposes, it is therefore desirable to use a gene region without evidence for selection (ROZAS and AGUADÉ 1994), which was difficult in our study, as nearly all gene regions demonstrated the deviation from neutrality when either Tajima's/Fu and Li's *D* or MKT were used.

ROZAS *et al.* (1999) proposed estimates of the inversion age according to the expansion model (SLATKIN and HUDSON 1991; ROGERS 1995) and to variation at the *rp49* gene region. When the

expansion model is applied to estimate the age of an inversion, it is assumed that nucleotide variation within each arrangement has not yet reached equilibrium and that it has accumulated independently. Therefore, polymorphic sites included in gene conversion tracts between arrangements have to be excluded from the analysis. Two criteria may confirm that nucleotide variation within the arrangement is still in the transient phase to equilibrium: first, the negative sign of Tajima's *D* statistics, which indicates an excess of rare variants, and second, the shape of the pairwise nucleotide difference distribution or mismatch distribution that is Poisson-like in expanding populations (HARPENDING *et al.* 1993; HARPENDING 1994). The negative Tajima's *D* and the Poisson-like distribution in all three assayed arrangements are consistent with the expansion model. The sudden expansion model basically depends on three parameters: θ_0, or initial theta; θ_1, or final theta; and $\tau = 2\mu t$ (units of mutation time, where μ is the mutation rate and *t* is the time in generations; ROGERS and HARPENDING 1992). We can estimate τ (ROGERS 1995) from the observed values of *K* (nucleotide diversity per sequence, that is, the average number of nucleotide differences between two sequences), considering that $\theta_0 = 0$ (due to the unique origin of inversions). When variation is null at the moment of the expansion, as in the case when an inversion originates, τ corresponds to the average number of nucleotide differences (π).

Furthermore, levels of silent variation in arrangements can be used to date the origin of them. As was found by PEGUEROLES *et al.* (2013) the age of inversions estimated using the two most divergent sequences seems more sensitive to differences in sample size or to differences in the genetic content between populations, as observed by the higher fluctuation of the estimated ages when mixing populations or estimating ages for each population separately. To avoid this potential bias the neutral mutation rate was estimated from the mean silent nucleotide diversity (π_{sil}). Silent nucleotide diversity at *Acph-1* was higher within $O_{\underline{3+4}}$ than within O_{ST} in the previous study of NAVARRO-SABATÉ, AGUADÉ and SEGARRA (1999a), which was in agreement with that previously found at the *rp49* gene region, but is not consistent with neither multilocus data on silent variation, nor the data for *Acph-1* gene in our observations and the study of PEGUEROLES *et al.* (2013) on *Fmr1*, where the significant differences in silent diversity were not detected between two gene arrangements. These estimates can be contrasted with the present data at the *Acph-1* and *Fmr1* gene regions as well as at the rest of the regions. The neutral mutation rate can be estimated from the rate of nucleotide substitutions in interspecific comparisons: $\mu = K_{sil}/2t_{div}$, where t_{div} is divergence

time between species. So, finally the inversion age could be calculated in Myr, using only silent substitutions, as $T=\pi_{sil}/2\mu*1000000$.

Ages were calculated using two divergence times between *D. subobscura* and *D. pseudoobscura*. The divergence time obtained by TAMURA, SUBRAMANIAN and KUMAR (2004) is more reliable since it is based on a large multilocus dataset and similar to that obtained by BECKENBACH, WEI and LIU (1993), based on transversions at the mitochondrial cytochrome oxidase II gene. Nonetheless the time estimated by RAMOS-ONSINS *et al.* (1998), although only based on the *rp49* gene and of the same order as for the nuclear *Gpdh* gene reported by WELLS (1996) allows the comparison with previous studies. *D. pseudoobscura* divergence time with respect to *D. subobscura* was estimated to be 17.7 Myr (TAMURA, SUBRAMANIAN and KUMAR 2004) or around 8 Myr (RAMOS-ONSINS *et al.* 1998). *D. pseudoobscura* is a native species from North America and its effective size could be quite similar to *D. subobscura* (PASCUAL, SCHUG and AQUADRO 2000).

Sequences of all gene regions were used for estimating the age of the inversions, since in all of them differentiation for all three arrangements was detected. To estimate the age of inversions, it is desirable to subtract all variation that has accumulated in the different chromosomal arrangements by processes other than mutation. Gene transfer between inversions tends to homogenize their genetic content and therefore counteracts the effect of mutation that tends to differentiate them by the independent accumulation of mutations (ROZAS and AGUADÉ 1994). So, the recombinant individuals (whenever due to crossover or to gene conversion) were excluded because only variation originated by mutation is useful for this analysis since variability originated by recombination causes an overestimation of the ages (ROZAS *et al.* 1999). The time estimates have been obtained under the assumption that all nucleotide variants observed in a particular chromosomal arrangement originated in that arrangement. However, some additional variants not included in the detected gene conversion or double crossover tracts (e.g., some shared polymorphisms) could have been transferred from other chromosomal arrangements; thus most probably for the $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements, the coalescent time could be overestimated.

The ages of inversions were estimated for each population separately (data not shown) and combining them since their origin should be unique (Table 4.14). Since older divergence times lead to estimate smaller mutation rates, the ages of the inversions estimated using Tamura's divergence time are sensitively older than using Ramos-Onsins' divergence time. The age estimates for O_{ST},

$O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ differed considerably among gene regions, ranging between 0.22-0.62 Myr for O_{ST}, 0.16-0.74 for $O_{\underline{3+4}}$ and 0.16-0.67 for $O_{\underline{3+4+8}}$ using Ramos-Onsins coalescence time estimation. Not always the same arrangement presented older times with all genes and thus we could consider that all arrangements arose at a similar time since the Sign test did not detect significant differences when comparing them (in all pairwise comparisons P>0.2). It is worth mentioning that due to the large standard deviations of the estimates of the number of nucleotide differences per site and to the undetected conversion tracts, these estimated times are rough estimates of the time of divergence. The mean age estimates obtained in the present study (0.39 Myr for O_{ST}, 0.43 Myr for $O_{\underline{3+4}}$ and 0.39 Myr for $O_{\underline{3+4+8}}$) were slightly older than those based on genes *Acph*-1 (0.26 Myr for O_{ST}, 0.31 Myr for $O_{\underline{3+4}}$ in NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a), *rp49* (0.24 Myr for O_{ST}, 0.33 Myr for $O_{\underline{3+4}}$ and 0.34 Myr for $O_{\underline{3+4+8}}$ in ROZAS *et al.* 1999), but similar to those obtained by PEGUEROLES *et al.* (2013) basing on *Fmr1* (0.48 Myr for O_{ST} and 0.41 Myr for $O_{\underline{3+4}}$) and using the same divergence time (RAMOS-ONSINS *et al*. 1998). Our age estimates from *Acph-1* and *Fmr1* differed from those previously published, being twofold higher in the case of *Acph-1* (NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a) and lower for *Fmr1* (PEGUEROLES *et al.* 2013) because variation at the noncoding regions of *Acph-1* and divergence time with *D. guanche* were used to estimate the age of the O_{ST} and $O_{\underline{3+4}}$ arrangements, while we used the silent variation in all gene regions and divergence time to *D. pseudoobscura*.

Table 4.14: Age of arrangements inferred using different times of divergence between *D. subobscura* and *D. pseudoobscura.*

	Age, Myr (TAMURA, SUBRAMANIAN and KUMAR 2004)			Age, Myr (RAMOS-ONSINS *et al.* 1998)		
	O_{ST}	O_{3+4}	O_{3+4+8}	O_{ST}	O_{3+4}	O_{3+4+8}
Acph-1	1.25	1.58	1.38	0.57	0.71	0.62
Ast	0.83	1.63	0.84	0.37	0.74	0.38
larp	0.78	0.82	0.97	0.4	0.31	0.34
reg-larp	1.14	0.57	0.25	0.38	0.31	0.27
CG5961	0.49	0.42	0.92	0.29	0.16	0.16
trus	0.65	0.36	0.35	0.22	0.19	0.41
Fmr1	0.54	0.77	0.72	0.25	0.35	0.33
reg-Fmr1	1.33	1.62	1.48	0.6	0.73	0.67
Concat.	0.88	0.82	0.94	0.4	0.37	0.43
mean	0.87	0.95	0.86	0.39	0.43	0.39
range	0.49-1.38	0.36-1.63	0.35-1.48	0.22-0.62	0.16-0.74	0.16-0.67

Differences also could be due to intrinsic characteristics of the different genes and to their distinct

genetic location. For example, genes *Acph*-1 and *rp49* are closely located inside the inversion O_3 near its distal break point (MUNTÉ *et al.* 2005), while *Fmr1* gene is located far away from them, inside the inversion O_4 near the distal break point. The sequenced part of the *Fmr1* gene in our study differed from that of PEGUEROLES *et al.* (2013), because we included a larger part of the intron. Since genes can differ in their selective pressure, mutation rates, or rate of recombination and each can affect estimates of their coalescence time, a multilocus approach as in TAMURA, SUBRAMANIAN and KUMAR (2004) is preferred to more precisely infer the age of inversions. Age estimates of O_{ST}, $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ arrangements were quite similar in the previous studies (ROZAS and AGUADÉ 1994; ROZAS *et al.* 1999; NAVARRO-SABATÉ, AGUADÉ and SEGARRA 1999a; PEGUEROLES *et al*. 2013) and O_{ST} was slightly younger than $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$. From multilocus analysis in our study we can conclude that all three arrangements are quite old and $O_{\underline{3+4}}$ is the older one (0.95 Myr on average), while O_{ST} and $O_{\underline{3+4+8}}$ are slightly younger and arose more or less at the same time (0.87 and 0.86 Myr on average respectively) consistently with the derived status of $O_{\underline{3+4+8}}$. These discrepancies could be due to differences in sample size, undetected recombination events, which would affect differentially methods employed in different studies, and gene localization relative to inversion break points.

4.4. THERMAL COADAPTATION

4.4.1. ASSOCIATION BETWEEN THERMAL PREFERENCE AND KNOCK OUT TEMPERATURE

The phenotypic correlation between T_p and T_{ko} was assessed from their partial correlation coefficient, holding constant the variables developmental temperature, sex, plate hour, and water bath (see Methods). In no case were the partial correlations statistically significant: inbred crosses $r_{Tp \cdot Tko}$ = 0.065, t = 1.21, df = 347, P = 0.226; outbred crosses $r_{Tp \cdot Tko}$ = -0.030, t = 0.79, df = 701, P = 0.429. Furthermore, as expected from the low values of the phenotypic correlation, the genetic (karyotypic) correlation for the outbred flies was also close to zero (r_k = -0.068, P = 0.914). The conclusion is that both traits are nearly orthogonal to each other (pooled $r_{Tp \cdot Tko}$ = 1.2 ×10^{-4}, t = 0.004, df = 1054, P = 0.997) and, hence, they will be analyzed separately in what follows.

4.4.2. Consanguinity and developmental effects

a) Thermal preference

Inbreeding and developmental temperature effects on T_p were simultaneously analyzed by contrasting isogenic *vs.* outbred homokaryotypic flies reared at both experimental temperatures (Figure 4.20). The factorial analysis of covariance (ANCOVA) only detected statistically significant differences for karyotypes, karyotype × inbreeding interaction, and karyotype × developmental temperature interaction effects (Table 4.15). Average (± SD) T_p was not different between rearing temperatures (flies reared at 18°C: 18.7°C ± 4.1°C; flies reared at 22°C: 18.8°C ± 3.1°C) or sexes (females: 19.0°C ± 3.6°C; males: 18.5°C ± 3.6°C), although in this last case the effect was marginally nonsignificant (P = 0.053). Permutation tests (see Methods) corroborated that the three assayed karyotypes differ in T_p (P = 0.001). Scheffé post hoc tests using the mean square of the nested "cross" effect as the error term showed that the thermal preference of O_{ST}/O_{ST} flies was significantly lower when compared to those of $O_{\underline{3+4}}/O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$ homokaryotypes, which did not differ between them. The difference was consistent for both isogenic and outbred flies (Figure 4.20). From the present data we can conclude that the preferred temperature ranges or "set point" (T_{set}) ranges (central 50% of preferred body temperatures; Hertz, Huey and Stevenson 1993) are bounded by 15.1°C - 20.5°C for O_{ST}/O_{ST} karyotypes, and 16.6°C - 22.2°C for the other two karyotypes.

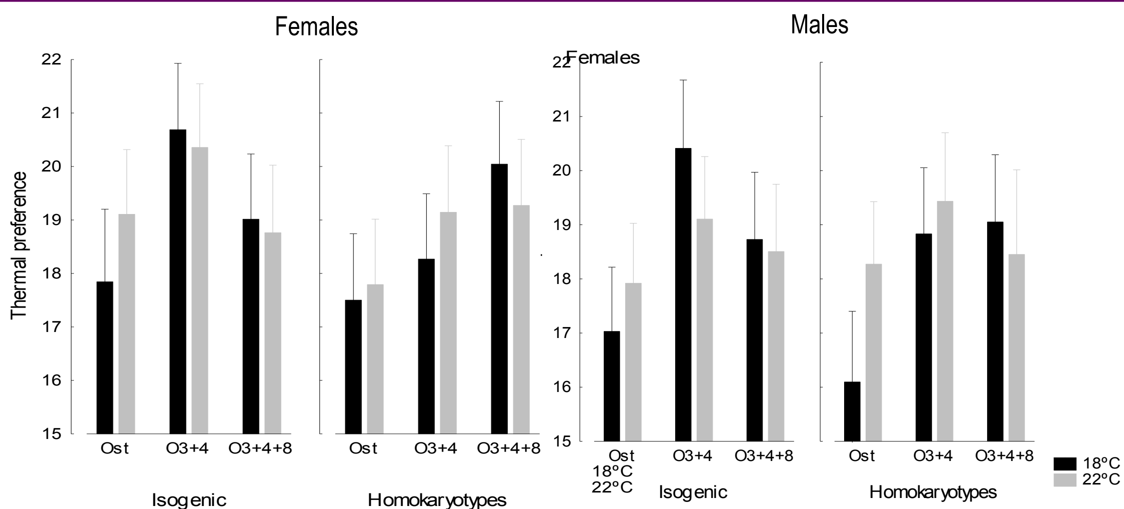


Figure 4.20: Inbreeding and temperature effects on thermal preference. Homokaryotypic averages for T_p (in °C with 95% confidence intervals) in inbred (left panels) and outbred (right panels) crosses according to sex and developmental temperature.

Table 4.16: Inbreeding and temperature effects on thermal preference.

Source of variation	d.f.	Mean Square	F	P
Covariate (plate hour)	1	25.502	2.07	0.151
Karyotype (κ)	2	231.515	18.29	<0.001
Cross $\subset \kappa$	15	12.676	1.03	0.425
Inbreeding (ι)	1	30.514	2.47	0.116
Temperature (τ)	1	4.119	0.33	0.564
Sex (ς)	1	46.227	3.74	0.053
$\kappa \times \iota$	2	40.337	3.27	0.039
$\kappa \times \tau$	2	40.031	3.24	0.040
$\kappa \times \varsigma$	2	6.195	0.50	0.606
$\iota \times \tau$	1	11.063	0.90	0.344
$\iota \times \varsigma$	1	6.257	0.51	0.477
$\tau \times \varsigma$	1	0.408	0.03	0.856
$\kappa \times \iota \times \tau$	2	17.477	1.42	0.243
$\kappa \times \iota \times \varsigma$	2	11.532	0.93	0.393
$\kappa \times \tau \times \varsigma$	2	7.600	0.62	0.541
$\iota \times \tau \times \varsigma$	1	12.123	0.98	0.322
$\kappa \times \iota \times \tau \times \varsigma$	2	4.245	0.34	0.709
Error	717	12.346		

Flies risen from inbred (isogenic) and outbred crosses of *Drosophila subobscura* reared at 18°C and 22°C. Karyotypes being compared are O_{ST}/O_{ST}, $O_{\underline{3+4}}/O_{\underline{3+4}}$, and $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$ ($\subset$ means "nested in").

The karyotype × inbreeding interaction arises from the somewhat different behavior between O_{ST}/O_{ST} and $O_{\underline{3+4}}/O_{\underline{3+4}}$ karyotypes on one side, and $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$ on the other: for the first two karyotypes T_p was slightly higher in inbred crosses when compared to their outbred counterparts, whereas the opposite was true for the $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$ karyotype. Average T_p was, however, almost identical for inbred (18.9°C ± 3.6°C) and outbred (18.5°C ± 3.6°C) flies. On the other hand, O_{ST}/O_{ST} flies raised at 22°C had a higher T_p than those rose at 18°C, but no clear trend was observed for $O_{\underline{3+4}}/O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$ karyotypes.

b) Knock out temperature

Knock out temperatures are plotted in Figure 4.21. The ANCOVA (Table 4.17) detected statistically significant differences for the effects of rearing temperature and sex. Flies reared at 18°C had a higher T_{ko} than flies reared at 22°C (mean ± SD: 33.3°C ± 2.1°C *vs*. 32.6°C ± 2.3°C), and females had a higher T_{ko} than males (33.4°C ± 1.9°C *vs*. 32.5°C ± 2.4°C). Even though T_{ko} was slightly lower for the isogenic lines when compared to their outbred counterparts (32.8°C ± 2.2°C *vs*. 33.1°C ± 2.2°C), inbreeding effects were clearly non-significant (P = 0.136).

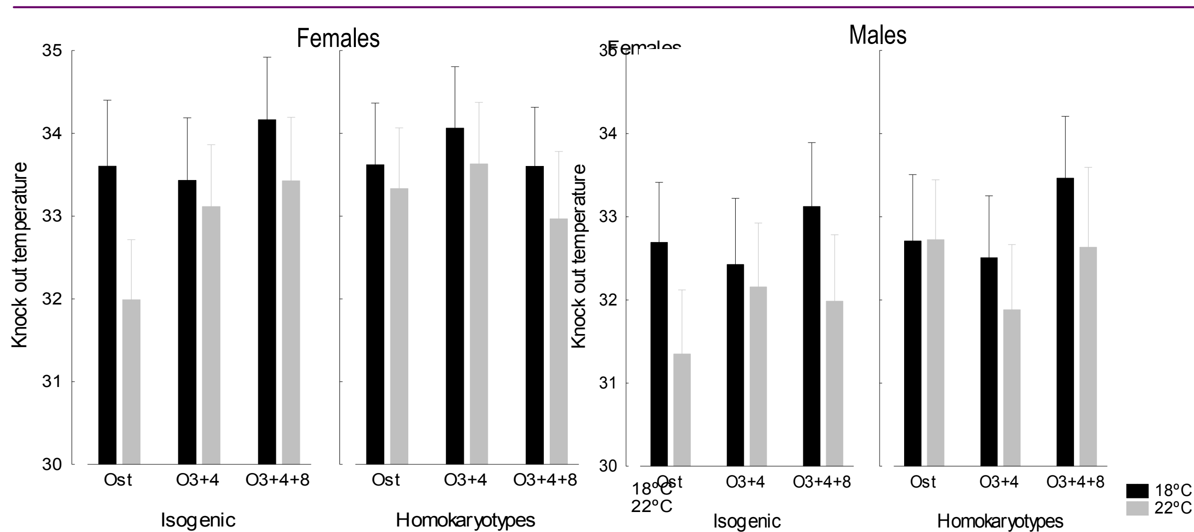


Figure 4.21: Inbreeding and temperature effects on knock out temperature. Homokaryotipic averages for T_{ko} (in °C with 95% confidence intervals) in inbred (left panels) and outbred (right panels) crosses according to sex and developmental temperature.

Table 4.17:Inbreeding and temperature effects on knockout temperature.

Source of variation	d.f.	Mean Square	*F*	*P*
Covariate (water bath)	1	103.117	24.04	<0.001
Karyotype (κ)	2	3.878	0.36	0.704
Cross $\subset \kappa$	15	11.027	2.57	0.001
Inbreeding (ι)	1	9.538	2.22	0.136
Temperature (τ)	1	77.034	17.96	<0.001
Sex (ς)	1	154.979	36.13	<0.001
$\kappa \times \iota$	2	4.176	0.97	0.378
$\kappa \times \tau$	2	1.999	0.47	0.628
$\kappa \times \varsigma$	2	8.106	1.89	0.152
$\iota \times \tau$	1	1.047	0.24	0.621
$\iota \times \varsigma$	1	0.435	0.10	0.750
$\tau \times \varsigma$	1	0.022	0.01	0.943
$\kappa \times \iota \times \tau$	2	7.798	1.82	0.163
$\kappa \times \iota \times \varsigma$	2	8.926	2.08	0.126
$\kappa \times \tau \times \varsigma$	2	1.693	0.39	0.674
$\iota \times \tau \times \varsigma$	1	0.241	0.06	0.813
$\kappa \times \iota \times \tau \times \varsigma$	2	0.159	0.04	0.964
Error	668	0.159		

Flies risen from inbred (isogenic) and outbred crosses of *Drosophila subobscura* reared at 18°C and 22°C. Karyotypes being compared are O_{ST}/O_{ST}, $O_{\underline{3+4}}/O_{\underline{3+4}}$, and $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$ ($\subset$ means "nested in").

4.4.3 Gene arrangement effects in the outbred lines

a) Thermal preference

The genetic and environmental (developmental temperature) contributions of chromosome *O* to T_p (and T_{ko}; below) was assessed from the outbred crosses including all possible karyotypes. Outbred crosses are obviously more relevant to the real situation because inbred genotypes are

homozygous for deleterious alleles, and also for alleles that might display heterozygote advantage in the original outbred population. The only statistically significant effects detected by the ANCOVA model (Table 4.18) were those arising from genetic differences among karyotypes (permutation tests corroborated that the three assayed karyotypes differ in T_p; P = 0.0018) and sexes, with females having a higher T_p (mean ± SD: 18.7°C ± 3.6°C) than males (18.0°C ± 3.6°C). As above, average T_p was slightly lower for flies reared at 18°C (18.1°C ± 4.0°C) than at 22°C (18.6°C ± 3.2°C), but the difference was marginally non-significant (P = 0.069).

The linear contrast between the two $O_{ST}/O^{*}_{\underline{3+4}}$ heterokaryotypes ($O^{*}_{\underline{3+4}}$ pools into a single class the arrangements that share $O_{\underline{3+4}}$; see Methods) reveals that $O_{ST}/O_{\underline{3+4}}$ and $O_{ST}/O_{\underline{3+4+8}}$ flies displayed a similar average T_p (18.5°C ± 3.8°C *vs*. 18.0°C ± 3.7°C, respectively). However, some differences were detected among the three $O^{*}_{\underline{3+4}}/O^{*}_{\underline{3+4}}$ karyotypes, which can be attributed to some under-dominance because average T_p for $O_{\underline{3+4}}/O_{\underline{3+4+8}}$ flies (18.1°C ± 3.4°C) was lower than that for the corresponding homokaryotypes ($O_{\underline{3+4}}/O_{\underline{3+4}}$: 18.9°C ± 3.5°C; $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$: 19.3°C ± 3.6°C). In any case, the main difference was between O_{ST} and $O^{*}_{\underline{3+4}}$ carriers, with mainly additive genetic effects (Figure 4.22). As already indicated, O_{ST}/O_{ST} flies clearly preferred lower temperatures than $O_{\underline{3+4}}/O_{\underline{3+4}}$ or $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$ flies.

Table 4.18: Karyotype and temperature effects on thermal preference.

Source of variation	d.f.	Mean Square	F	P
Covariate (plate hour)	1	147.947	11.84	<0.001
Karyotype (κ)	5	60.774	4.97	0.002
O_{ST}/O^{*}_{3+4}	1	0.853	0.07	0.793
O_{3+4}/O^{*}_{3+4}	2	42.884	3.51	0.043
O_{ST}/O_{ST} , O_{ST}/O^{*}_{3+4} , O^{*}_{3+4}/O^{*}_{3+4}	2	106.330	8.70	0.001
additive effect	1	205.854	16.85	<0.001
dominance effect	1	3.532	0.29	0.595
Cross $\subset \kappa$	30	12.220	0.98	0.502
Temperature (τ)	1	41.328	3.31	0.069
Sex (ς)	1	91.221	7.30	0.007
$\kappa \times \tau$	5	19.791	1.58	0.162
$\kappa \times \varsigma$	5	10.805	0.86	0.505
$\tau \times \varsigma$	1	4.948	0.40	0.529
$\kappa \times \tau \times \varsigma$	5	8.863	0.71	0.617
Error	691	12.498		

Flies risen from inbred (isogenic) and outbred crosses of *Drosophila subobscura* reared at 18°C and 22°C. Karyotypes being compared are O_{ST}/O_{ST}, $O_{\underline{3+4}}/O_{\underline{3+4}}$, and $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$. $O^{*}_{\underline{3+4}}$ stands for $O_{\underline{3+4}}$ + $O_{\underline{3+4+8}}$. ($\subset$ means "nested in").

b) Knock out temperature

The ANCOVA for T_{ko} (Table 4.19) did not detect any difference among karyotypes, in accordance with the previous findings for the inbred crosses. Similarly, the main differences arose between developmental temperature (flies reared at 18°C: 33.6°C ± 1.9°C; flies reared at 22°C: 32.8°C ±

2.3°C) and sex (females: 33.7°C ± 1.8°C; males: 32.7°C ± 2.4°C). The genetic correlation between T_p and T_{ko} after pooling $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ was r_p =-0.130 (P = 0.917).

Table 4.19: Karyotype and temperature effects on knockout temperature.

Source of variation	d.f.	Mean Square	*F*	*P*
Covariate (plate hour)	1	101.377	25.87	<0.001
Karyotype (κ)	5	4.295	0.57	0.724
O_{ST}/O^*_{3+4}	1	11.598	1.52	0.228
O_{3+4}/O^*_{3+4}	2	0.016	0.002	0.998
O_{ST}/O_{ST}, O_{ST}/O^*_{3+4}, O^*_{3+4}/O^*_{3+4}	2	4.872	0.64	0.536
additive effect	1	0.015	0.001	0.965
dominance effect	1	8.632	1.13	0.296
Cross $\subset \kappa$	30	7.641	1.95	0.002
Temperature (τ)	1	107.075	27.33	<0.001
Sex (ς)	1	180.874	46.16	<0.001
$\kappa \times \tau$	5	7.576	1.93	0.087
$\kappa \times \varsigma$	5	8.777	2.24	0.049
$\tau \times \varsigma$	1	1.650	0.42	0.517
$\kappa \times \tau \times \varsigma$	5	2.329	0.59	0.704
Error	654	3.918		

Flies risen from inbred (isogenic) and outbred crosses of *Drosophila subobscura* reared at 18°C and 22°C. Karyotypes being compared are O_{ST}/O_{ST}, $O_{\underline{3+4}}/O_{\underline{3+4}}$, and $O_{\underline{3+4+8}}/O_{\underline{3+4+8}}$. $O^*_{\underline{3+4}}$ stands for $O_{\underline{3+4}}$ + $O_{\underline{3+4+8}}$. ($\subset$ means "nested in").

Again, the conclusion is that these two traits are uncorrelated. Figure 4.22 plots the genotypic values in the additive-dominance scales for T_p and T_{ko}, together with their statistical significance obtained from the appropriate contrasts (Table 4.18, 4.19).

c) Average effects on thermal preference

Our experiment only provides an estimation of the gene (chromosome *O*) action on T_p and does not allow inferences to the base population. It is possible, however, to obtain estimates of the average effects, or "statistically additive effects", by taking into account the gene action and allelic (chromosome arrangement) frequencies in the natural populations (FALCONER and MACKAY 1996). Assuming that the chromosome arrangement effects are roughly the same along the cline (for a measure of climatic temperatures along the Palaearctic cline see Figure 1 in BALANYÀ *et al.* 2006), Table 4.20 gives the average effects (females and males pooled) estimated from the frequencies of the different arrangements in European populations spanning about 17° latitude (SOLÉ *et al*. 2002, BALANYÀ *et al.* 2004). The interpretation is that flies inheriting O_{ST} chromosome will choose a temperature ranging from around 0.31°C - 0.45°C below the average temperature chosen by the population (conversely, flies carrying warm-climate chromosome arrangements will choose a temperature ranging from around 0.03°C - 0.52°C above the average).

Figure 4.22: Karyotypic values in the additive-dominance scale. Deviation values for thermal preference (T_p) and knockout temperature (T_{ko}) were measured after pooling arrangements $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ into a single class ($O^*_{\underline{3+4}}$), and the coordinate point (0, 0) was taken as the midparent (i.e., the average of T_p and T_{ko} for the two karyotypes O_{ST}/O_{ST} and $O^*_{\underline{3+4}}/O^*_{\underline{3+4}}$). Females (upper panel) and males (lower panel) are plotted separately because the interaction karyotype × sex was statistically significant for T_{ko} (Table 4.19). In the original scale the (0, 0) point corresponds to an average T_p of 18.31°C for females and 17.91°C for males, and an average T_{ko} of 33.58°C for females and 32.61°C for males. Open squares give the values for all six karyotypes to appreciate their dispersion from the midparent, as well as their dispersion from the pooled O_{ST}/ O^*_{3+4} and O^*_{3+4}/O^*_{3+4} karyotypes (black circles). Statistical significance for additive (a_{Tp}, a_{Tko}) and dominance (d_{Tp}, d_{Tko}) effects are given in Tables 4.18 and 4.19. Note also that the phenotypic ($r_{Tp \cdot Tko}$ = -0.030) and genetic (r_k = -0.068, r_p = -0.130; see Methods) correlations were non-significantly different from zero (see text for details).

Table 4.20: Average effect of chromosome *O* on thermal preferences (°C).

		Frequency		Average effect	
Population	Coordinates	O_{ST}	O^*_{3+4}	O_{ST}	Rest
Málaga (Spain)	36°43'N–4°25'W	0.080	0.407	-0.4506	0.0392
Punta Umbría (Spain)	37°10'N–6°57'W	0.066	0.410	-0.4494	0.0318
Calviá (Spain)	39°33'N–2°29'E	0.057	0.590	-0.4485	0.0271
Riba-roja (Spain)	39°33'N–0°34'W	0.148	0.324	-0.4530	0.0787
Queralbs (Spain)	42°13'N–2°10'E	0.290	0.493	-0.4395	0.1795
Lagrasse (France)	43°05'N–2°37'E	0.330	0.590	-0.4312	0.2124
Montpellier (France)	43°36'N–3°53'E	0.362	0.557	-0.4232	0.2401
Villars (France)	45°26'N–0°44'E	0.389	0.581	-0.4155	0.2645
Leuk (Switzerland)	46°19'N–7°39'E	0.595	0.365	-0.3267	0.4800
Vienna (Austria)	48°13'N–16°22'E	0.625	0.270	-0.3095	0.5158
Tübingen (Germany)	48°32'N–9°04'E	0.606	0.351	-0.3205	0.4930
Louvain-la-Neuve (Belgique)	50°43'N–4°37'E	0.397	0.540	-0.4130	0.2719
Groningen (The Netherlands)	53°13'N–6°35'E	0.502	0.405	-0.3733	0.3763

$O^*_{\underline{3+4}}$ pools gene arragements $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ used in the present work. Together with O_{ST}, their combined frequency is ≥ 0.90 in central European populations and drops to approximately 0.50 in south-western Europe, where arrangement $O_{\underline{3+4}+7}$ is also frequent. However, from previous data (REGO *et al.* 2010) no difference in T_p is detected between $O_{\underline{3+4}+7}$ and $O^*_{\underline{3+4}}$, which justifies their pooling and allows estimating average effects assuming two gene arrangements: O_{ST} and the rest. Gene arrangement frequencies where taken from the "new collections" in SOLÉ *et al.* (2002) and BALANYÀ *et al.* (2004).

Combined with our previous results with chromosome *A* (which is the sex chromosome and additive values can be estimated using males' T_p; REGO *et al.* 2010), where gene arrangement A_{ST} exhibits a similar latitudinal pattern as O_{ST} and flies carrying A_{ST} also display a laboratory thermal preference towards colder temperature, the conclusion is that flies inheriting simultaneously A_{ST} and O_{ST} will choose temperatures ranging from approximately 0.5°C - 1.0°C below the average (these estimates assume perfect additivity).

The present results with isogenic lines and their crosses corroborate and extend our previous work with wild flies from south-western Europe (REGO *et al.* 2010). They confirm that arrangements on chromosome *O* have a biometrical effect on thermal preference in a laboratory temperature gradient, with cold-climate O_{ST} carriers displaying a lower T_p than their warm-climate $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ counterparts. In addition, T_p and T_{ko} were again found to be uncorrelated, and we can now discard a potential genetic covariance between both traits arising from linkage disequilibrium between genes affecting thermal preference and candidate genes for heat shock resistance (i.e., *Hsp68* and *Hsp70*; MOLTÓ *et al.* 1992, CUENCA *et al.* 1998) located inside, or close to, the chromosome regions covered by the inversions analyzed here (see Introduction). In other words, we conclude that variation on *O* chromosome arrangements does not have any effect on knock out temperature (but see below). Note, however, that this does not imply that genes on chromosome *O* have no effect on T_{ko} (actually, statistically significant differences were detected among crosses

within karyotypes; Table 4.19); it simply indicates that any allelic variation of putative genes influencing this trait is not in linkage disequilibrium with inversions on this chromosome.

The new findings were: (i) a lack of inbreeding depression for both T_p and T_{ko}; (ii) a lack of phenotypic plasticity for T_p according to the temperature at which the flies were raised (18°C and 22°C); and (iii) a substantial effect of developmental temperature on T_{ko}. The absence of inbreeding depression for T_p agrees with the genetic analysis from outbred flies, where a dominance effect after pooling chromosome arrangements $O_{\underline{3+4}}$ and $O_{\underline{3+4+8}}$ into a single class ($O^*_{\underline{3+4}}$) was absent (Table 4.18; note that the differences detected among the three $O^*_{\underline{3+4}}/O^*_{\underline{3+4}}$ karyotypes, and attributed to some underdominance, could not be appreciated in the inbreeding analysis because it only included inbred and outbred homokaryotypes). On the other hand, the lack of inbreeding depression for T_{ko} is expected and does not mean anything here, simply because no "gene" effects linked to chromosomal arrangements on chromosome *O* were detected. At first sight this might be surprising because a well-characterized cellular defense mechanism once environmental temperature approaches the upper thermal limits is the heat shock response, and in *D. melanogaster* the major inducible heat shock protein *Hsp70* appears to be the primary protein involved in thermotolerance (PARSELL, TAULIEN and LINDQUIST 1993; SØRENSEN, KRISTENSEN and LOESCHCKE 2003). Recent work, however, questions the pervasive role of *Hsp70* in the mediation of the heat stress response and suggests that it may be life-stage specific, being important in larvae but not in adults (JENSEN *et al.* 2010). Our results are apparently consistent with the lack of association between *Hsp70* and adult heat resistance (but see further discussion below), although also raise a caveat to the conclusion that there is no covariance between T_p and T_{ko}. Thus, it could be the case that *Hsp70* variation across karyotypes is associated with juvenile tolerance to heat stress, an important trait in *Drosophila* particularly in summer when larval feeding patches can become lethally hot (FEDER, BLAIR and FIGUERAS 1997). This possibility warrants further analysis.

An important concern here is that *Hsp70* production might not be inducible in the dynamic experimental protocol we used to estimate upper thermal tolerance, where temperature increased 0.1°C min^{-1}. One apparently compelling reason for this is that the estimated maximum thermal limits that *D. melanogaster* can tolerate decrease from approximately 39.9°C with heating rate 0.5°C min^{-1} to 38.7°C with heating rate 0.1°C min^{-1} (CHOWN *et al.* 2009), a puzzling result because slower heating rates should allow individuals to acclimatize to new temperatures and also because

slow heating rates pre-exposes individuals to non-lethal high temperatures ("hardening"), which increases heat shock resistance (HOFFMANN, SØRENSEN and LOESCHCKE 2003). We have recently discussed why these conflicting outcomes arise, and suggest that the contribution of other stressors (e.g. higher desiccation in long thermal tolerance assays associated with slow warming rates) can potentially overshadow thermal acclimation effects in dynamic assays with varying heating rates (REZENDE, TEJEDO and SANTOS 2011). In other words, we challenge the idea that induced thermotolerance does not occur in dynamic assays with slow heating rates. At this stage this is just speculative because *Hsp70* production was not measured in our flies, but the problem is important because *Drosophila* adults are likely to experience slow heating rates in nature of 0.06 - 0.1°C min^{-1} (CHOWN *et al.* 2009, MITCHELL and HOFFMANN 2010) and further empirical studies are required to explain the apparently inconsistent findings.

Considering that inversion polymorphisms in other chromosomes associated with thermotolerance exhibit latitudinal variation in the expected direction (REGO *et al*. 2010), it is unclear why different arrangements in chromosome *O* seem to have negligible effects on this trait in our flies. There are at least three possible explanations. First, it might be the case that there is not enough genetic differentiation for *Hsp*70 between $O_{\underline{3+4}}$ and O_{ST} chromosome arrangements. Second, assuming that there is indeed genetic differentiation it might happen that there is no association between *Hsp70* protein levels and adult thermotolerance in *D*. *subobscura*, in accordance with what has been recently suggested to occur in *D*. *melanogaster* (JENSEN *et al*. 2010). Third, flies carrying different arrangements may differ in their plasticity to heat shock, resulting in contrasting patterns of basal and induced heat tolerance (see REZENDE, TEJEDO and SANTOS 2011). Consequently, the initial assessment of the co-adaptation hypothesis could be somewhat flawed if the potential fitness benefits of heat-induced thermotolerance under extreme field conditions (LOESCHCKE and HOFFMANN 2007) were not appropriately assessed in our experiments. These alternatives were investigated in the other study of CALABRIA *et al.* (2012) where it was assessed whether or not warm- and cold-climate chromosomal arrangements in *D*. *subobscura* are differentiated for *Hsp70* protein expression levels. It was shown that flies carrying the warm-climate chromosome arrangement $O_{\underline{3+4}}$ have higher basal protein levels of *Hsp70* than their cold-climate O_{ST} counterparts, but this difference disappears after heat hardening. $O_{\underline{3+4}}$ carriers are also more heat tolerant, although it is difficult to conclude from our results that this is causally linked to their higher basal levels of *Hsp70*. The observed patterns were consistent with the thermal co-adaptation

hypothesis, and suggest that the interplay between behaviour and physiology underlies latitudinal and seasonal shifts in inversion frequencies.

The pooled average T_p here was (mean ± SD) 18.4°C ± 3.6°C (T_{set}: 15.4°C - 21.2°C; these figures include only outbred lines) and about the same at both rearing temperatures. The difference with the previous estimate for wild-flies from Adraga (16.6°C, T_{set}: 12.4°C - 20.4°C; REGO *et al*. 2010) does not seem to be overreached, and could be partially explained by the fact that the present flies were genetically homogeneous for all chromosomes from the *ch-cu* marker strain but chromosome *O* (recall that the sex chromosome *A* also had a significant effect on T_p; REGO *et al.* 2010). This strain has a long history of maintenance at 18°C in the laboratory. In any case, our estimates remain substantially lower than that from Huey and Pascual (23.7°C, T_{set}: 21.2°C - 25.9°C; HUEY and PASCUAL 2009), and the difference cannot be accounted by flies' rearing temperature. No reasonable explanation for the discrepancy can be offered at this moment, but the additional result that developmental temperature substantially affected T_{ko} makes us confidently conclude that our estimates are indeed closer to the actual T_p of the species. Flies reared at 22°C showed lower heat resistance than their counterparts reared at 18°C (32.8°C *vs.* 33.6°C; outbred lines), which could be a consequence of their smaller size due to the inverse relationship between body size and developmental temperature (MORETEAU *et al.* 1997; SANTOS, BRITES and LAAYOUNI 2006). However, resistance to heat does not seem to be associated with body size (BUBLI, IMASHEVA and LOESCHCKE 1998). The association between T_{ko} and wing size have been also analyzed from the previous experiment in our laboratory where both traits were recorded (FRAGATA *et al.* 2010) and found no relationship whatsoever. Most likely, 22°C was a suboptimal and potentially stressful temperature for our flies, making them to be weaker and less resistant to the heat shock. Note, however, that this conclusion might not be extrapolated to wild flies that harbor higher levels of genetic variability than our chromosomal lines.

To interpret the interplay between thermal preference and heat stress resistance, an understanding of the environmental temperatures experienced by *D. subobscura* along climatic gradients is required. As far as we are aware, the only data available on T_b for active flies along a latitudinal gradient (spanning 12°) come from the work of HUEY and PASCUAL (2009) in western North America. They found that mean T_b varies by as much as 21°C (from 8°C to 29°C), and that the temporal activity of flies during the day did not match predictions from optimal temperature

regulation or desiccation avoidance. Temperatures of maximum activity in summer (Figure 2 in HUEY and PASCUAL 2009) - when wild flies are smaller probably due to their higher developmental temperatures and/or crowding conditions (KARI and HUEY 2000); and crowding is known to affect adult thermal stress resistance in *Drosophila* (SØRENSEN and LOESCHCKE 2001) – are dangerously close to the T_{ko} obtained here for the outbred flies raised at 22°C. This suggests that active *D. subobscura* flies can experience extreme conditions in the wild, and one would expect flies' activity to be correlated with heat resistance under these conditions if behavior and physiology were coadapted. Some evidence indicates that diurnal activity patterns in summer can vary according to inversion polymorphism, and chromosome arrangements on the *O* chromosome seem to behave as expected from our data: O_{ST} is more frequent towards the evening while chromosomes carrying gene arrangement $O_{\underline{3+4}}$ are most frequently sampled at midday (SAVKOVIC, STAMENKOVIC-RADAK and ANDJELKOVIC 2004). This behavioral thermoregulation, however, would not confer less susceptibility to high temperatures because the genetic basis of both traits does not seem to allow for the building up of "coadaptation". It is well known from basic population genetics theory that genetic covariance between traits can arise when alleles at different loci are associated (linkage disequilibrium), and this critically depends on relatively low recombination rates (HEDICK 2000). The lack of association between T_p and T_{ko} in *D. subobscura* is fully consistent with their genetic basis as independently segregating chromosomes are involved (REGO *et al.* 2010). Yet, a correlation between these traits can be expected at the interpopulational level due to patterns of correlated selection (rather than genetic correlations) across a latitudinal gradient because of the congruent latitudinal clinal variation for chromosome arrangements on the *E* (which influences T_{ko}; REGO *et al.* 2010), and on the *A* and *O* chromosomes (which influence T_p; REGO *et al.* 2010, this work).

We now speculate that the mismatch between T_p and T_{ko} could apparently generate an interesting dynamics in the population frequencies of different chromosome arrangements on chromosome *O*. Suppose the daily activity of flies in the warmest months of the year follows the previously described pattern; that is, flies carrying gene arrangement $O_{\underline{3+4}}$ are more active at midday and, therefore, have a higher risk of a heat shock than O_{ST} and are selected against. On the other hand, assuming T_p corresponds closely with temperatures that maximize fitness $O_{\underline{3+4}}$ flies likely enjoy a fitness advantage in summer. The net effect would be a compromise between "behavior unresponsiveness" and general performance, which means that chromosome arrangements on chromosome *O* may or may not cycle seasonally according to average environmental temperature

(i.e., O_{3+4} could be expected to increase in frequency in summer and decrease in winter if general performance is what matters). Interestingly, both patterns have been detected: consistent seasonal cycling at a north-western population in Spain (RODRÍGUEZ-TRELLES, ALVAREZ and ZAPATA 1996) and apparently no seasonal variation at a north-eastern population also in Spain (DE FRUTOS and PREVOSTI 1984). The point here is that parallel seasonal changes should also be detected for chromosome *A* since it also affects T_p (REGO *et al.* 2010). In accordance with this prediction, no seasonal cycling was detected for chromosome A in the northeastern population, but unfortunately no information is available for the other population because chromosome *O* was the only chromosome scored. It would be very interesting to see what happens for chromosome *A* in the cycling population.

For ectotherms facing spatiotemporal variation in environmental temperature theory predicts that a coevolution between thermal preference and physiological performance can occur (ANGILLETTA 2009). In the widespread species *D. subobscura* behavioral thermoregulation and heat tolerance are "coadapted" in the sense that flies carrying cold-climate (warm-climate) chromosome arrangements tend to choose colder (warmer) temperatures and have lower (higher) heat stress tolerance (REGO *et al.* 2010). We have analyzed the genetic basis of these thermal traits using isochromosomal lines for the *O* chromosome. This chromosome was known to affect thermal preference (REGO *et al.* 2010), and also harbours several genes involved in the heat shock response (*Hsp68* and *Hsp70;* MOLTÓ *et al.* 1992; CUENCA *et al.* 1998). These genes are located inside of, or close to, the chromosome regions covered by inversions that show conspicuous northwest-southwest latitudinal clines in Palaearctic populations, as well as seasonal fluctuations that are in agreement with the latitudinal patterns (REZENDE *et al.* 2010). Our results corroborate that arrangements on chromosome *O* affect adult thermal preference: flies inheriting the cold-climate O_{ST} chromosome are predicted to choose a temperature around 0.31°C - 0.45°C below the average temperature chosen by the population and, conversely, flies inheriting the warm-climate O_{3+4} and O_{3+4+8} chromosomes are expected to choose a temperature ranging from around 0.03°C - 0.52°C above the average. However, these chromosome arrangements did not have any differential effect on adult heat tolerance. We conclude that thermal preference and heat tolerance in *D. subobscura* appear to be genetically independent and, therefore, any latitudinal correlation between both traits would likely reflect a pattern of correlated selection across populations rather than within-population genetic correlations.

PART 5

CONCLUSIONS

"That is not said right," said the Caterpillar.

"Not quite right, I'm afraid," said Alice, timidly; "some of the words have got altered."

"It's wrong from beginning to end," said the Caterpillar decidedly, and there was silence for some minutes.

Lewis Carroll,

Alice's Adventures in Wonderland

CONCLUSIONS

1. Latitudinal clines were detected for chromosomal arrangements O_{ST}, $O_{\underline{3+4}}$ and $O_{\underline{3+4}+7}$. In $O_{\underline{3+4+8}}$ arrangement, the sign of the regression coefficients with latitude were the same as previously described despite no significant clines were found.

2. Significant shifts in frequencies for "warm-" and "cold-adapted" gene arrangements were found for the populations of Montpellier and Valencia with former increasing and last decreasing as a probable consequence of climate change.

3. Thirty candidate genes for thermal adaptation were localized on the *O* chromosome: nine in *Segment I* and 21 in *Segment II*.

4. When comparing the physical positions of the 52 gene regions mapping in the *O* chromosome of *D. subobscura* with the homologous chromosomes of *D. melanogaste*r and *D. pseudoobscura* a poor synteny and high reorganization in the order of markers were found, but chromosomal elements were highly conserved.

5. The genetic content for a particular chromosomal arrangement is the same in two populations. Therefore, gene flow among populations is high.

6. However, significant genetic differentiation was found when comparing three different arrangements in all gene regions, the O_{ST} was the most differentiated one. Recombination networks when using the concatenated dataset, regulatory region of *larp* and *Acph-1* showed three clades corresponding to the three arrangements.

7. There were no correlation between nucleotide diversity or genetic differentiation and the distance of gene location from inversion break point.

8. Linkage disequilibrium among nucleotide pairs was high along the whole inversion and despite of the distance between them.

9. Recombination events were found in each gene region along the entire inversion independently from the distance to the nearest inversion break point.

10. The population expansion was detected in almost all gene regions especially for O_{3+4} arrangement.

11. Footprints of significant positive selection were found for the coding positions of gene *Acph-1* in O_{3+4} arrangement.

12. The Local Adaptation hypothesis fits better to our data over Coadaptation hypothesis and would explain the maintenance of the *O* chromosome arrangements in *D. subobscura*.

13. The age of three chromosomal arrangements was quite similar, varying due to selected gene region and time of divergence with *D. pseudoobscura* implemented. The O_{3+4} arrangement was slightly older (0.95 Myr or 0.43 Myr on average) than O_{ST} and O_{3+4+8}, which apparently arose at the same time (0.87 Myr or 0.39 Myr; 0.86 or 0.39 Myr on average respectively).

14. The results on thermal traits essay corroborate and extend the previous work with wild flies from south-western Europe: Cold-climate O_{ST} displays a lower T_p than warm-climate O_{3+4} and O_{3+4+8} and T_p and T_{ko} were again found to be uncorrelated.

15. The lack of inbreeding depression for both T_p and T_{ko} and lack of phenotypic plasticity for T_p according to the temperature at which the flies were raised (18ºC and 22ºC) were found.

16. There was a substantial effect of developmental temperature on T_{ko}.

BIBLIOGRAPHY

ADAMS, M.A., CELNIKER, S.E., HOLT, R.A. *et al.* 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**:2185–95.

AGUADÉ, M. 1988a. Nucleotide sequence comparison of the *rp49* gene region between *Drosophila subobscura* and *D. melanogaster*. *Mol Biol Evol* **5**(4):433-41.

AGUADÉ, M. 1988b. Restriction map variation at the *Adh* locus of *Drosophila melanogaster* in inverted and noninverted chromosomes. *Genetics* **119**:135–40.

AGUADÉ, M. 1989**.** Variation in natural populations of *Drosophila* as revealed by four-cutter analysis. *Genome* **31**:784-7.

ALEXANDER, L.V., ZHANG, X., PETERSON, T.C., CAESAR, J., GLEASON, B., KLEIN TANK, A., HAYLOCK, M., COLLINS, D., TREWIN, B., RAHIMZADEH, F., TAGIPOUR, A., AMBENJE, P., RUPA KUMAR, K., REVADEKAR, J., GRIFFITHS, G., VINCENT, L., STEPHENSON, D., BURN, J., AGUILAR, E., BRUNET, M., TAYLOR, M., NEW, M., ZHAI, P., RUSTICUCCI, M., VAZQUEZ-AGUIRRE, J.L. 2006. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research* **111**:D05109.

ÁLVAREZ, G., MARTÍNEZ, P., ZAPATA, C., SANTOS, M., FONTDEVILA, A. 1981 Genetic analysis of modifier variability in *Drosophila subobscura*. *Experientia* **37:**1150-51.

ANANINA, G., PEIXOTO, A.A., BITNER-MATHÉ, B.C., SOUZA, W.N., BASSO DA SILVA, L., VALENTE, V.L.S., KLACZKO, L.B. 2004. Chromosomal inversion polymorphism in *Drosophila mediopunctata*: seasonal, altitudinal, and latitudinal variation. *Genetics and Molecular Biology* **27**(1):61-69.

ANDERSON, W.W., ARNOLD, J., BALDWIN, D.G., BECKENBACH, A.T., BROWN, C.J., BRYANT, S.H., COYNE, J.A., HARSHMAN, L.G., HEED, W.B., JEFFERY, D.E. 1991. Four decades of inversion polymorphism in *Drosophila pseudoobscura*. *PNAS* **88**:10367-71.

ANDOLFATTO, P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**(7062):1149-52.

ANDOLFATTO, P. 2008. Controlling type-I error of the McDonald-Kreitman test in genome wide scans for selection on noncoding DNA. *Genetics* **180**(3):1767-71.

ANDOLFATTO, P., DEPAULIS, F., NAVARRO, A. 2001. Inversion polymorphism and nucleotide variability in *Drosophila*. *Genet Res* **77**:1-8.

ANDOLFATTO, P, NORDBORG, M. 1998. The effect of gene conversion on intralocus associations. *Genetics* **148**(3):1397-9.

ANDOLFATTO, P., WALL, J.D., KREITMAN, M. 1999. Unusual haplotype structure at the proximal breakpoint of *In(2L)t* in a natural population of *Drosophila melanogaster*. *Genetics* **153**:1297–311.

ANGILLETTA, M.J. 2009.Thermal Adaptation. A Theoretical and Empirical Synthesis. *Oxford: Oxford Univ Press*.

ARADOTTIR, G.I., ANGUS, R.B. 2004. A chromosomal analysis of some water beetle species recently transferred from *Agabus* Leach to Ilybius Erichson, with particular reference to the variation in chromosome number shown by *I. montanus* Stephens (*Coleoptera: Dytiscidae*). *Hereditas* **140**:185-92.

ARAÚZ, P.A., MESTRES, F., PEGUEROLES, C., ARENAS, C., TZANNIDAKIS, G., KRIMBAS, C.B., SERRA L. 2008 Tracking the origin of the American colonization by *Drosophila subobscura*: genetic comparison between Eastern and Western Mediterranean populations. *J Zool Syst Evol Res* **47**(1):25-34.

ARAÚZ, P.A. 2009. Estudio comparativo de poblaciones del Mediterráneo occidental y oriental de *Drosophila subobscura* mediante marcadores genéticos y su relación con la colonización de América. *Tesis Doctoral. Departament de Genètica, Universitat de Barcelona.*

ARAÚZ, P.A., PEGUEROLES, C., ZIVANOVIC, G., FERNÁNDEZ-IRIARTE, P.J., RUIZ-MARTÍN, H., BALANYÀ, J., SERRA, L., MESTRES, F. 2009. Recent origin of a chromosomal inversion revealed by its association with nucleotide haplotypes. *Hereditas* **146**(5):257-9.

ARAÚZ, P.A., PERIS-BONDIA, F., LATORRE, A., SERRA, L., MESTRES, F. 2011. Molecular evidence to suggest the origin of a colonization: *Drosophila subobscura* in America. *Genetica* **139**:1477–86.

ARBOLEDA, C.E. 2008. Polimorfismo y divergencia en genes implicados en la resistencia al frio en *Drosophila*. *Tesis Doctoral. Departament de Genética, Universitat de Barcelona.*

ARGEMÍ, M., MESTRES, F., PREVOSTI, A., SERRA, L. 2003. Microevolutionary dynamics of a community of *Drosophilids. J Zool Syst Evol Res* **41**(1):57-63.

AYALA, F.J., COLUZZI, M. 2005. Chromosome speciation: Humans, *Drosophila* and mosquitoes. *PNAS* **102**:6535-42.

AYALA, F.J., SERRA, L., PREVOSTI, A. 1989. A grand experiment in evolution: the *Drosophila subobscura* colonization of the America. *Genome* **31**:246-55.

BALANYÀ, J., HUEY, R.B., GILCHRIST, G.W., SERRA, L. 2009. The chromosomal polymorphism of *Drosophila subobscura*: a microevolutionary weapon to monitor global change. *Heredity* **103**:364-67.

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**:1773-5.

Balanyà, J., Serra, L., Gilchrist, G.W., Huey, R.B., Pascual, M., Mestres, F., Solé, E. 2003. Evolutionary pace of chromosomal polymorphism in colonizing populations of *Drosophila subobscura*: An evolutionary time series. *Evolution* **57**:1837-45.

Balanyà, J., Solé, E., Oller, J.M., Sperlich, D., 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.

Ballard, J.W.O., Kreitman, M. 1994. Unraveling selection in the mitochondrial genome of *Drosophila*. *Genetics* **138**:757–72.

Beckenbach, A.T., Prevosti A. 1986. Colonization of North America by the European species *D. subobscura* and *D. ambigua*. *Am Midl Nat* **115:**1018.

Beckenbach, A.T., Wei, Y.W., Liu, H. 1993. Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. *Mol Biol Evol* **10**:619–34.

Betràn, E., Santos, M., Ruiz, A. 1998. Antagonistic pleiotropic effect of second chromosome inversions on body size and early life-history traits in *Drosophila buzzatii*. *Evolution* **52**:144-54.

Betràn, E., Rozas, J., Navarro, A., 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., Kim, I., Hoffmann, A.A., Feder, M.E. 2002. Response to natural and laboratory selection at the *Drosophila hsp70* genes. *Evolution* **56**:1796-801.

Bhutkar, A., Schaeffer, S.W., Russo, S.M., Xu, M., Smith, T.F., Gelbart, W.M. 2008. Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics* **179**:1657–80.

Blagden, S.P., Gatt, M.K., Archambault, V., Lada, K., Ichihara, K., Lilley, K.S., Inoue, Y.H., Glover, D.M. 2009. *Drosophila Larp* associates with poly(A)-binding protein and is required for male fertility and syncytial embryo development. *Dev Biol* **334**(1):186-97.

Bradshaw, W.E., Holzapfel, C.M. 2008. Genetic response to rapid climate change: it's seasonal timing that matters. *Molecular Ecology* **17**:157-66.

Brncic, D. 1972. Seasonal fluctuations of inversion polymorphism in *Drosophila flavopilosa* and the relationships with certain ecological factors. *Univ Texas Publ* **7213**:103-16.

Brncic, D., Budnik, M., 1980. Colonization of *Drosophila subobscura* Collin in Chile. *Drosoph Inf Serv* **55**:20.

Brncic, D., Prevosti, A., Budnik, M., Monclus, M., Ocaña, J. 1981. Colonization of *D. subobscura* in Chile. I. First population and cytogenetic studies. *Genetica* **56:**3–9.

Bubli, O.L., Imasheva, A.G., Loeschcke, V. 1998. Selection for knockdown resistance to heat in *Drosophila melanogaster* at high and low larval densities. *Evolution* **52**:619-25.

Burla, H., Götz, G. 1965. Veränderlichkeit der chromosomalen polymorphismus bei *Drosophila subobscura.* Genética **36:**83-104.

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

Cáceres, M., Puig, M., Ruiz, A. 2001. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res* **11**:1353–64.

Cáceres, M., Ranz, J.M., Barbadilla, A., Long, M., Ruiz, A. 1999. Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285**:415-18.

Cadevall, M. 1988. La estructura de la teoría de la evolución (Enrahonar. Monografies 3), *Publicacions de la Universitat Autonoma de Barcelona, Bellaterra.*

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

Calabria, G., Dolgova, O., Rego, C., Castañeda, L.E., Rezende, E.L., Balanyà, J., Pascual, M., Sørensen, J.G., Loeschcke, V., Santos, M. 2012. Hsp70 protein levels and thermotolerance in *Drosophila subobscura*: a reassessment of the thermal co-adaptation hypothesis. *J Evol Biol* **25**:91-700.

Caracristi, G., Schlötterer, C. 2003. Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Mol Biol Evol* **20**:792-9.

Carson, H.L. 1946. The selective elimination of inversion dicentric chromatids during meiosis in the egg of *Sciara impatiens*. *Genetics* **31**:95-113.

Carson, H.L. 1953. The effect of inversions on crossingover in *Drosophila robusta*. *Genetics* **38:**168-86.

Carson, H. L. 1958. The population genetics of *Drosophila robusta. Advanced Genentics* **9**:1-40.

Carson, H.L., Lockwood, J.P., Craddock, E.M. 1990. Extinction and recolonization of local populations on a growing shield volcano. *PNAS* **87**:7055-7.

Casals, F., Navarro, A. 2007. Chromosomal evolution: inversions: the chicken or the egg? *Heredity* **99**:479-80.

Casey, T.M., Hegel, J.R. 1981. Caterpillar setae: insulation for an ectotherm. *Science* **214**:1131-3.

Castañeda, L.E., Balanyà, J., Rezende, E., Santos, M. 2013. Vanishing chromosomal inversion clines in *Drosophila subobscura* from Chile: is behavioral thermoregulation to blame? *Am Nat* **182**(2), online in advance of print. doi:10.1086/671057

Charlesworth, J., Eyre-Walker, A. 2008. The McDonald-Kreitman Test and slightly deleterious mutations. *Mol Biol Evol* **25**(6):1007-15.

Chovnick, A. 1973. Gene conversion and transfer of genetic information within the inverted regions of inversion heterozygotes. *Genetics* **75**:123-31.

Chown, S.L., Jumbam, K.R., Sørensen, J.G., Terblanche, J.S. 2009. Phenotypic variance, plasticity and heritability estimates of critical thermal limits depend on methodological context. *Funct Ecol* **23**:133-40.

Cirera, S., Aguadé, M. 1998. Molecular evolution of a duplication: the sex-peptide (*Acp70A*) gene region of *Drosophila subobscura* and *Drosophila madeirensis. Mol Biol Evol* **15:**988–96.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B. *et al.* 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**:203–18.

Codina, M., Pérez, M.M. 1980. Primeros datos sobre la distribución estacional de las especies de *Drosophila* en Cataluña. *Bol Real Soc Esp Hist Nat* **78**:143-54.

Collier, G., MacIntyre, R.J., Featherston, D. 1990. Evolutionary relationships of enzymes in the genus *Drosophila*. In: Ecological and evolutionary genetics of *Drosophila* (eds. Baker, J.S.F., Starmer, W.T. and MacIntyre, R.J.). *Plenum Press, New York*:419-46.

Coluzzi, M. 1982. Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes. In Liss A. (Ed.). *Mechanisms of Speciation. New York*:143-53.

Comeron, J. M. 1997. Estudi de la variabilitat nucleotídica a *Drosophila*: Regió *Xdh* a *D. subobscura.* Ph.D. Thesis. Universitat de Barcelona.

Coyne, J.A., Bundgaard, J., Prout, T. 1983. Geographic variation of tolerance to environmental stress in *Drosophila pseudoobscura*. *Am Nat* **122**:474-88.

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**:113-26.

David, J.R. 1962. A new medium for rearing *Drosophila* in axenic conditions. *Dros Inf Serv* **36:**128.

Deutsch, C.A., Tewksbury, J.J., Huey, R.B, Sheldon, K.S., Ghalambor, C.K., Haak, D.C., Martin, P.R. 2008. Impacts of climate warming on terrestrial ectotherms across latitude. Thermal safety margin. *PNAS* **105**:6668-72.

Dillon, M.E., Wang, G., Garrity, P.A., Huey, R.B. 2009. Thermal preference in *Drosophila*. *J Therm Biol* **34**:109-19.

DOBZHANSKY, TH. 1937. Genetics and the Origin of Species. *Columbia Univ. Press, New York.*

DOBZHANSKY, TH. 1948. Genetics of natural populations XVIII. Experiments on chromosomes of *Drosophila pseudoobscura* from different geographic regions. *Genetics* **33**:588-602.

DOBZHANSKY, TH. 1950. Genetics of natural populations XIX. Origin of heterosis throught natural selection in populations of *Drosophila peusoobscura*. *Genetics* **35**:288-302.

DOBZHANSKY, TH. 1962. Rigid *vs*. flexible chromosomal polymorphisms in *Drosophila. Am Nat* **96**:321-8.

DOBZHANSKY, TH. 1970. Genetics of the Evolutionary Process. *New York: Columbia Univ Press.*

DOBZHANSKY, TH., AYALA, F.J. 1973. Temporal frequency changes of enzyme and chromosomal polymorphisms in natural populations of *Drosophila*. *PNAS USA* **70**:680-3.

DOBZHANSKY, TH., EPLING, C. 1948. The suppression of crossing-over in inversion heterozygotes of *Drosophila pseudoobscura*. *PNAS USA* **34**:137-41.

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.

DROSOPHILA 12 GENOMES CONSORTIUM. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*: **450**(7167):203-18.

DRUMMOND, A.J., ASHTON, B., BUXTON, S., CHEUNG, M., COOPER, A., DURAN, C., FIELD, M., HELED, J., KEARSE, M., MARKOWITZ, S., MOIR, R., STONES-HAVAS, S., STURROCK, S., THIERER, T., WILSON, A. 2011. Geneious v5.4, Available from http://www.geneious.com/.

DUBININ, N.P., TINIAKOV, G.G. 1946. Inversion gradients and natural selection in ecological races of *Drosophila funebris*. *Genetics* **31:**537-45.

DUCKWORTH, R.E. 2009. The role of behavior in evolution: a search for mechanism. *Evol Ecol* **23**:513-31.

DWORKIN, I. 2005. Canalization, cryptic variation and developmental buffering: a critical examination and analytical perspective. *Variation, A Central Concept in Biology. Elsevier. Chapter* **8**: 131-58.

EDGINGTON, E.S. 1995. Randomization Tests. 3rd edition. *New York: Marcel Dekker*.

EGEA, R., CASILLAS, S., BARBADILLA, A. 2008. Standard & Generalized McDonald and Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res* 36: W157–W162.

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**: 675-88.

Etterson, J.R., Shaw, R.G. 2001. Constraint to adaptive evolution in response to global warming. *Science* **294**:151-4.

Eyre-Walker, A. 2002. Changing effective population size and McDonald-Kreitman test. *Genetics* **162**(4):2017-24.

Falconer, D.S., Mackay, T.F. 1996. Introduction to Quantitative Genetics. 4th edition. *Essex, England: Longman*.

Feder, M.E., Blair, N., Figueras, H. 1997. Oviposition site selection: unresponsiveness of *Drosophila* to cues of potential thermal stress. *Anim Behav* **53**: 585-8.

Feder, M.E., Hofmann, G.E. 1999. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu rev physiol* **61**:243-82.

Fontdevila, A., Zapata, C., Alvarez, G., Sanchez, L., Méndez, J., Enriquez, I. 1983. Genetic coadaptation in the chromosomal polymorphism of *Drosophila subobscura*. I. Seasonal changes of gametic disequilibrium in a natural population. *Genetics* **105**:935-55.

Fragata, I., Balanyà, J., Rego, C., Matos, M., Rezende, E.L., Santos, M. 2010. Contrasting patterns of phenotypic variation linked to chromosomal inversions in native and colonizing populations of *Drosophila subobscura*. *J Evol Biol* **23**:112-23.

de Frutos, R., Kimura, K., Peterson, K.R. 1989. *In situ* hybridization of *Drosophila* polytene chromosomes with digoxigenin dUTP labeled probes. *Trends Genet* **5:**366.

de Frutos, R., Kimura, K., Peterson, K.R. 1990. *In situ* hybridization of *Drosophila* polytene chromosomes with digoxigenin dUTP labeled probes. *Methods Mol Cell Biol* **2**:32-6.

de Frutos, R., Prevosti, A. 1984. Temporal changes of chromosomal polymorphism in natural populations of *Drosophila subobscura*. *Genetica* **63**:181-7.

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

Gao, J.J., Watabe, H.A., Aotsuka, T., Pang, J.F., Zhang, Y.P. 2007. Molecular phylogeny of the *Drosophila obscura* species group, with emphasis on the Old World species. *BMC Evol Biol.* **7**:87.

Gibson, G., Wagner, G. 2000. Canalization in evolutionary genetics: a stabilizing theory? *BioEssays* **22**(4):372–80.

Gilbert, S.F. 2003. Developmental Biology, 7th Edition. *Sinauer Associates, Inc., Sunderland, MAISBN 0-87893-258-5*

Gilchrist, G., Jeffers, L., West, B., Folk, D., Suess, J., Huey, R.B. 2008. Clinal patterns of desiccation and starvation resistance in ancestral and invading populations of *Drosophila subobscura*. *Evol Appl* **1**: 513-23.

GILLESPIE, J.H. 1991. The causes of molecular evolution. *Oxford University Press, New York.*

GÓMEZ-BALDÓ, L., LATORRE, A., SERRA, L., MESTRES, F. 2008. Molecular variation in the *Odh* gene in Chilean natural populations of *Drosophila subobscura. Hereditas* **145**(4)*:*154-62.

GONZÁLEZ, J., RANZ, J.M., RUIZ, A. 2002. Chromosomal elements evolve at different rates in *Drosophila* genome. *Genetics* **161**:1137–54.

GOSTELI, M. 1990. Chromosomal polymorphism in natural populations of *Drosophila subobscura*. *Ph.D.Thesis, Zurich*.

GOSTELI, M. 1990. Chromosomal polymorphism in natural populations of *D. subobscura* near Zurich, Switzerland: a contribution to longterm comparisons. *Genetica* **81:**199–204.

GRIFFITHS, A.J.F, MILLER, J.H., SUZUKI, D.T., LEWONTIN, R.C., GELBART, W.M. 2000. An Introduction to Genetic Analysis. 7th edition. Inversions*. New York: W. H. Freeman.*

HALL, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**:95-8.

HARPENDING, H.C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* **66**:591–600.

HARPENDING, H.C., SHERRY, S.T., ROGERS, A.R., STONEKING, M. 1993. The genetic structure of ancient human populations. *Curr Anthropol* **34**:483–96.

HASEGAWA, M., CAO, Y., YANG, Z. 1998. Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. *Mol Biol Evol* **15**:1499-1505.

HASSON, E., EANES, W.F. 1996. Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. *Genetics* **144**:1565–75.

HEDICK, P.W. 2000. Genetics of Populations. 2nd edition. *Boston: Jones and Bartlett Publ*.

HERTZ, P.E., HUEY, R.B., STEVENSON, R.D. 1993. Evaluating temperature regulation by field-active ectotherms: the fallacy of the inappropriate question. *Am Nat* **142**:796-818.

HILL, W.G., ROBERTSON, A. 1966. The effect of linkage on limits to artificial selection. *Genet Res* **8**:369-294.

HILL, W.G., ROBERTSON, A. 1968. Linkage disequilibrium in finite populations. *Theor Appl Genet* **38**:226–31.

HILLIKER, A.J., HARAUZ, G., REAUME, A.G., GRAY, M., CLARK, S.H., *CHOVNICK, A.* 1994. Meiotic gene conversion tract length distribution within the *rosy* locus of *Drosophila melanogaster. Genetics* **137:**1019-26.

Hoffmann, A.A. 2010. Physiological climatic limits in *Drosophila*: patterns and implications. *J Exp Biol* **213**:870-80.

Hoffmann, A.A., Parsons, P.A. 1991. Evolutionary Genetics and Environmental Stress. *Oxford University Press, New York.*

Hoffmann, A.A., Riesenberg, **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**:21-4.

Hoffmann, A.A., Sgro, C.M., Weeks, A.R. 2004. Chromosomal inversion polymorphisms and adaptation. *Trends in Ecology and Evolution* **19:**482-8.

Hoffmann, A.A., Sørensen, J.G., Loeschcke, V. 2003. Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *J Therm Biol* **28**:175-216.

Hoffmann, A.A., Weeks, A.R. 2007. Climatic selection on genes and traits after 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica* **129**:133-47.

Hofmann, G.E., Todgham, A.E. 2010. Living in the now: physiological mechanisms to tolerate a rapidly changing environment. *Annu Rev Physiol* **72**: 127-45.

Hudson, R.R. 1990. In Oxford Surveys in Evolutionary Biology, eds. Futuyma, D.J., Antonovics, J. *Oxford Univ. Press, Oxford* **7**:1-44.

Hudson, R.R. 2000. A new statistic for detecting genetic differentiation. *Genetics* **155**:2011–4.

Hudson, R.R. 2001. Two-locus sampling distributions and their application. *Genetics* **159**:1805-17.

Hudson, R.R., Boos, D.D., Kaplan, N.L. 1992. A statistical test for detecting geographic subdivision. *Mol Biol Evol* **9**:138–51.

Hudson, R.R., Slatkin, M., Maddison, W.P. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**:583–9.

Huey, R.B., Bennett, A.F. 1987. Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. *Evolution* **41**:1098-115.

Huey, R.B., Hertz, P.E., Sinervo, B. 2003. Behavioral drive versus behavioral inertia in evolution: a null model approach. *Am Nat* **161**:357-66.

Huey, R.B., Pascual, M. 2009. Partial thermoregulatory compensation along a latitudinal cline in a rapidly evolving invasive species. *Drosophila subobscura* in North America. *Ecology* **90**:1715-20.

Huey, R.B., Stevenson, R.D. 1979. Integrating Thermal Physiology and Ecology of Ectotherms: A Discussion of Approaches. *Am Zool* **19**:357-66.

HUSON, D.H., BRYANT, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23** (2):254–67. Software available from www.splitstree.org.

IBNSOUDA, S., SCHWEISGUTH, F., DE BILLY, G., VINCENT, A. 1993. Relationship between expression of *serendipity alpha* and cellularisation of the *Drosophila* embryo as revealed by interspecific transformation. *Development* **119**(2):471-83.

IPCC WORKING GROUP I Third Assessment Report, Climate Change 2001: *The Scientific Basis*:1-20.

ICHIHARA, K., SHIMIZU, H., TAGUCHI, O., YAMAGUCHI, M., INOUE, Y.H. 2007. A *Drosophila* orthologue of *larp* protein family is required for multiple processes in male meiosis. *Cell Struct Funct* **32**(2):89-100.

JENSEN, L.T., COCKERELL, F.E., KRISTENSEN, T.N., RAKO, L., LOESCHCKE, V., MCKECHNIE, S.W., HOFFMANN, A.A. 2010. Adult heat tolerance variation in *Drosophila melanogaster* is not related with *Hsp70* expression. *J Exp Zool* **313A**:35-44.

JUKES, T.H., CANTOR, C.R. 1969. Evolution of protein molecules. Mammalian Protein Metabolism (Munro, H.M. [ed.]). *Academic Press, New York:* 21–120.

KARI, J.S., HUEY, R.B. 2000. Size and seasonal temperature in free-ranging *Drosophila subobscura*. *J Therm Biol* **25**:267-72.

KASTRITSIS, C.D., CRUMPACKER, D.W. 1966. Gene arrangements in the third chromosome of *Drosophila pseudoobscura*. I. Configurations with tester chromosomes. *J Hered* **57**:150–8.

KELLY, J.K. 1997. A test of neutrality based on interlocus associations. *Genetics* **146**:1197–206.

KENNEDY, P., NACHMAN, M.W. 1998. Deleterious mutations at the mitochondrial *ND3* gene in South American marsh rats (*Holochilus*). *Genetics* **150**:359–68.

KINGSOLVER, J.G., WATT, W.B. 1983. Thermoregulatory strategies in Colias butterflies: thermal stress and limits to adaptation in temporally varying environments. *Am Nat* **121**:32-5.

KIMURA, M. 1983. The Neutral Theory of Molecular Evolution. *Cambridge Univ Press, Cambridge.*

KIRKPATRICK, M., BARTON, N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**:419-34.

KLINGENBERG, C.P. 2003. Developmental instability as a research tool: using patterns of fluctuating asymmetry to infer the developmental origins of morphological integration. Developmental instability: causes and consequences (Polak, M. [ed.]). *New York, Oxford University Press*: 427–42.

KLINGENBERG, C.P. 2004. Integration, modules and development: molecules to morphology to evolution. Phenotypic integration: studying the ecology and evolution of complex phenotypes (Pigliucci, M., Preston, K. [ed.]). *New York, Oxford University Press*: 213–30.

Knibb, W.R., Oakeshott, J.G., Gibson, J.B. 1981. Chromosome inversion polymorphism in *Drosophila melanogaster*. I. Latitudinal clines and associations between inversions in Australasian populations. *Genetics* **98:**833–47.

Koske, Th., Maynard-Smith, J. 1954. Genetics and cytology of *Drosophila subobscura*. X. The fifth linkage group. *J Genet* **52:**521-41.

Krebs, R.A. 1999. A comparison of *Hsp70* expression and thermotolerance in adults and larvae of three *Drosophila* species. *Cell Stress Chaperones* **4**:243-9.

Krebs, R.A., Feder, M.E. 1997. Natural variation in the expression of the heat-shock protein *Hsp70* in a population of *Drosophila melanogaster* and its correlation with tolerance of ecologically relevant thermal stress. *Evolution* **51**:173-9.

Krimbas, C.B. 1964. The genetics of *D. subobscura* populations. II Inversion polymorphism in a population from Holland. *Z. Verebungsl*. **95**:125-8.

Krimbas, C.B. 1992. The inversion polymorphism of *Drosophila subobscura. Drosophila* inversion polymorphism (Krimbas, C.B., Powell, J.R. [ed.]). *Boca Raton (CRC Press)*:127-220.

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

Krimbas, C.B., Loukas, M. 1980. The inversion polymorphism of *Drosophila subobscura*. *Evol Biol* **12:**163-234.

Krimbas, C.B., Powell, J.F. 1992. *Drosophila* Inversion Polymorphism. *CRC Press, Boca Raton, Florida.*

Kunze-Mühl, E., Müller, E. 1958. Weitere Untersuchungen über die chromosomale Struktur und die natürlichen Strukturtypen von *Drosophila subobscura* Coll. *Chromosoma* **9**:559-70.

Kunze-Mühl, E., Müller, E., Sperlich, D. 1958. Qualitative, quantitative und jahreszeitliche Untersuchungen über den chromosomalen Polymorphismus natürlicher Populationen von *Drosophila subobscura* in der Umgebung von Wien. *Z. Vererbungsl*. **89**:636-46.

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

Laayouni, H., Hasson, E., Santos, M., Fontdevila, A. 2003. Inversion polymorphism and nucleotide variability in different regions of the secondchromosome in *Drosophila buzzatii*. *Mol Biol Evol* **20**: 931-44.

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:**83–6.

Lankinen, P., Pinsker, W. 1977. Allozyme constitution of standard strains of *Drosophila subobscura*. *Experientia* **33:**1301.

Larruga, J. M., Pinsker, W. 1984. The importance of geographic isolation, structural rearrangement and speciation for the genetic divergence of chromosome *O* in *Drosophila subobscura*, *Drosophila guanche*, and *Drosophila madeirensis*. *Z zool Syst Evolut-forsch* **22**:103-13.

Lemeunier, F., Aulard, S. 1992. Inversion polymorphism in *Drosophila melanogaster*. *Drosophila* Inversion Polymorphism (Krimbas, C. B., Powell, J. R. [ed.]). *Boca Raton (CRC Press)*:339-406.

Levitan, M. 1978. Studies of linkage in populations. IX. The effect of altitude on X-chromosomal arrangement combinations in *Drosophila robusta*. *Genetics* **82:**751–63.

Levitan, M. 1992. Chromosome variation in *Drosophila robusta*. *Drosophila* Inversion Polymorphism (Krimbas, C. B., Powell, J. R. [ed.]). *Boca Raton (CRC Press)*:221-338.

Levitan, M. 2003. Climatic factors and increased frequencies of southern chromosome forms in natural populations of *Drosophila robusta*. *Evol Ecol Res* **5**:597-604.

Lewontin, R.C. 1970. The Units of Selection. *Annu Rev Ecol Syst* **1**:1-18.

Loeschcke, V., Hoffmann, A.A. 2007. Consequences of heat hardening on a field fitness component in *Drosophila* depend on environmental temperature. *Am Nat* **169**:175-83.

Loukas, M., Krimbas, C.B., Vergini, Y. 1979. The genetics of *D. subobscura* populations. IX. Studies on linkage disequilibrium in four natural populations. *Genetics* **93**:497-523.

Lutterschmidt, W.I., Hutchinson, V.H. 1997. The critical thermal maximum: data to support the onset of spasms as the definitive end point. *Can J Zool* **75**:1553-60.

MacIntyre, R.J. 1966. The genetics of an acid phosphatase in *Drosophila melanogaster* and *Drosophila simulans*. *Genetics* **53**:461-74.

Mainx, F., Koske, T., Smital, E. 1953. Untersuchungen über die chromosomale Struktur europäischer Vertreter der *Drosophila obscura* Gruppe. *Z Indukt Abstammungs Vererbungsl* **85**:354-72.

Mather, K., Jinks, J.L. 1977. Introduction to Biometrical Genetics. *London: Chapman and Hall.*

Mather, K., Jinks, J.L. 1982. Biometrical Genetics: The Study of Continuous Variation. 3rd edition. *London: Chapman and Hall.*

MathWorks Inc: Matlab, V. 7.5.0.342 (R2007b). 2007. The Language of Technical Computing [http://www.mathworks.com].

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**:2462-3.

MARTINEZ-SEBASTIAN, M.J., LATORRE, A., DE FRUTOS, R. 1984. Chromosomal polymorphism of *Drosophila subobscura* populations sampled at different altitudes. *Genética Ibérica* **36**:1-11.

MAYNARD-SMITH, J. 1956. Acclimatization to high temperatures in inbred and outbred *Drosophila subobscura*. *J Genet* **54**:497-505.

MAYNARD-SMITH, J., HAIGH, J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* **23:**23–35.

MCCOLL, G., HOFFMANN, A.A., MCKECHNIE, S.W. 1996. Response of two heat shock genes to selection for knockdown heat resistance in Drosophila melanogaster. *Genetics* **143**:1615-27.

MCDONALD, J.H., KREITMAN, M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila. Nature* **351**(6328):652-4.

MCFARQUHAR, A.M., ROBERTSON, F.W. 1963. The lack of evidence for co-adaptation in crosses between geographical races of *Drosophila subobscura* Coll. *Genetic Research* **4**:104-31.

MCGAUGH, S.E., NOOR, M.A.F. 2012. Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Phil Trans R Soc B* **367**:422-9.

MCVEAN, G., AWADALLA, P., FEARNHEAD, P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**:1231-41.

MEIKLEJOHN, C.D., HARTL, D.L. 2002. A single mode of canalization. *Trends Ecol Evol* **17**:468-73.

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:**625-41.

MESTRES, F., ABAD, L., SABATER-MUÑOZ, B., LATORRE, A., SERRA, L. 2004. Colonization of America by *Drosophila subobscura*: association between *Odh* gene haplotypes, lethal genes and chromosomal arrangements. *Genes Genet Syst* **79**(4):233-44.

MESTRES, F., BALANYÀ, J., ARENAS, C., SOLÉ, E., SERRA, L. 2001. Colonization of America by *Drosophila subobscura*: heterotic effect of chromosomal arrangements revealed by the persistence of lethal genes. *PNAS* **98**(16):9167-70.

MESTRES, F., BALANYÀ, J., PASCUAL, M., ARENAS, C., SOLÉ, E., SERRA L. 2005. Lethal genes and the colonization of America by *D. subobscura*. *Cur Top Genet* **1**:31–57.

MESTRES, F., BALANYÀ, J., SEGARRA, C., PREVOSTI, A., SERRA, L. 1994. O chromosome inversion polymorphism in Northern and Atlantic Europe and its implications in the American colonization by *Drosophila subobscura*. *Z Zool Syst Evolut-forsch.* **32:**108-16.

MESTRES, F., PEGUEROLES, G., PREVOSTI, A., SERRA, L. 1990. Colonization of America by *Drosophila subobscura*: lethal genes and the problem of the O_5 inversion. *Evolution* **44**:1823-36.

MESTRES, F., SANZ, J., SERRA, L. 1998. Chromosomal structure and recombination between inversions in *Drosophila subobscura. Hereditas* **128**:105-13.

MESTRES, F., SERRA, L., AYALA, F.J. 1995. Colonization of the Americas by *Drosophila subobscura*: lethal gene allelism and association with chromosomal arrangements. *Genetics* **140**:1297-305.

MESTRES, F., SERRA, L. 2008. The *Va/Ba* balanced lethal strain: thirty years of research in *Drosophila subobscura* (1977–2007). *J Zool Syst Evol Res* **46**(2):190-91.

METTLER, L.E., VOELKER, R.A., MUKAI, T. 1977. Inversion clines in populations of *Drosophila melanogaster. Genetics* **87:**169–76.

MITCHELL, K.A., HOFFMANN, A.A. 2010. Thermal ramping rate influences evolutionary potential and species differences for upper thermal limits in *Drosophila*. *Funct Ecol* **24**:694-700.

MOLTÓ, M.D., PASCUAL, L., MARTÍNEZ-SEBASTIÁN, M.J., DE FRUTOS, R. 1992. Genetic analysis of the heat shock response in three *Drosophila* species of the *obscura* group. *Genome* **35**:870-80.

MONCLÚS, M. 1964. Distribución y ecología de Drosofílidos en España. I. Especies de *Drosophila* en la región catalana. *Genética Ibérica* **16**:143-65.

MORETEAU, B., MORIN, J.P., GIBERT, P., PETAVY, G., PLA, E., DAVID, J.R. 1997. Evolutionary changes of nonlinear reaction norms according to thermal adaptation: a comparison of two *Drosophila* species. *C R Acad Sci Paris* **320**:833-41.

MOUSSET, S., BRAZIER, L., CARIOU, M.L., CHARTOIS, F., DEPAULIS F., VEUILLE, M. 2003. Evidence of a high rate of selective sweeps in African *Drosophila melanogaster. Genetics* **163:**599–609.

MÜLLER, H.J. 1940. Bearings of the '*Drosophila'* work on systematics. In: Huxley, J. (ed). The new systematics. *Clarendon, Oxford*:185–268.

MUNTÉ, A., AGUADÉ, M., SEGARRA, C. 2000. Nucleotide variation at the yellow gene region is not reduced in *Drosophila subobscura*: a study in relation to chromosomal polymorphism. *Mol Biol Evol* 17(12):1942-55.

MUNTÉ, A., ROZAS, J., AGUADÉ, M., SEGARRA, C. 2005. Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. *Genetics* **69**: 1573-81.

NACHMAN, M.W., BOYER, S.N., AQUADRO, C.F. 1994. Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. *Proc. Natl. Acad. Sci. USA* **91**:6364–68.

NACHMAN, M.W., DROWN, W.M., STONEKING, M., AQUADRO, C. F. 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* **142**:953–63.

NAVARRO, A., BARBADILLA, A., RUIZ, A. 2000. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in *Drosophila*. *Genetics* **155**:685-98.

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

Navarro-Sabaté, A., Aguadé, M., Segarra, C. 1999a. The relationship between allozyme and chromosomal polymorphism inferred from nucleotide variation at the *Acph-1* gene region of *Drosophila subobscura. Genetics* **153**: 871–89.

Navarro-Sabaté, A., Aguadé, M., Segarra, C. 1999b. The *Acid phosphatase-1* gene region in the *Drosophila* species of the *subobscura* cluster. *Hereditas* **130**(1): 65-75.

Navarro-Sabaté, A., Aguadé, M., Segarra, C. 2003. Excess of nonsynonymous polymorphism at *Acph-1* in different gene arrangements of *Drosophila subobscura. Mol Biol Evol* **20**(11): 1833-43.

Nei, M. 1987. Molecular Evolutionary Genetics. *Columbia University Press, New York.*

Nei, M., Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**:418–26.

Nielsen, R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**(6):641-7.

Nijhout, H.F., Davidowitz, G. 2003. Developmental perspectives on phenotypic plasticity, canalization, and fluctuating asymmetry. Developmental Instability: Causes and Consequences. edited by (Polak, M. [ed.]). *MIT Press*: 3-13.

Nóbrega, C., Khadem, M., Aguadé, M., Segarra, C. 2008. Genetic exchange versus genetic differentiation in a medium-sized inversion of *Drosophila*: the A_2/A_{ST} arrangements of *Drosophila subobscura*. *Mol Biol Evol* **25**:1534-43.

Noor, M.A.F., Grams, K.L., Bertucci, L.A., Reiland, J. 2001. Chromosomal inversions and the reproductive isolation of species. *PNAS* **98**: 12084-8.

Nosil, P., Crespi, B.J., Sandoval, C.P., Kirkpatrick, M. 2006. Migration and the genetic covariance between habitat preference and performance. *Am Nat* **167**:E66-E78.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst* **23**:263–86.

Orengo, D.J., Prevosti, A. 1994. Preadult competition between *Drosophila subobscura* and *Drosophila pseudoobscura. J Zool Syst Evol Res* **32**(1):44-50.

Papaceit, M., Aguadé, M., Segarra, C. 2006. Chromosomal evolution of elements B and C in the *Sophophora* subgenus of *Drosophila*: evolutionary rate and polymorphism. *Evolution* **60**:768–81.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics* **37**:637-69.

PARSELL, D.A., TAULIEN, J., LINDQUIST, S. 1993. The role of heat-shock proteins in thermotolerance. *Phil Trans Roy Soc London Series B* **339**:279-86.

PASCUAL, M., AQUADRO, CH.F., SOTO, V., SERRA, L. 2001. Microsatellite variation in colonizing and Palearctic populations of *Drosophila subobscura. Mol Biol Evol* **18**:731-40.

PASCUAL, M., AYALA, F.J., PREVOSTI, A., SERRA, L. 1993. Colonization of North America by *Drosophila subobscura:* Ecological analysis of three communities of drosophilids in California. *J Zool Syst Evol Res* **31**(3): 216-26.

PASCUAL, M., BALANYÀ, J., LATORRE, A., SERRA, L. 1997. Analysis of the variability of *D. azteca* and *D. athabasca* populations revealed by randomly amplified polymorphic DNA. *J Zool Syst Evol Res* **35**:159–64.

PASCUAL, M., CHAPUIS, M.P., MESTRES, F., BALANYÀ, J., HUEY, R.B., GILCHRIST, G.W., SERRA, L., ESTOUP, A. 2007. Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Mol Ecol* **16:**3069-83.

PASCUAL, M., SCHUG, M.D., AQUADRO, C.F. 2000. High density of long dinucleotide microsatellites in *Drosophila subobscura*. *Mol Biol Evol* **17**:1259-67.

PEGUEROLES, C. 2010. Estudi de recombinació en *D. subobscura* a nivel genomic. *Tesi Doctoral. Universitat de Barcelona.*

PEGUEROLES, C., AQUADRO, C.F, MESTRES, F., PASCUAL, M. 2013. Gene flow and gene flux shape evolutionary patterns of variation in *D. subobscura. Heredity* **110**:520-9.

PEGUEROLES, C., ARAÚZ, P., PASCUAL, M., MESTRES, F. 2010a. A recombination survey using microsatellites: the O chromosome of *Drosophila subobscura*. *Genetica* **138**(7):795-804.

PEGUEROLES, C., ORDÓÑEZ, V., MESTRES, F., PASCUAL, M. 2010b. Recombination and selection in the maintenance of the adaptive value of inversions. *Journal of Evolutionary Biology* **23**(12): 2709–17.

PENTZOS-DAPONTE, A. 1964. Qualitative und quantitative Untersuchungen über den chromosomalen Polymorphismus natürlicher Populationen von *Drosophila subobscura* in der Umgebung von Thessaloniki–Griechenland. *Z Vererbungsl* **95**:129-44.

PENTZOS-DAPONTE, A., SPERLICH, D. 1965. Vitalitsprufungen an intra- und interlokalen Heterozygoten von *Drosophila subobscura*. *Z VererbLehre* **97**:46-54.

PINA-MARTINS, F., PAULO, O.S. 2008. CONCATENATOR: Sequence data matrices handling made easy. *Molecular Ecology Resources* **8**:1254-5.

PINSKER, W., BÖHM, I. 1989. Allozyme divergence between different gene arrangements in *Drosophila subobscura*. *J Evol Biol* **2**:353-66.

PINSKER, W., SPERLICH, D. 1981. Geographic pattern of allozyme and inversion polymorphism on

chromosome *O* of *Drosophila subobscura* and its evolutionary origin. *Genetica* **57**:51-64.

Pool, J.E., Aquadro, C.F. 2006. History and structure of Sub-Saharian populations of *Drosophila melanogaster. Genetics* **174**:915-29.

P , A., P , D., Anderson, W.W. 1995. Interchromosomal exchange of genetic information between gene arrangements on the third chromosome of *Drosophila pseudoobscura*. *Mol Biol Evol* **12**:938–43.

Powell, J.R. 1992. Inversion polymorphisms in *Drosophila pseudoobscura* and *Drosophila persimilirs*. *Drosophila* Inversion Polymorphism (Krimbas, C. B., Powell, J. R. [ed.]). *Boca Raton (CRC Press)*:73-126.

Powell, J.R. 1997. Progress and prospects in Evolutionary Biology: the *Drosophila* model*. Oxford University Press, New York*.

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

Prevosti, A. 1966. Chromosomal polymorphism in western Mediterranean populations of *Drosophila subobscura*. *Genet Res* **7**:149-58.

Prevosti, A. 1971. Chromosomal polymorphismin *Drosophila subobscura* Coll.: populations from the Canary Islands. *Genet Iber* **23**:69-84.

Prevosti, A. 1972. Chromosomal polymorphism in *Drosophila subobscura* populations from the Madeira Islands. *Genet Iber* **24**:11-21.

Prevosti, A. 1974. Chromosomal inversion polymorphism in the southwestern range of *Drosophila subobscura* distribution area. *Genetica* **45**:111-24.

Prevosti, A., de Frutos, R., Alonso, G., Latorre, A., Monclus, M., Martinez, M.-J. 1984. Genetic differentiation between natural populations of *D. subobscura* in the Western Mediterranean area with respect to chromosomal variation. *Genet Sel Evol* **16**:143-56.

Prevosti, A., Garcia, M.P., Serra, L., Aguade, M., Ribb, G., Sagarra, E. 1983. Association between isozyme alleles and chromosomal arrangements in European populations and Chilean colonizers of *Drosophila subobscura*: 171-191. In Ratazzi, M.C., Scandalios, J.G., Whitt, G.S. (eds.). Isozymes: Current Topics in Biological and Medical Research. Vol. **10**: Genetics and Evolution. *Alan R. Liss, New York.*

Prevosti, A., Ribó, G., Serra, L., Aguadé, M., Balanyà, J., Monclús, M., Mestres, F. 1988. Colonization of America by *Drosophila subobscura*: Experiment in natural populations that supports the adaptive role of chromosomal inversion polymorphism. *PNAS* **85**:5597-600.

Prevosti, A., Serra, L., Ribó, G., Aguadé, M., Sagarra, E., Monclús, M., García, M. P. 1985. The colonization of *Drosophila subobscura* in Chile. II. Clines in the chromosomal arrangements. *Evolution* **39**:838-44.

PREVOSTI, A., SERRA, L., SEGARRA, C., AGUADÉ, M., RIBÓ, G., MONCLÚS, M. 1990. Clines of chromosomal arrangements of *Drosophila subobscura* in South America evolve closer to Old World patterns. *Evolution* **44**:218-21.

QUINTANA, A., PREVOSTI A. 1990a. Genetic and environmental factors in the resistance of *Drosophila subobscura* adults to high temperature shock. I. Breeding temperature and crowding. *Theor Appl Genet* **79**:103-7.

QUINTANA, A., PREVOSTI, A. 1990b. Genetic and environmental factors in the resistance of *Drosophila subobscura* adults to high temperature shock. II. Modification of heat resistance by indirect selection. *Theor Appl Genet* **80**: 847-51.

QUINN, W.G., HARRIS, W.H., BENZER, S. 1974. Conditioned behavior in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **71**:708–12.

RAND, D.M., DORFSMAN, M., KANN, L.M. 1994. Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. *Genetics* **138**:741–56.

RAND, D.M., KANN, L.M. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol Biol Evol* **13**:735–48.

RAMOS-ONSINS, S.E., ROZAS, J. 2002. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* **19**:2092-100.

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. *Mol Phylogenet Evol* **9:**33–41.

RANGANATH, H.A., KRISHNAMURTHY, N.B. 1978. Chromosomal morphism in *Drosophila nasuta.* Coexistence of heteroselection and flexibility in the polymorphic system of South Indian populations. *Genetica* **48**:215-21.

REDDY, G.S., KRISHNAMURTHY, N.B. 1974. Altitudinal gradients in frequencies of three common inversions in *Drosophila ananasse*. *Drosophila Information Service* **51**:136-7.

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**:385-97.

REZENDE, E.L., BALANYÀ, J., RODRÍGUEZ-TRELLES, F., REGO, C., FRAGATA, I., MATOS, M., SERRA, L., SANTOS, M. 2010. Climate change and chromosomal inversions in *Drosophila subobscura. Climate Research* **43**:103-14.

REZENDE, E.L., TEJEDO, M., SANTOS, M. 2011. Estimating the adaptive potential of critical thermal limits: methodological problems and evolutionary implications. *Funct Ecol* **25**(1):111-21.

**RICHARDS, S., LIU, Y., BETTENCOURT, B.R., HRADECKY, P., LETOVSKY, S., NIELSEN, R., THORNTON, K., HUBISZ, M.J., CHEN, R., MEISEL, R.P., COURONNE, O., HUA, S., SMITH, M.A., ZHANG, P., LIU,

J., **Bussemaker, H.J., van Batenburg, M.F., Howells, S.L., Scherer, S.E., Sodergren, E., Matthews, B.B., Crosby, M.A., Schroeder, A.J., Ortiz-Barrientos, D., Rives, C.M., Metzker, M.L., Muzny, D.M., Scott, G., Steffen, D., Wheeler, D.A., Worley, K.C., Havlak, P., Durbin, K.J., Egan, A., Gill, R., Hume, J., Morgan, M.B., Miner, G., Hamilton, C., Huang, Y., Waldron, L., Verduzco, D., Clerc-Blankenburg, K.P., Dubchak, I., Noor, M.A., Anderson, W., White, K.P., Clark, A.G., Schaeffer, S.W., Gelbart, S.W., Weinstock, G.M., Gibbs, R.A.** 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene and cis-element evolution. *Genome Res* **15**:1–18.

Riesenberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution* **16**:351-8.

Roberts, P.A. 1962. Interchromosomal effects and the relation between crossing over and nondisjunction. *Genetics* **47**:1691-709.

Roberts, P.A. 1976. The genetics of chromosome aberration. *The Genetics and Biology of Drosophila*. *Academic Press, London*:67-184.

Rodríguez-Trelles, F. 1993. Analisis estacional de sistemas geneticos multiloci y estructura genetica en una poblacion natural de *Drosophila subobscura*. *Ph.D. thesis. Universidad de Santiago de Compostela.*

Rodríguez-Trelles, F. 2003. Seasonal cycles of allozyme-by-chromosomal inversion gametic disequilibrium in *Drosophila subobscura*. *Evolution* **57**:839-48.

Rodríguez-Trelles, F., Alvarez, G., Zapata, C. 1996. Time-series analysis of seasonal changes of the O inversion polymorphism of *Drosophila subobscura*. *Genetics* **142**:179-87.

Rodríguez-Trelles, F., Rodríguez, M.A. 1998. Rapid microevolution and loss of chromosomal diversity in *Drosophila* in response to climate warming. *Evol Ecol* **12**:829-38.

Rodríguez–Trelles, F., Rodríguez, M.A., Sheiner, S.M. 1998. Tracking the genetic effects of global warming: *Drosophila* and other model systems. *Conserv Ecol* **2**:2.

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

Rogers, A.R. 1995. Genetic evidence for a pleistocene population explosion. *Evolution* **49**:608–15.

Rogers, A.R., Harpending, H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* **9**:552–69.

Root, T.L., Price, J.T., Hall, K.R., Schneider, S.H. 2003. Fingerprints of global warming on wild animals and plants. *Nature* **421**:57-60.

Rozas, J., Aguadé, M. 1990. Evidence of extensive genetic exchange in the *rp49* region among polymorphic chromosome inversions in *Drosophila subobscura*. *Genetics* **126**:417-26.

Rozas, J., Aguadé, M. 1993. Transfer of genetic information in the *rp49* region of *Drosophila subobscura* between different chromosomal gene arrangements. *Proc Natl Acad Sci USA* **90:**8083–7.

Rozas, J., Aguadé, M. 1994. Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila*. *PNAS* **91**:11517-21.

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.

Rozas, J., Segarra, C., Zapata, C, Álvarez, G., Aguadé, M. 1995. Nucleotide polymorphism at the *rp49* region of *Drosophila subobscura*: Lack of geographic subdivision within chromosomal arrangements in Europe. *Journal of Evolutionary Biology* **8**:355-67.

Rozen, S., Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols* in the series *Methods in Molecular Biology*. *Humana Press, Totowa, NJ*: 365-86.

Russo, C.A.M., Takezaki, N., Nei, M. 1995. Molecular phylogeny and divergence time of Drosophilid species. *Mol Biol Evol* 12**(3)**: 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. 2007. Thermal evolution of gene expression profiles in *Drosophila subobscura*. *BMC Evol Biol* **7**:42.

Santos, M. 2009. Recombination load in a chromosomal inversion polymorphism of *Drosophila subobscura*. *Genetics* **181**:803-9.

Santos, M., Brites, D., Laayouni, L. 2006. Thermal evolution of pre-adult life history traits, geometric size and shape, and developmental stability in *Drosophila subobscura*. *J Evol Biol* **19**:2006-21.

Santos, M., Céspedes, W., Balanyà, J., Trotta, V., Calboli, F.C.F., Fontdevila, A., 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**: 258-73.

Santos, M., Iriarte, P.F., Cespedes, W. 2005. Genetics and geometry of canalization and developmental stability in *Drosophila subobscura*. *BMC Evol Biol* **5**:7.

Santos, J., Serra, L., Solé, E., Pascual, M. 2010. FISH mapping of microsatellite loci from *Drosophila subobscura* and its comparison to related species. *Chromosome res* **18**(2):213-26.

SAVKOVIC, V., STAMENKOVIC-RADAK, M., ANDJELKOVIC, M. 2004. Diurnal variability of gene arrangement frequencies in *Drosophila subobscura* populations from two habitats. *J Zool Syst Evol Res* **42**:208-14.

SAYEED, O., BENZER, S. 1996. Behavioral genetics of thermosensation and hygrosensation in *Drosophila*. *Proc Natl Acad Sci* **93**:6079–84.

SCHAEFFER, S.W., ANDERSON, W.W. 2005. Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. *Genetics* **171**:1729-39.

SCHAEFFER, S.W., BHUTKAR, A., MCALLISTER, B.F., MATSUDA, M., MATZKIN, L.M., O'GRADY, P.M., ROHDE, C., VALENTE, V.L.S., AGUADÉ, M., ANDERSON, W.W., EDWARDS, K., GARCIA, A.C.L., GOODMAN, J., HARTIGAN, J., KATAOKA, E., LAPOINT, R.T., LOZOVSKY, E.R., MACHADO, C.A, NOOR, M.A.F., PAPACEIT, M., REED, L.K., RICHARDS, S., RIEGER, T.T., RUSSO, S.M., SATO, H., SEGARRA, C., SMITH, D.R., SMITH, T.F., STRELETS, V., TOBARI, Y.N., TOMIMURA, Y., WASSERMAN, M., WATTS, TH., WILSON, R., YOSHIDA, K., MARKOW, TH.A., GELBART, W.M., KAUFMAN, TH.C. 2008. Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* **179**:1601–55.

SCHAEFFER, S.W., GOETTING-MINESKY, M.P., KOVACEVIC, M., PEOPLES, J.R., GRAYBILL, J.L., MILLER, J.M., KIM, K., NELSON, J.G., ANDERSON, W.W. 2003. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. *PNAS* **100**: 8319–24.

SCHÄFER, M. A., MAZZI, D., KLAPPERT, K., KAURANEN, H., VIEIRA, J., HOIKKALA, A., RITCHIE, M. G., SCHLÖTTERER, C. 2010. A microsatellite linkage map for *Drosophila montana* shows large variation in recombination rates, and a courtship song trait maps to an area of low recombination. *J Evol Biol* **23**:518-27.

SCHULTZ, J., REDFIELD, H. 1951. Interchromosomal effects on crossing-over in *Drosophila*. *Cold Spring Harbor Symposia on Quantitative Biology* **16**:175-97.

SEGARRA, C., AGUADÉ, M. 1992. Molecular organization of the *X* chromosome in different species of the *obscura* group of *Drosophila*. *Genetics* **130**:513–21.

SEGARRA, C., RIBÓ, G., AGUADÉ M. 1996. Differentiation of Muller's Chromosomal Elements D and E in the *Obscura* Group of *Drosophila*. *Genetics* **144**(1):139-46.

SERRA, L. 2009. La microevolució en acció: marcacors genètics del canvi climatic. *Cent cinquanta anys després de L'origen de les espècies, de Darwin* (Navarro, A., Segarra, C. [ed.]). *Treballs de la SCB* **60***:* 181-94.

SERRA, L., PEGUEROLES, G., MESTRES, F. 1987. Capacity of dispersal of a colonizing species: *Drosophila subobscura*. *Genetica* **73**:223-35.

SHARP, P. M., LI,W-H. 1989. On the rate of DNA sequence evolution in *Drosophila***.** *J Mol Evol* **28**:398-402.

SHORROCKS, B. 1982. The breeding sites of temperate woodland *Drosophila*. In: Ashburner, M., Carson, H. L. and Thompson, J. N., Jr (eds) *The Genetics and Biology of Drosophila* **3b**: 385–428. Academic Press, London.

SIMOES, P., CALABRIA, G., PICAO-OSÓRIO, J., BALANYÀ, J., PASCUAL, M. 2012. The genetic content of chromosomal inversions across a wide latitudinal gradient. *Plos one* **7**(12):e51625.

SINCLAIR, B.J., VERNON, P., KLOK, C.J., CHOWN, S.L. 2003. Insects at low temperatures: an ecological perspective. *Trends Ecol Evol* **18**: 257-62.

SINNOT, E.W., DUNN, L.C., DOBZHANSKY, T. 1958. Principles of genetics. *New York: McGraw Hill*: 278.

SMITH, N.G., EYRE-WALKER, A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022–4.

SLATKIN, M., HUDSON, R.R. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**:555–62.

SOKAL, R.R., ROHLF, F.J. 1995. Biometry. 3rd edition. *New York: Freeman*.

SOLÉ, E. 2002. Anàlisi dels canvis cromosòmics a llarg termini en poblacions naturals de *Drosophila subobscura* i la seva relació amb el possible canvi climàtic global. *Tesi Doctoral. Departament de Genètica, Universitat de Barcelona.*

SOLÉ, E., BALANYÀ, J., SPERLICH, D., SERRA, L. 2002. Long-term changes of the chromosomal inversion polymorphism of *Drosophila subobscura*. I. Mediterranean populations from South-western Europe. *Evolution* **56**: 830-5.

SOLER, M. 2002. Selección Natural y Adaptación. *Evolución. La base de la Biología. Proyecto Sur de Ediciones* (Soler, J.J. [ed.]). Capitulo **7**:127-57.

SØRENSEN, J.G., KRISTENSEN, T.N., LOESCHCKE, V. 2003. The evolutionary and ecological role of heat shock proteins. *Ecol Lett* **6**:1025-37.

SØRENSEN, J.G., 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**:1301-7.

SPERLICH, D., FEUERBACH-MRAVLAG, H. 1974. Epistatic gene interaction, crossing-over and linked and unlinked inversions of *Drosophila subobscura*. *Evolution* **28:**67-75.

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**:835-48.

SPERLICH, D., PFRIEM, P. 1986. Chromosomal polymorphism in natural and experimental populations. The genetics and biology of *Drosophila* (Ashburner, M., Carson, H.L., Thompson Jr.J. [ed.]). *Academic Press*: 257-309.

SPSS INC: SPSS for Windows 2006 [http://www.spss.com].

STALKER, H.D. 1976. Chromosome studies in wild populations of *Drosophila melanogaster. Genetics* **82**:323–47.

STATSOFT INC. 2009. STATISTICA (data analysis software system), version 9.0. www.statsoft.com.

STEINBERG, A.G. 1936. The effect of autosomal inversions on crossing over in the X-chromosome of *D. melanogaster. Genetics* **21**: 615-24.

STEVENSON, R.D. 1985. The relative importance of behavioral and physiological adjustments controlling body temperature in terrestrial ectotherms. *Am Nat* **126**: 362-86.

STRICKBERGER, M.W. 1976. Genética, 2ª edición. *Omega, Barcelona.*

STUMM-ZOLLINGER, E. 1953. Vergleichende Untersuchungen über die Inversionshäufigkeit bei *Drosophila subobscura* in Populationen der Schweiz und Südwesteuropas. *Z. indukt. Abstamm.-u. Vererb.-Lehre* **85**: 382-407.

STUMP, A.D., POMBI, M., GOEDDEL, L., RIBEIRO, J.M.C., WILDER, J.A., RORRE, A.D., BESANSKY, N.J. 2007. Genetic exchange in *2La* inversion heterokaryotypes of *Anopheles gambiae*. *Insect Mol Biol* **16**:703-9.

STURTEVANT, A.H. 1921. A case of rearrangement of genes in *Drosophila*. *PNAS* **7**:235-7.

STURTEVANT, A.H. 1926. A crossover reducer in *Drosophila melanogaster* due to inversion of a section of the third chromosome. *Biologisches Zentralblatt* **26**: 697-702.

STURTEVANT, A.H., BEADLE, G.W. 1936. The relations of inversions in the X chromosomes of *Drosophila melanogaster* to crossing-over and disjunction. *Genetics* **28**: 554-604.

STURTEVANT, A.H., PLUNKETT, C.R. 1926. Sequence of corresponding third chromosome genes in *Drosophila melanogaster* and *D. simulans*. *Biology Bulletin* **50**:56-60.

TADIN-STRAPPS, M., WARBURTON, D., BAUMEISTER, F.A.M., FISCHER, S.G., YONAN, J., GILLIAM, T.C., CHRISTIANO, A.M. 2004. Cloning of the breakpoints of a *de novo* inversion of chromosome 8, inv (8)(p11.2q23.1) in a patient with Ambras syndrome. *Cytogenet Genome Res* **107**:68–76.

TAJIMA, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–95.

TAKAHATA, N. 1993. Relaxed natural selection in human populations during the Pleistocene. *Jpn J Genet* **68**:539–47.

TAMURA, K., SUBRAMANIAN, S., KUMAR, S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* **21**:36-44.

Tan, C.C. 1935. Salivary gland chromosomes in the two races of *Drosophila pseudoobscura*. *Genetics* **20**:392–402.

Tan, C.C. 1937. The cytological maps of the autosomes in *Drosophila pseudoobscura*. *Zeitschrift fur Zellforschung u. mikr. Anatomie* **26**:439–62.

The UniProt Consortium. 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 40: D71-D75.

Thompson, J.D., Higgins, D.G., Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**:4673–80.

Torgerson, D.G., Boyko, A.R., Hernandez, R.D, Indap, A., Hu. X., White, Th.J., Shinsky, J.J., Cargill, M., Adams, M.D., Bustamante, C.D., Clark, A.G. 2009. Evolutionary Processes Acting on Candidate *cis*-Regulatory Regions in Humans Inferred from Patterns of Polymorphism and Divergence. *PLoS Genet* **5**(8):e1000592.

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**:691-93.

Van 't Land, J., Van Putten, P., Zwaan, B., Kamping, A., Van Delden, W. 1999. Latitudinal variation in wild populations of *Drosophila melanogaster*: heritabilities and reaction norms. *J Evol Biol* **12**:222–32.

Waddington, C.H. 1942. Canalization of development and the inheritance of acquired characters. *Nature* **150** (3811):563–5.

Wallace, B. 1953. On coadaptation in *Drosophila. Am Nat* **87:**343-58.

Wan, L., Dockendorff, T.C., Jongens, T.A., Dreyfuss, G. 2000. Characterization of dFMR1, a *Drosophila melanogaster* Homolog of the Fragile X Mental Retardation Protein. *Mol Cell Biol* **20**(22):8536-47.

Wasserman, M. 1968. Recombination-Induced Chromosomal Heterosis. *Genetics* **58**:125-39.

Weinreich, D.M., Rand, D.M. 2000. Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. *Genetics* **156**:385–99.

Weir, B.S. 1996. Genetic Data Analysis II. *Sinauer Associates, Sunderland, MA.*

Wells, R.S. 1996. Nucleotide variation at the *Gpdh* locus in the genus *Drosophila*. *Genetics* **143**: 375–84.

Wesley, C.S., Eanes, W.F. 1994. Isolation and analysis of the breakpoint sequences of chromosome inversion *In(3L) Payne* in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **91**:3132-6.

White, M.J.D. 1968. Models of Speciation. *Science* **159**:1065-70.

WHITE, M.J.D., MORLEY, F.H.W. 1955. Effects of pericentric rearrangements on recombination in grasshopper chromosomes. *Genetics* **40**:604-19.

WILLIAMS, J.G.K., KUBELIK, A.R., LIVAK, K.J., RAFALSKI, J.A., TINGEY, S.V. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* **18:**6531–5.

WILLMER, P.G., UNWIN S. 1981. Field analyses of insect heat budget: reflectance, size and heating rates. *Oecologia* **50**:250-5.

WISE, C.A., SRAML, M., EASTEAL, S. 1998. Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. *Genetics* **148**:409–21.

WRIGHT, S., DOBZHANSKY, TH., HOVANITZ, W. 1942. Genetics of natural populations. VII. The allelism of lethals in the third chromosome of *D. pseudoobscura*. *Genetics* **27**:363–94.

YAMAMOTO, A.H. 1994. Temperature preference of *Drosophila immigrans* and *D. virilis*: Intra- and inter-population genetic variation. *Jpn J Genet* **69**:67-76.

ZAPATA, C., ALVAREZ, G. 1992. The detection of gametic disequilibrium between allozyme loci in natural populations of *Drosophila*. *Evolution* **46**:1900-17.

ZIVANOVIC, G. 2007. Seasonal changes in chromosomal inversion polymorphism in a *Drosophila subobscura* natural population from a south-eastern European continental refugium of the last glaciation period. *Russ J Genet* **43**:1344–9.

ZIVANOVIC, G., ARENAS, C., MESTRES, F. 2007. The genetic structure of Balkan populations of *Drosophila subobscura*. *Hereditas* **144**:120-8.

Supplementary Table S1: The genes mapped on the chromosome *O* of *D. subobscura*, their cytological localization and genomic position in *D. pseudoobscura* and *D. melanogaster*.

Gene region in *D. melanogaster*	Symbol	Name of the gene	Locatization in *D.subobscura*	Primers used for in situ hybridization	Size of the probe (bp)	Reference	Gene ID and genomic position in *D. pseudoobscura*	Genomic localization in *D. melanogaster*
CG4316	*Sb, Sbd*	*Stubble*	*O(77B)*	5'-GTGGCTATCCAGTTCTCGTT-3' 5'-CCAGAACTTCAAGATCAGCC-3'	2288	*Our study*	*GA18102: 2: 26,116,311-26,121,335; (46B)*	*89B4-89B6*
CG6072	*sra*	*sarah*	*O(77B)*	5'-GCCAGAGCTGCCAGTAGATC-3' 5'-CGGCCAAGTCCAATAACAAT-3'	726	*Our study*	*GA19337: 26,059,846-26,060,739; (46B)*	*89B7*
CG10889			*O(77B)*	5'-TTGTGAAGTGCAGGAACAGC-3' 5'-GATGGAAGTTGGAGCTCTGC-3'	1776	*Our study*	*GA10620: 2: 27,825,978-27,827,753; (45A)*	*92C1-92C2*
CG9434	*Fst*	*Frost*	*O(77C)*			*Arboleda 2008*		*85E2*
CG9768	*Hkb*	*huckebein*	*O(78A)*	5'-CAGGACGGTGAAGAACGAG-3' 5'-GGCATATGCGTCTTCATGTG-3'	954	*Our study*	*GA22020: 2: 24,774,746-24,775,699; (47C)*	*82A4*
CG31045	*Mhc l*	*Myosin heavy chain-like*	*O(78A)*	5'-ACACAGCATGTTCCAACCAA-3' 5'-GTGCGAGACTCCTCCAGTTC-3	2091	*Our study*	*GA15963: 2:26,051,097-26,075,113; (46B)*	*89B2-89B3*
CG3772	*cry*	*cryptochrome*	*O(79A)*	5'- AGTTTTCCATCGCCCATAAA-3' 5'- AGCGCAGTTCCTCGCTATC-3'	2003	*Our study*	*GA17677: 2: 29,503,099-29,505,499; (43E)*	*91F11*
CG42599	*Pif1A*	*PFTAIRE-interacting factor 1A*	*O(79D)*	5'-AATGTATCACAAGGAGAACG-3' 5'-CTCCTGGTAGTACTGCAGAT-3'	2241	*Our study*	*GA11190: 2: 27,295,914-27,297,834; (45C)*	*85B1*
CG11988	*neur*	*neuralized*	*O(79D)*	5'-CATCATCCGCATCAGCAGTC-3' 5'-GAGTTGTGTTGCAGCATCTT-3	788	*Our study*	*GA11314: 2: 27,342,727-27,354,633; (45C)*	*85C2-85C3*
CG9749	*Abi*	*Abelson interacting protein*	*O(80A)*	5'-ACCTTCTCCTTGTGAATATG-3' 5'-GTGAGTAGTAATTAGGTTCG-3'	534	*Our study*	*GA22007: 2: 13,447,192-13,449,265; (54E)*	*88A9*
CG1028	*Antp*	*Antennapedia*	*O(80C)*			*Segarra et al. 1996*	*GA10215: 2: 19,646,190-19,660,064; 50A*	*84A6-84B2*
CG9379	*by*	*blistery*	*O(82B)*	5'-CTGGTTATCCTTGGAGCTGT-3' 5'-AAGAATGCCTATGGCCTGGT-3	590	*Laayouni et al. 2007; our study*	*GA20869: 2: 22,243,802-22,249,233; (48D)*	*85D22*
CG31530			*O(82C)*	5'-TCGGAGGATTCTCATTTTGG-3' 5'-AACTCCGTGCAGGGTATGAC-3'	2089	*Our study*	*GA16305: 2: 22,219,020-22,234,100; (48D)*	*82D4-82D5*
CG31534			*O(82D-83A)*			*Our study*	*GA16309: 2: 24,119,334-24,134,403; (47E)*	*82D2*
CG18599			*O(83C)*	5'-CCTGGCATTGCATAACACAC-3' 5'-TTGTCCTCCTCCTGCAGACT-3'	1231	*Our study*	*GA15009: 2: 19,374,884-19,376,291; (50B)*	*90F11-91A1*
CG6598	*Odh, Fdh*	*Formaldehyde dehydrogenase*	*O(84B)*			*Mestres et al. 2004*	*GA19711: 2: 28,114,639-28,116,062; (44C)*	*86C7*
CG17184			*O(84C)*	5'-GAAAGATACGCCCTCCATGA-3' 5'-TTGCATTTTGACAGCCACAT-3'	1012	*Our study*	*GA14372: 2: 28,616,445-28,617,789; (44B)*	*86D8*
CG12537	*rdx*	*roadkill*	*O(85E)*	5'-GAACTCCCTGCTGTTGAAGC-3' 5'-TCCTCGCACATCACTTTCAG-3'	2185	*Our study*	*GA22127: 2: 13,547,216-13,617,295; (54E)*	*88A3-88A5*
CG2244	*MTA1-like*	*MTA1-like*	*O(86A)*	5'-GATAAAATCCGAAGGCACGA-3' 5'-TGATCTGGTAGCGCTGTTTG-3'	2155	*Our study*	*CG12170: 2:11,200,625-11,223,015; (55C)*	*83B7*

For *D. pseudoobscura* approximate gene locations (in parentheses) were determined using their genomic positions from Flybase server and physical locations of several previously localized genes as landmarks (without parentheses).

Supplementary Table S1 (continued).

Gene region in *D. melanogaster*	Symbol	Name of the gene	Localization in *D.subobscura*	Primers used for in situ hybridization	Size of the probe (bp)	Reference	Gene region and genomic position in *D. pseudoobscura*	Hybridization signal in *D. melanogaster*
CG7642	*Xdh, ry*	*rosy, Xanthine dehydrogenase*	*O(86C)*			*Segarra et al. 1996*	*GA20500: 2: 11,261,531-11,266,780; 55C*	*87D9*
CG18290	*Act87E*	*Actin 87E*	*O(86D)*			*Laayouni et al. 2007*	*GA14877: 2: 16,707,867-16,709,548; (52E)*	*87E11*
CG16901	*sqd*	*squid*	*O(86E)*	5'-AACTAACCTTGTTCCTCTCC-3' 5'-TTACACACGCTTCGTCAGTT-3'	2892	*Our study*	*GA14206: 2: 16,910,400-16,920,357; (52D)*	*87F5-87F6*
CG9764	*yrt*	*yurt*	*O(86E)*	5'-CTGGACATCATCGAGAAGGA-3' 5'-ACATTGGCCAGCTTCACTTG-3'	2143	*Laayouni et al. 2007; our study*	*GA22017: 2: 16,711,119-16,715,402; (52E)*	*87E11*
CG5670	*Atpα*	*Na pump α subunit*	*O(87C)*	5'-TCATAAGATCTCTCCTGAGG-3' 5'-GCAATATCCTCAACGGTCTC-3'	1959	*Laayouni et al. 2007; our study*	*GA19046: 2: 12,645,276-12,672,065; (55A)*	*93A4-93B2*
CG2867	*Prat*	*Phosphoribosylamidotransferase*	*O(88A)*	5'-ACCGGAGCAGCAACAACA-3' 5'-TGATCATGTCTTTTATTGGCATTT-3'	2012	*Our study*	*GA15494: 2: 27,159,956-27,161,485; (45D)*	*84E5*
CG6129	*Rootletin*	*Rootletin*	*O(88E)*	5'-CAATGCACTCGAAAAGCGTA-3' 5'-GAATGCTGCAGTCCTTCTCC-3'	2179	*Our study*	*GA19376: 2: 14,487,340-14,495,059; (54C)*	*95E1*
CG5436	*Hsp68*	*Heat shock protein 68*	*O(88E)-O(89B)*			*Moltó et al. 1992; Cuenca et al. 1998*	*GA18881: 2: 14,410,445-14,412,702; (54C)*	*95D11*
7 copies, for example CG5834	*i.e. Hsp70Bbb*	*Heat shock protein 70*	*O(89A-94A)*			*Molto et al. 1992; Cuenca et al. 1998*	*GA19632: 2: 5,523,773-5,525,918; (58E)*	*7 copies, i.e. 87B14*
CG11516	*Ptp99A*	*Protein tyrosine phosphatase 99A*	*O(89C)*	5'-ACCACCGACAAGGAGAACAC-3' 5'-GCAATCAGCTCGCTATCCTC-3'	2189	*Our study*	*GA30065: 2: 2,069,276-2,173,200; (61B)*	*99A7-99B1*
CG7899	*Acph1*	*Acid phosphatase 1*	*O(91C)*			*Aguadé 1988b*	*GA20670: 2:921,927-923,587; 62B*	*99D1*
CG7939	*rp49, RpL32*	*Ribosomal protein L32*	*O(91C)*			*Rozas and Aguade 1993*	*GA20704: 2: 978,884-979,746; 62B*	*99D3*
CG17957	*Sry-α*	*Serendipity α*	*O(91C)*			*Ibnsouda et al. 1993*	*GA14740: 2: 975,270-976,970; (62B)*	*99D3*
CG14247	*(S25)*		*O(92B)*			*Múnte et al. 2005*	*GA12857: 2: 4,917,866-4,918,423; (59A)*	*97D1*
CG13633	*Ast*	*Allatostatin*	*O(93B)*			*Múnte et al. 2005*	*GA12425: 2: 1,646,898-1,647,423; (62A)*	*96A20*
CG14066	*larp*	*La related protein*	*O(93C)*			*Our study*	*GA12736: 2: 7,926,064-7,928,574; (58A)*	*98C3-98C4*
CG5333	*trus*	*toys are us*	*O(94A)*			*Múnte et al. 2005*	*GA18809: 2: 5,510,314-5,512,165; (58E)*	*87B11*
CG5961	*(P154)*		*O(94A)*			*Múnte et al. 2005*	*GA19260: 2: 5,512,241-5,513,634; (58E)*	*87B11*

For *D. pseudoobscura* approximate gene locations (in parentheses) were determined using their genomic positions from Flybase server and physical locations of several previously localized genes as landmarks (without parentheses).

Supplementary Table S1 (continued).

Gene region in *D. melanogaster*	Symbol	Name of the gene	Localization in *D.subobscura*	Primers used for in situ hybridization	Size of the probe (bp)	Reference	Gene region and genomic position in *D. pseudoobscura*	Hybridization signal in *D. melanogaster*
CG10325	*abd-A*	*abdominal A*	*O(94E)*			*Múnte et al. 2005*		*89E2*
CG31012	*Cindr*	*CIN85 and CD2AP orthologue*	*O(95A)*			*Múnte et al. 2005*	*GA26879: 2: 529,310-541,652; (62C)*	*100A6*
CG1455	*CanA1*	*Calcineurin A1*	*O(95B)*			*Múnte et al. 2005*	*GA26888: 2: 275,898-283,329; (62C)*	*100B1*
CG5595	*Sce*	*Sex combs extra*	*O(95D)*	5'-AAGCCGCAGGAGATAATCAC-3' 5'-ACCTGATGCAACGTCTGGTT-3'	1192	*Our study*	*GA18995: 2: 3,865,180-3,866,647; (59D)*	*98B1*
CG5003			*O(96A)*	5'-GTTGGGCATATTGGAGGAGA-3' 5'-ATCGTCGAAATCGAATGGAG-3'	2086	*Our study*	*GA18586: 2: 4,102,498-4,105,371; (59C)*	*98B5-98B6*
CG7951	*sima*	*similar*	*O(96C)*	5'-GTTCTGTGTTCCGCCGTATT-3' 5'-AGTGCCAATGGGTTAGAACG-3'	2040	*Our study*	*GA20714: 2: 993,088-1,056,824; (62B)*	*99D3-99D7*
CG17998	*Gprk2*	*G protein-coupled receptor kinase 2*	*O(97A)*	5'-CCTGACAGCTTTTGCATTGA-3' 5'-AAACAATGAACCCGAACAGC-3'	2119	*Our study*	*GA14759: 2: 4,467,218-4,505,425; (59C)*	*100C3-100C4*
CG5394	*Aats-glupro*	*Glutamyl-prolyl-tRNA synthetase*	*O(97B)*	5'-GTTAGCAGCTGGCCGTACTC-3' 5'-TGTTGGTCATGTTCAAGCGA-3'	1692	*Our study*	*GA18849: 2: 14,290,349-14,296,706; (54C)*	*95D1*
CG5650	*Pp1-87B*	*Protein phosphatase 1 at 87B*	*O(97C)*	5'-TTCATGCTAGTTAGCTGTGA-3' 5'-GCTTCAAGATCTGGAACGAG-3'	1018	*Our study*	*GA19032: 2: 5,476,843-5,477,751; (58E)*	*87B9-87B10*
CG6203	*Fmr1*	*fragile X mental retardation 1*	*O(97E)*	5'-ACAGCCAAGTCGTTCTACCA-3' 5'-CCATTCACCAGACCTTCCTT-3'	2280	*Our study*	*GA19437: 2: 5,011,163-5,015,966; (59A)*	*85F10-85F12*
CG10091	*GstD9*	*Glutathione S transferase D9*	*O(98B)*	5'-AATCCTCAGCACACGATTCC-3' 5'-TTCTTCGCATTCTCGTACCA-3'	430	*Our study*	*GA10065: 2: 5,423,583-5,424,233; (58E)*	*87B8*
CG11422	*Obp83b*	*Odorant-binding protein 83b*	*O(98D)*			*Sanchez-Gracia and Rozas 2011*	*GA10996: 2: 15,991,865-15,992,405; (53B)*	*83D1*
CG11421	*Obp83a*	*Odorant-binding protein 83a*	*O(98D)*			*Sanchez-Gracia and Rozas 2011*	*GA10995: 2: 15,994,913-15,995,635; (53B)*	*83D1*
CG1213			*O(98D)*			*Our study*	*GA11424: 2: 16,113,575-16,115,183; (53A)*	*83C5*
CG32934	*CecII*	*Cecropin II*	*O(99C)*			*Segarra et al. 1996*	*GA17203: 2: 1,471,977-1,472,227; 62C*	*99E2*

For *D. pseudoobscura* approximate gene locations (in parentheses) were determined using their genomic positions from Flybase server and physical locations of several previously localized genes as landmarks (without parentheses).

Supplementary Table S2: Haplotypes of the eight gene regions in *D. subobscura*.

Following tables show the nucleotide polymorphic sites for each gene region in the *D. subobscura* lines. The name of the gene region and the number of position in the alignment are indicated at the top of the table.

Different lines are grouped according to their gene arrangement and population. In the superior part there are lines with the O_{ST} arrangement, then the O_{3+4} and in the inferior part the O_{3+4+8} arrangement. Inside each arrangement first lines correspond to the population of Barcelona (their names start with "Ba") and the following ones to the population of Málaga (the names strat with "Ma").

Dots indicate nucleotides identical to the reference sequence while dashes indicate gaps. Noncoding regions appear shaded. The nonsynonymous changes in exonic regions marked with asterisk (*). The gene conversion tracts are indicated with boxes.

Acph-1		6	10	21	25	26	30	46	51	62	68	70	101	111	135	136	139	179*	201	249	264	268*	285	293*	348	349	350	355	383	388
Bam73	st	G	G	A	G	T	T	C	A	G	C	T	C	T	C	A	T	G	C	C	C	G	C	A	T	A	T	G	G	T
Baf58	st	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Baf75	st	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Baf145	st	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Baf147	st	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Baf169	st	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	st	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	G
Baf200	st	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G
Baf203	st	A	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Mam16	st	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G
Maf36	st	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf37	st	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf81	st	.	.	-	.	.	.	.	.	.	.	A	T	C	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G
Maf85	st	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf85.2	st	.	A	-	.	.	.	.	.	.	.	A	T	C	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf116	st	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Bam16	3+4	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	G
Bam56	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G
Baf23	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G
Baf32	3+4	.	.	-	T	A	.	.	.	.	.	A	T	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	G
Baf46	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	A	.	.	.	.	.	.	.	.	A	G
Baf69	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Baf81	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G
Baf148	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G
Baf167	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Baf185	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Mam16	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Mam27	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G
Maf1	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf5	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf11	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf20	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf22	3+4	.	.	-	.	.	A	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G
Maf34	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf78	3+4	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Bam6	3+4+8	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G
Bam13	3+4+8	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	G
Bam32	3+4+8	.	.	-	T	A	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Baf14	3+4+8	A	.	C	.	.	.	.	G	.	T	.	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G
Baf20	3+4+8	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Baf21	3+4+8	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G
Baf53	3+4+8	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G
Baf166	3+4+8	.	.	C	.	.	.	.	G	A	.	.	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G
Baf159	3+4+8	.	.	-	.	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G
Baf196	3+4+8	.	.	C	.	.	.	.	G	A	.	.	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G
Mam28	3+4+8	.	.	-	T	A	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf17	3+4+8	.	.	-	T	A	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf13	3+4+8	.	.	.	.	.	.	.	.	.	.	.	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Maf25	3+4+8	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G
Maf27	3+4+8	.	.	C	.	.	.	.	G	.	.	.	T	C	A	.	.	.	.	.	.	.	.	G	.	.	.	T	.	G

Supplementary Tables S2

Acph-1		6	10	21	25	26	30	46	51	62	68	70	101	111	135	136	139	179*	201	249	264	268*	285	293*	348	349	350	355	383	388	412	429	436	479
Bam73	O_{ST}	G	G	A	G	T	T	C	A	G	C	T	C	T	C	A	T	G	C	C	C	G	C	A	T	A	T	G	G	T	C	T	A	A
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf147	O_{ST}	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	G	.	.	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	.	.	.
Baf203	O_{ST}	A	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Mam16	O_{ST}	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	T
Maf36	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Maf37	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Maf81	O_{ST}	.	.	-	.	.	.	.	.	.	.	A	T	C	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Maf85.2	O_{ST}	.	A	-	.	.	.	.	.	.	.	A	T	C	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Bam16	O_{3+4}	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	G	.	.	.	.
Bam56	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.
Baf23	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	G	.	.	T
Baf32	O_{3+4}	.	.	-	T	A	.	.	.	.	.	A	T	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	G	.	.	.	T
Baf46	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	A	.	.	.	.	.	.	.	.	A	G	.	.	.	T
Baf69	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf81	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.
Baf148	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.
Baf167	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf185	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Mam16	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Mam27	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	.
Maf1	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Maf5	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Maf11	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Maf20	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Maf22	O_{3+4}	.	.	-	.	.	A	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.	.
Maf34	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Maf78	O_{3+4}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Bam6	O_{3+4+8}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	T
Bam13	O_{3+4+8}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	G	G	.	.	T
Bam32	O_{3+4+8}	.	.	-	T	A	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	A	.	T
Baf14	O_{3+4+8}	A	.	C	.	.	.	.	G	.	T	.	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	T
Baf20	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	.
Baf21	O_{3+4+8}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	G	.	.	T
Baf53	O_{3+4+8}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.	.
Baf166	O_{3+4+8}	.	.	C	.	.	.	.	G	A	.	.	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	T
Baf159	O_{3+4+8}	.	.	-	.	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.
Baf196	O_{3+4+8}	.	.	C	.	.	.	.	G	A	.	.	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	T
Mam28	O_{3+4+8}	.	.	-	T	A	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T
Maf17	O_{3+4+8}	.	.	-	T	A	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	T
Maf13	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	T	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	.	T	T
Maf25	O_{3+4+8}	.	.	-	.	.	.	.	.	.	.	A	T	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	G	C	.	T
Maf27	O_{3+4+8}	.	.	C	.	.	.	.	G	.	.	.	T	C	A	.	.	.	.	.	.	.	.	G	.	.	.	T	.	G	.	.	.	T

Supplementary Tables S2

Acph-1		485	520*	549	554	566	573	574	577	580	581	583	585	588	591	593	596	601	602	612	620	626	641	643*	653	668	680	686	698	716	725	728	740	743
Bam73	O_{ST}	T	G	C	C	C	A	A	C	A	A	A	C	G	C	A	A	G	T	C	T	G	C	A	C	C	G	G	T	T	G	C	G	C
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	C	.	T	T
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	-	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	G	T	T
Baf147	O_{ST}	.	.	.	.	.	G	C	.	T	C	.	.	.	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	G	T	T
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	G	T	T
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	C	G	T	T
Baf203	O_{ST}	.	.	.	.	.	G	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	T	T
Mam16	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	T	.	.	.	.	C	.	T	T
Maf36	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	T
Maf37	O_{ST}	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	A	.	.	T	.	.	G	.	.	.	.	T	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	A	A	C	.	C	.	T	T
Maf85.2	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	A	T	A	A	C	.	C	.	T	T
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	G	T	T
Bam16	O_{3+4}	.	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	C	.	.	.	.	.
Bam56	O_{3+4}	.	.	.	A	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.
Baf23	O_{3+4}	.	.	.	.	G	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	C	.	C	.	.	.
Baf32	O_{3+4}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	C	.	C	.	.	.
Baf46	O_{3+4}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.
Baf69	O_{3+4}	.	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	.	.	.	.	.	T	.	.	C	.	C	.	.	.
Baf81	O_{3+4}	.	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.
Baf148	O_{3+4}	.	.	.	A	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.
Baf167	O_{3+4}	.	.	.	.	A	.	.	T	.	.	G	.	.	.	.	T	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.
Baf185	O_{3+4}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	A	.	.	C	.	.	.
Mam16	O_{3+4}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	C	.	C	.	.	.
Mam27	O_{3+4}	.	.	.	.	A	.	.	T	.	.	G	.	.	.	.	T	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.
Maf1	O_{3+4}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	C	.	.	.	.	.
Maf5	O_{3+4}	A	.	.	A	A	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.
Maf11	O_{3+4}	.	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	C	.	C	.	.	.
Maf20	O_{3+4}	A	.	.	A	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.
Maf22	O_{3+4}	.	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	C	.	.	.	.	.
Maf34	O_{3+4}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	C	.	C	.	.	.
Maf78	O_{3+4}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.
Bam6	O_{3+4+8}	A	A	.	A	A	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.
Bam13	O_{3+4+8}	A	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.
Bam32	O_{3+4+8}	A	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.
Baf14	O_{3+4+8}	A	.	.	A	A	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	C	.	.	.	.	T
Baf20	O_{3+4+8}	.	A	.	A	A	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.
Baf21	O_{3+4+8}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	A	.	.	.	T	.	A	.	.	C	.	.	.
Baf53	O_{3+4+8}	.	.	.	.	A	.	.	T	.	.	G	.	.	.	.	T	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.
Baf166	O_{3+4+8}	A	.	.	A	A	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.
Baf159	O_{3+4+8}	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.
Baf196	O_{3+4+8}	A	.	.	A	A	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.
Mam28	O_{3+4+8}	A	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.
Maf17	O_{3+4+8}	A	.	.	A	A	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.
Maf13	O_{3+4+8}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	.	.	.	.	A	T	.	.	C	.	C	.	.	.
Maf25	O_{3+4+8}	A	.	.	.	A	.	.	T	.	.	G	.	.	.	T	.	A	C	.	A	A	.	.	.	T	.	A	.	C	C	.	.	.
Maf27	O_{3+4+8}	A	.	.	A	A	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.

Supplementary Tables S2

Acph-1		755	758	767	773	782	785	788	797	798*	800*	806	807	808	809	818	827	863	866	872	878	884	893	902	906	921*	923	935	941	945*	951*	962	963*	965
Bam73	O_{ST}	C	A	C	A	A	T	A	C	A	A	T	C	C	G	T	C	T	T	A	G	T	C	T	T	G	T	C	T	C	C	G	C	C
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
Baf75	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	C	.	.	.	A	T	.	.	.	.
Baf145	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
Baf147	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	T	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
Baf200	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	T	.	.	C	T	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.
Baf203	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.
Mam16	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
Maf36	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
Maf37	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	C	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Maf85	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.
Maf85.2	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.
Maf116	O_{ST}	.	G	A	T	G	.	.	.	.	.	.	T	.	.	C	T	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.
Bam16	O_{3+4}	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	.	.	.	.
Bam56	O_{3+4}	T	G	A	T	G	.	.	.	G	.	.	.	.	.	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	.	.	.	T
Baf23	O_{3+4}	.	.	.	.	.	A	G	.	.	.	C	.	.	A	.	.	C	A	.	.	.	.	.	.	T	C	.	.	T	.	.	.	.
Baf32	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	.	.	.	T
Baf46	O_{3+4}	.	.	.	.	.	.	G	T	.	.	.	.	.	A	.	.	C	A	.	.	A	.	C	.	.	C	.	.	T	.	.	.	.
Baf69	O_{3+4}	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	C	A	.	.	A	.	C	.	.	C	.	.	T	.	.	.	.
Baf81	O_{3+4}	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	.	.	.	T
Baf148	O_{3+4}	T	G	A	T	G	.	.	.	G	.	.	.	.	.	.	.	C	A	G	.	A	.	C	.	.	C	T	.	T	.	.	.	T
Baf167	O_{3+4}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	C	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Baf185	O_{3+4}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Mam16	O_{3+4}	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	A	.	.	.
Mam27	O_{3+4}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	C	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Maf1	O_{3+4}	.	.	.	.	.	.	G	T	.	.	.	.	.	A	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	.	.	.	.
Maf5	O_{3+4}	.	.	.	.	.	A	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Maf11	O_{3+4}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Maf20	O_{3+4}	T	G	A	T	G	.	.	.	G	.	.	.	.	.	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	.	.	.	T
Maf22	O_{3+4}	.	.	.	.	.	.	G	T	.	.	.	.	.	A	.	.	C	A	.	.	A	.	.	.	.	C	T	.	T	.	.	.	.
Maf34	O_{3+4}	.	.	.	.	.	.	G	T	.	.	.	.	T	A	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	A	.	.	.
Maf78	O_{3+4}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Bam6	O_{3+4+8}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Bam13	O_{3+4+8}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Bam32	O_{3+4+8}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	.	.	.	.
Baf14	O_{3+4+8}	.	G	.	.	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	A	.	.	.	T	C	.	.	T	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Baf21	O_{3+4+8}	.	G	.	.	.	.	G	T	.	.	.	.	.	A	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	.	A	.	.
Baf53	O_{3+4+8}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	C	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Baf166	O_{3+4+8}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	.	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Baf159	O_{3+4+8}	.	.	.	.	.	A	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	.	.	T	C	.	.	T	.	.	.	.
Baf196	O_{3+4+8}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	.	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Mam28	O_{3+4+8}	.	G	.	.	.	.	G	T	.	.	.	.	.	A	.	.	C	A	.	.	A	G	.	.	.	C	.	.	T	.	.	.	.
Maf17	O_{3+4+8}	.	.	.	.	.	.	G	T	.	.	C	.	.	A	.	.	C	A	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.
Maf13	O_{3+4+8}	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	C	A	.	.	A	.	C	.	.	C	.	.	T	.	.	.	.
Maf25	O_{3+4+8}	.	G	.	.	.	.	G	T	.	.	.	.	.	A	.	.	C	A	.	.	A	.	C	.	.	C	T	.	T	.	A	.	.
Maf27	O_{3+4+8}	.	.	.	.	.	.	G	T	.	G	C	.	.	A	.	.	C	.	.	.	A	.	.	.	.	C	.	.	T	.	.	.	.

Supplementary Tables S2

Acph-1		986	992	1007	1010*	1019	1034	1046	1061*	1096*	1099*	1100	1101*	1124	1145	1148*	1161*	1203	1205	1218	1220	1225*	1226	1241	1259	1262	1265	1283	1292	1309*	1317	1319	1324	1342
Bam73	O_{ST}	T	A	C	T	C	C	T	C	G	A	G	G	T	G	C	G	C	G	C	G	A	C	T	C	C	A	A	G	T	C	C	C	A
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G
Baf147	O_{ST}	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Baf200	O_{ST}	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Mam16	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Maf36	O_{ST}	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	T	A	.	.	.	.	T	.	.	.	.	.	T	.	.
Maf37	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Maf81	O_{ST}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	.	.	.	.	.	.	T	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	.	C	.	T	.	.	.	.	.	T	.	.
Maf85.2	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	.	C	.	T	.	.	.	.	.	T	.	.
Maf116	O_{ST}	C	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Bam16	O_{3+4}	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	G	.	.	.	.	.	.	.	.	.	T	.	.
Bam56	O_{3+4}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	.	.	.	.	.	.	T	.	.
Baf23	O_{3+4}	C	.	T	A	.	T	.	.	C	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	.	.	.	.	.	.	T	T	.
Baf32	O_{3+4}	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	.	.	.	.	.	T	.	.
Baf46	O_{3+4}	C	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	C	.	.	.	A	T	.	.
Baf69	O_{3+4}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	C	.	.	.	.	T	.	.
Baf81	O_{3+4}	C	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	A	.	T	.	.	.	.	.	.	.	.	T	.	.
Baf148	O_{3+4}	C	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	A	.	T	.	.	.	.	.	.	.	.	T	.	.
Baf167	O_{3+4}	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	T	C	.	.	.	A	T	.	.
Baf185	O_{3+4}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	C	.	.	.	.	T	.	.
Mam16	O_{3+4}	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	.	.	C	.	.	A	T	.	.
Mam27	O_{3+4}	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	A	T	.	.
Maf1	O_{3+4}	C	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	C	.	.	.	A	T	.	.
Maf5	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	G	.	.	.	.	.	C	.	.	.	T	.	.
Maf11	O_{3+4}	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	C	.	.	.	.	T	.	.
Maf20	O_{3+4}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	.	.	.	.	.	.	T	.	.
Maf22	O_{3+4}	C	G	.	.	T	T	A	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	C	.	.	.	.	T	.	.
Maf34	O_{3+4}	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	C	.	.	.	A	T	.	.
Maf78	O_{3+4}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	.	.	.	.	.	A	T	.	.
Bam6	O_{3+4+8}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	T	.	.
Bam13	O_{3+4+8}	C	G	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	T	.	.
Bam32	O_{3+4+8}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	A	.	.	T	.	.
Baf14	O_{3+4+8}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	C	C	A	.	.	T	.	.
Baf20	O_{3+4+8}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	T	.	.
Baf21	O_{3+4+8}	.	.	.	.	T	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	T	.	.
Baf53	O_{3+4+8}	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	A	.	A	T	.	.
Baf166	O_{3+4+8}	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	T	.	.
Baf159	O_{3+4+8}	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	.	.	C	.	.	.	T	.	.
Baf196	O_{3+4+8}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	T	.	.
Mam28	O_{3+4+8}	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	T	.	.
Maf17	O_{3+4+8}	C	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	T	C	C	A	.	.	G	.	.
Maf13	O_{3+4+8}	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	T	.	.	.	.	.	.	.	.	T	.	.
Maf25	O_{3+4+8}	.	.	.	.	T	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	T	.	.
Maf27	O_{3+4+8}	C	.	.	.	T	T	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	C	.	T	.	.

Supplementary Tables S2

Acph-1	1355	1358	1366	1368	1373	1387	1420	1421	1438	1447	1453	1473	1483	1507	1558	1560	1563	1627	1637	1653	1655	1668	1677	1679	1693	1711	1772	1773*	1775*	1804	1807	1808	1859
Bam73 O_{ST}	A	A	C	T	C	A	C	G	T	A	G	C	A	C	C	G	A	G	G	A	G	A	G	A	C	T	C	T	T	C	G	A	G
Baf58 O_{ST}	.	.	.	.	.	.	.	.	.	G	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75 O_{ST}	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	T	.	T	.	.	.	.	.	.	.	.	.	.	.
Baf145 O_{ST}	T	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	G	T	A	.	.	T	.	G	A	.	.	.	.	.	.	.	.
Baf147 O_{ST}	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf169 O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179 O_{ST}	.	.	.	.	.	.	.	.	C	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf200 O_{ST}	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.
Baf203 O_{ST}	.	.	.	.	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam16 O_{ST}	.	.	.	.	.	.	.	C	.	G	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Maf36 O_{ST}	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.
Maf37 O_{ST}	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf81 O_{ST}	.	.	.	.	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Maf85 O_{ST}	.	.	.	.	.	.	T	.	.	G	.	T	.	.	A	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	G	.
Maf85.2 O_{ST}	.	.	.	.	.	.	T	.	.	G	.	T	.	.	A	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	G	.
Maf116 O_{ST}	.	.	.	.	.	.	.	C	.	G	.	.	.	.	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	A	.	.	.
Bam16 O_{3+4}	.	.	.	C	.	.	.	.	C	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Bam56 O_{3+4}	.	.	.	.	.	.	.	.	C	G	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	A	.	.	.
Baf23 O_{3+4}	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf32 O_{3+4}	.	.	.	.	.	T	.	.	C	G	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	A	A	.	.
Baf46 O_{3+4}	.	.	.	.	.	T	.	.	C	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf69 O_{3+4}	.	.	.	C	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	A	.	.	.
Baf81 O_{3+4}	.	.	.	.	.	.	.	.	C	G	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf148 O_{3+4}	.	.	.	.	.	.	.	.	C	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.
Baf167 O_{3+4}	.	.	.	.	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf185 O_{3+4}	.	.	.	C	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	A	.	.	.
Mam16 O_{3+4}	.	.	.	.	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	A	.	.	.
Mam27 O_{3+4}	.	.	.	C	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	A	.	.	.
Maf1 O_{3+4}	.	.	.	.	.	T	.	.	C	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	A	.	.	.
Maf5 O_{3+4}	.	.	.	.	.	T	.	.	C	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Maf11 O_{3+4}	.	.	.	C	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	A	.	.	.
Maf20 O_{3+4}	.	.	.	C	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Maf22 O_{3+4}	.	.	.	.	.	.	.	.	C	G	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	T	.	.	A	.	.	.
Maf34 O_{3+4}	.	.	.	.	.	T	.	.	C	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Maf78 O_{3+4}	.	.	.	.	.	T	.	.	C	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Bam6 O_{3+4+8}	.	G	.	.	.	.	.	.	C	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A
Bam13 O_{3+4+8}	.	.	.	.	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	A	.	.	.
Bam32 O_{3+4+8}	.	.	.	.	.	.	.	.	C	G	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf14 O_{3+4+8}	.	.	.	.	.	T	.	.	C	G	.	.	.	.	A	T	.	.	.	T	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf20 O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A
Baf21 O_{3+4+8}	.	.	A	.	.	G	.	.	C	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf53 O_{3+4+8}	.	.	.	.	A	.	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf166 O_{3+4+8}	.	.	.	.	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf159 O_{3+4+8}	.	.	.	.	.	.	.	.	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf196 O_{3+4+8}	.	.	.	.	.	.	.	.	C	G	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	A	.	.	.
Mam28 O_{3+4+8}	.	.	.	.	.	T	.	.	C	G	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	A	.	.	.
Maf17 O_{3+4+8}	.	.	.	.	.	.	.	.	C	G	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Maf13 O_{3+4+8}	.	.	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	A	.	.	.
Maf25 O_{3+4+8}	.	.	.	.	.	G	.	.	C	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	A	.	.	.
Maf27 O_{3+4+8}	.	.	.	.	.	.	.	.	C	G	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.

Supplementary Tables S2

Acph-1		1861	1864	1865	1868*	1879
Bam73	O_{ST}	G	G	C	T	C
Baf58	O_{ST}	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.
Baf145	O_{ST}	.	.	A	.	.
Baf147	O_{ST}	.	T	.	.	.
Baf169	O_{ST}	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.
Baf200	O_{ST}	.	T	.	.	.
Baf203	O_{ST}	.	.	.	.	.
Mam16	O_{ST}	.	T	.	.	.
Maf36	O_{ST}	.	.	.	.	.
Maf37	O_{ST}	.	.	.	C	.
Maf81	O_{ST}	.	.	.	.	.
Maf85	O_{ST}	.	T	.	.	.
Maf85.2	O_{ST}	.	T	.	.	.
Maf116	O_{ST}	.	.	.	C	.
Bam16	O_{3+4}	C	.	.	.	.
Bam56	O_{3+4}	.	.	.	.	.
Baf23	O_{3+4}	.	.	.	.	.
Baf32	O_{3+4}	.	.	.	.	.
Baf46	O_{3+4}	C	.	.	.	.
Baf69	O_{3+4}	.	.	.	.	.
Baf81	O_{3+4}	.	.	.	.	.
Baf148	O_{3+4}	.	.	.	.	.
Baf167	O_{3+4}	.	.	.	.	.
Baf185	O_{3+4}	.	.	.	.	.
Mam16	O_{3+4}	.	.	.	.	T
Mam27	O_{3+4}	.	.	.	.	.
Maf1	O_{3+4}	.	.	.	.	.
Maf5	O_{3+4}	.	.	.	.	.
Maf11	O_{3+4}	.	.	.	.	.
Maf20	O_{3+4}	.	.	.	.	.
Maf22	O_{3+4}	.	.	.	.	.
Maf34	O_{3+4}	.	.	.	.	.
Maf78	O_{3+4}	C	.	.	.	.
Bam6	O_{3+4+8}	.	.	.	.	.
Bam13	O_{3+4+8}	.	.	.	.	.
Bam32	O_{3+4+8}	.	.	.	.	.
Baf14	O_{3+4+8}	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	.	.
Baf21	O_{3+4+8}	.	.	.	.	.
Baf53	O_{3+4+8}	.	.	.	.	.
Baf166	O_{3+4+8}	.	.	.	.	.
Baf159	O_{3+4+8}	.	.	.	.	.
Baf196	O_{3+4+8}	.	.	.	.	.
Mam28	O_{3+4+8}	.	.	.	.	.
Maf17	O_{3+4+8}	.	.	.	.	.
Maf13	O_{3+4+8}	.	.	.	.	.
Maf25	O_{3+4+8}	.	.	.	.	.
Maf27	O_{3+4+8}	.	.	.	.	.

Supplementary Tables S2

Ast		39	148	179	189	190	203	224	229	248	264	347	371	392	393	396	397	454	457	463	466	477	482	487	495	503	509	530	532	540	549	550	551	564	566	576
Bam73	O_{ST}	T	T	A	G	C	A	A	T	G	G	C	A	G	A	C	T	T	A	C	G	A	T	G	T	G	G	C	G	G	G	T	C	C	G	T
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	.	.	A
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf200	O_{ST}	.	.	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf37	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.
Bam16	O_{3+4}	A	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Baf23	O_{3+4}	A	A	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	.	T	A	.	C	.	.	.	.	T	C	.	C	.	.	.	.	.
Baf69	O_{3+4}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	T	A	.	.	.	A	.	C	.	.	.	.	T	.	A	C	.	T	.	.	.
Baf46	O_{3+4}	A	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.	.	T	.	A	.	C	.	.	.	.	.	.	.	C	.	.	.	.	.
Baf81	O_{3+4}	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	.	.	.	.	T	.	.	C	.	.	.	.	.
Baf148	O_{3+4}	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Baf185	O_{3+4}	A	.	.	.	.	.	.	C	.	C	.	.	.	.	T	A	.	.	.	A	.	C	.	.	A	.	T	.	.	C	.	.	.	.	.
Mam16	O_{3+4}	A	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	T	.	A	.	C	.	.	.	.	.	.	.	C	.	.	.	.	.
Mam27	O_{3+4}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	T	A	.	.	.	A	.	C	.	.	.	C	.	.	.	C	.	.	.	.	.
Mam37	O_{3+4}	A	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	T	.	A	.	C	.	.	.	.	.	.	.	C	.	.	.	.	.
Maf1	O_{3+4}	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	.	.	.	.	T	.	.	C	.	.	.	.	.
Maf5	O_{3+4}	A	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.	.	.	.	A	.	C	.	.	.	.	.	.	.	C	.	.	.	.	.
Maf11	O_{3+4}	A	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	A	.	C	T	.	.	.	.	.	.	C	.	.	.	.	.
Maf22	O_{3+4}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	C	.	.	.	.	T	.	.	C	.	.	.	.	.
Maf34	O_{3+4}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Maf78	O_{3+4}	A	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	T	.	A	.	C	.	.	.	.	.	.	.	C	.	.	.	.	.
Bam6	O_{3+4+8}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Bam13	O_{3+4+8}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	A	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Bam32	O_{3+4+8}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Baf14	O_{3+4+8}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Baf21	O_{3+4+8}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	C	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Baf53	O_{3+4+8}	A	.	.	.	.	.	G	C	.	C	.	.	.	.	.	.	.	.	.	A	.	C	.	.	.	.	T	.	.	C	.	.	.	.	.
Baf159	O_{3+4+8}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Baf166	O_{3+4+8}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Baf196	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf17	O_{3+4+8}	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	.	C	.	.	.	.	.
Maf25	O_{3+4+8}	A	.	.	.	.	C	.	.	.	.	.	.	.	C	.	.	.	T	.	A	.	C	.	.	.	.	.	.	.	C	.	.	.	.	.
Maf56	O_{3+4+8}	A	.	.	.	.	C	.	.	.	.	.	.	.	C	.	.	.	T	.	A	.	C	.	.	.	.	.	.	.	C	.	.	.	.	.

Supplementary Tables S2

Ast		582	592	595	603	604	608	612	620	632	642	644	650	661	674	679	682	701	716	775	777	781	784	792	801	808	812	841	852	880	882	885	895	900	907	909
Bam73	O_{ST}	C	G	C	G	G	C	T	G	T	C	A	G	G	G	A	A	C	T	T	A	T	C	A	A	G	C	T	T	C	T	A	T	A	G	G
Baf58	O_{ST}	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
Baf75	O_{ST}	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	T	T	.	.	C	C	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.
Baf169	O_{ST}	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf200	O_{ST}	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	T	.	A	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Maf37	O_{ST}	.	.	.	.	A	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	T	.	.	.	.	.	.	T	.	.	G	.	A	.	T	T	.	.	.	.	C	.	.	.	A	.	.	.	.	.	.	A	.	.	.
Baf23	O_{3+4}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	T	T	.	.	C	C	C	.	.	.	.	.	A	.	.	.	.	A	.	.	.
Baf69	O_{3+4}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.
Baf46	O_{3+4}	T	.	.	.	.	A	.	T	.	.	.	.	.	.	T	T	.	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	T	.
Baf81	O_{3+4}	.	.	.	.	A	.	.	.	G	.	.	.	.	.	.	.	.	.	C	C	C	.	.	.	A	.	.	.	.	.	.	.	.	.	.
Baf148	O_{3+4}	T	.	.	.	.	.	.	T	.	A	G	.	A	.	T	T	.	.	.	.	C	A	.	T	A	.	.	.	.	.	.	.	.	.	.
Baf185	O_{3+4}	T	.	.	.	.	.	.	T	.	.	.	.	.	A	T	T	.	.	C	C	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.
Mam16	O_{3+4}	T	.	.	.	.	.	.	T	.	.	G	.	A	.	T	T	.	.	C	C	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.
Mam27	O_{3+4}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	C	C	C	.	.	.	.	.	A	C	.	.	.	A	.	.	.
Mam37	O_{3+4}	T	.	.	.	.	.	.	T	.	.	G	.	A	.	T	T	.	.	C	C	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.
Maf1	O_{3+4}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	T	T	.	.	.	C	C	.	.	.	A	.	.	.	.	.	.	.	.	.	.
Maf5	O_{3+4}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	T	T	.	.	C	C	C	.	.	.	.	.	A	.	.	C	G	.	G	.	.
Maf11	O_{3+4}	T	.	.	C	.	.	.	T	.	.	.	.	.	.	T	T	.	.	C	C	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf22	O_{3+4}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	T	T	.	.	C	C	C	.	.	.	.	.	A	.	.	.	.	A	.	.	.
Maf34	O_{3+4}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	T	T	.	.	C	C	C	.	.	.	.	.	A	.	.	C	G	.	G	.	.
Maf78	O_{3+4}	T	.	.	.	.	.	.	T	.	.	G	.	A	.	T	T	.	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	T	.
Bam6	O_{3+4+8}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	A	.	.	C	G	.	G	.	.
Bam13	O_{3+4+8}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	T	T	.	.	.	C	C	.	.	.	.	.	A	.	.	C	G	.	G	.	.
Bam32	O_{3+4+8}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	C	.	.	.	.	.	A	.	.	C	G	.	G	.	.
Baf14	O_{3+4+8}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	C	C	C	.	.	.	.	.	A	.	T	C	G	.	G	.	.
Baf21	O_{3+4+8}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	C	C	C	.	.	.	.	.	A	.	.	C	G	.	G	.	.
Baf53	O_{3+4+8}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	T	T	.	.	.	C	C	.	.	.	A	.	.	.	.	.	.	.	.	.	.
Baf159	O_{3+4+8}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	T	T	.	.	.	.	C	.	.	.	.	.	A	.	.	C	G	.	G	.	.
Baf166	O_{3+4+8}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	T	T	.	.	.	.	C	.	.	.	.	.	A	.	.	C	G	.	G	.	.
Baf196	O_{3+4+8}	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf17	O_{3+4+8}	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	C	.	.	.	.	.	A	.	.	C	G	.	G	.	.
Maf25	O_{3+4+8}	T	.	.	.	.	.	G	T	.	.	.	.	.	.	T	T	.	.	C	C	C	.	G	.	.	T	A	.	.	C	G	.	G	.	.
Maf56	O_{3+4+8}	T	.	.	.	.	.	G	T	.	.	.	.	.	.	T	T	.	.	C	C	C	.	G	.	.	T	A	.	.	C	G	.	G	.	.

Supplementary Tables S2

Ast		918	923	959	970	1006	1039	1040	1050	1065	1066	1070	1081	1092	1106	1176	1242	1251	1253	1273	1278	1306	1319	1344	1367	1388	1393	1396	1415	1419	1450	1463	1511	1522*	1535*	1536*
Bam73	O_{ST}	T	T	G	G	C	C	A	C	T	C	G	G	G	C	T	G	A	C	A	A	T	A	A	T	G	C	C	G	C	C	G	C	C	C	C
Baf58	O_{ST}	.	.	.	.	T	.	T	.	.	A	.	.	.	.	G	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	.	.	.	T	.	.	.	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
Baf169	O_{ST}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf179	O_{ST}	.	.	.	.	T	.	.	.	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
Baf200	O_{ST}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf37	O_{ST}	A	C	.	A	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	A	C	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	G	.
Baf23	O_{3+4}	A	C	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	G	.
Baf69	O_{3+4}	A	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	G	.
Baf46	O_{3+4}	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Baf81	O_{3+4}	A	C	.	A	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Baf148	O_{3+4}	A	C	.	.	T	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Baf185	O_{3+4}	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	A	.	C	.	.	.	.	.	.	.	.	.	T	G	.
Mam16	O_{3+4}	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	T	.	T	.	.	.	.	.	.	A	.	.	.	.	.	T	G	.
Mam27	O_{3+4}	A	C	.	.	T	.	.	.	.	.	.	C	.	.	G	.	T	.	.	.	.	G	C	.	.	.	.	.	.	.	.	.	T	G	.
Mam37	O_{3+4}	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	T	.	T	.	.	.	.	.	.	A	.	.	.	.	.	T	G	.
Maf1	O_{3+4}	A	C	.	.	T	.	.	.	.	.	C	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Maf5	O_{3+4}	A	C	C	.	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	T	G	.
Maf11	O_{3+4}	A	C	.	.	T	.	.	A	.	.	.	A	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	A	.	.	C	.	T	G	.
Maf22	O_{3+4}	A	C	.	.	T	A	.	.	.	.	.	A	.	.	.	T	T	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	T	G	A
Maf34	O_{3+4}	A	C	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
Maf78	O_{3+4}	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	T	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	G	.
Bam6	O_{3+4+8}	A	C	.	.	T	.	.	.	G	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Bam13	O_{3+4+8}	A	C	.	.	T	.	.	.	G	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Bam32	O_{3+4+8}	A	C	.	.	T	.	.	.	G	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Baf14	O_{3+4+8}	A	C	.	.	T	.	.	.	G	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Baf21	O_{3+4+8}	A	C	.	.	T	.	.	.	G	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Baf53	O_{3+4+8}	A	C	.	.	T	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	G	.
Baf159	O_{3+4+8}	A	C	.	.	T	.	.	.	G	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Baf166	O_{3+4+8}	A	C	.	.	T	.	.	.	G	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	C	.	T	G	.
Baf196	O_{3+4+8}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Maf17	O_{3+4+8}	A	C	.	.	T	.	.	.	G	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Maf25	O_{3+4+8}	A	C	.	.	T	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	G	.
Maf56	O_{3+4+8}	A	C	.	.	T	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	G	.

Supplementary Tables S2

Ast		1538	1555*	1577	1583	1628	1649	1682	1700	1724	1731*	1739	1849	1869	1870
Bam73	O_{ST}	G	C	T	C	G	C	A	G	C	T	G	G	A	T
Baf58	O_{ST}	.	G	.	.	.	.	.	A	.	.	.	.	.	.
Baf75	O_{ST}	.	G	A	.	.	G	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	G	.	.	.	G	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	G	A	.	.	G	.	.	.	.	.	.	.	.
Baf169	O_{ST}	.	G	A	.	.	.	G	.	.	.	.	.	.	.
Baf179	O_{ST}	.	G	.	.	.	.	.	A	.	.	.	.	.	.
Baf200	O_{ST}	.	G	A	.	.	G	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	G	A	.	.	G	.	.	.	.	.	.	.	.
Maf37	O_{ST}	.	G	.	.	.	G	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	G	.	.	.	G	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	G	A	.	.	G	.	.	.	.	.	.	.	A
Maf116	O_{ST}	.	G	A	.	.	G	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	A	G	G	A	.	G	G	.	A	.	.	.	.	.
Baf23	O_{3+4}	A	G	G	A	.	G	G	.	A	.	.	.	.	.
Baf69	O_{3+4}	A	G	G	A	.	G	.	.	A	.	.	.	.	.
Baf46	O_{3+4}	A	G	G	A	.	G	.	.	.	.	A	.	.	.
Baf81	O_{3+4}	A	G	G	A	A	G	.	.	.	.	.	.	T	.
Baf148	O_{3+4}	A	G	G	A	.	G	.	.	.	.	.	.	.	.
Baf185	O_{3+4}	A	G	G	A	.	G	G	.	.	.	.	.	.	.
Mam16	O_{3+4}	A	G	G	A	.	G	.	.	.	.	.	.	.	.
Mam27	O_{3+4}	A	G	G	A	.	G	.	.	.	.	.	.	.	.
Mam37	O_{3+4}	A	G	G	A	.	G	G	.	A	.	.	.	.	.
Maf1	O_{3+4}	A	G	G	A	A	G	.	.	.	.	.	.	T	.
Maf5	O_{3+4}	A	G	G	A	.	G	G	.	A	.	.	.	.	.
Maf11	O_{3+4}	A	G	G	A	.	G	.	.	A	.	.	.	.	.
Maf22	O_{3+4}	A	G	G	A	.	G	.	.	.	.	A	.	.	.
Maf34	O_{3+4}	.	G	.	.	.	G	.	.	.	.	.	.	.	.
Maf78	O_{3+4}	A	G	G	A	.	G	.	.	.	.	A	.	.	.
Bam6	O_{3+4+8}	A	G	G	A	A	G	.	.	.	.	.	.	.	.
Bam13	O_{3+4+8}	A	G	G	A	.	G	.	.	.	.	A	.	.	.
Bam32	O_{3+4+8}	A	G	G	A	.	G	.	.	.	.	A	.	.	.
Baf14	O_{3+4+8}	A	G	G	A	.	G	.	.	.	.	.	.	.	.
Baf21	O_{3+4+8}	A	G	G	A	.	G	.	.	.	.	A	.	.	.
Baf53	O_{3+4+8}	A	G	G	A	.	G	.	.	.	.	A	.	.	.
Baf159	O_{3+4+8}	A	G	G	A	.	G	G	.	.	.	A	.	.	.
Baf166	O_{3+4+8}	A	G	G	A	A	G	.	.	.	.	.	.	.	.
Baf196	O_{3+4+8}	.	G	A	.	.	G	G	.	.	.	.	.	.	.
Maf17	O_{3+4+8}	A	G	G	A	.	G	.	.	.	.	A	.	.	.
Maf25	O_{3+4+8}	A	G	G	A	.	G	.	.	.	C	A	.	.	.
Maf56	O_{3+4+8}	A	G	G	A	.	G	.	.	.	.	.	A	.	.

Supplementary Tables S2

larp		9	96	152	157	160	167	168	171	174	175	182	183	193	201	213	214	218	240*	350	415*	452	459*	479*	499*	500*	501*	513*	516*	540*	543*	554	557	564*	565*	570*
Bam73	O_{ST}	G	T	C	C	T	C	C	T	C	A	C	A	A	C	C	C	C	G	G	G	T	G	G	A	G	G	C	G	G	G	T	G	G	T	G
Baf58	O_{ST}	.	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
Baf100	O_{ST}	.	.	.	.	G	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	A	.	.	.	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Baf203	O_{ST}	.	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Maf37	O_{ST}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Maf84	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	.	A	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	A	T	.	.	C	A	.	T	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Maf85.2	O_{ST}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	T	.	T	.	.	.	C	.	.	.	.	A	T	.	.	C	A	.	T	.	.
Bam16	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Bam56	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Baf23	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Baf32	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Baf46	O_{3+4}	.	.	.	.	.	.	.	.	T	T	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	T
Baf69	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Baf81	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Baf88	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Baf148	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Baf167	O_{3+4}	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Baf185	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.
Mam16	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Mam27	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Mam37	O_{3+4}	A	.	.	.	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Maf1	O_{3+4}	A	.	.	.	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Maf5	O_{3+4}	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	A	A	.	C	.
Maf11	O_{3+4}	.	.	.	.	.	.	T	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Maf20	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Maf34	O_{3+4}	.	.	.	.	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Maf78	O_{3+4}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Bam6	O_{3+4+8}	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Bam13	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Bam32	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Baf14	O_{3+4+8}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Baf20	O_{3+4+8}	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Baf21	O_{3+4+8}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Baf53	O_{3+4+8}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Baf159	O_{3+4+8}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	C	.
Baf166	O_{3+4+8}	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Baf196	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Mam25	O_{3+4+8}	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	T	.	T	.	.	T	A	.	.	.	.	A	A	.	.	.
Mam28	O_{3+4+8}	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.
Mam31	O_{3+4+8}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	.	A	A	.	.	.
Maf13	O_{3+4+8}	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	.	A	A	.	.	.
Maf17	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	A	.	.	.

Supplementary Tables S2

larp		572	578	584	731*	749	758	764	798	809	843*	860	866	871*	930*	931*	935	944	945*	987*	1069*	1096*	1123*	1176*	1202	1224*	1227*	1246*	1276*	1301	1320*	1409	1490	1542	1580
Bam73	O_{ST}	A	A	C	G	G	C	G	C	T	C	A	G	G	G	T	T	G	G	G	C	T	A	A	C	T	T	C	C	C	G	C	C	C	A
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf100	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf145	O_{ST}	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	T	.
Maf37	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf84	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85.2	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Bam56	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf23	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf32	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	A	.	.	.	.	.
Baf46	O_{3+4}	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf69	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.
Baf81	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.	.	.	.
Baf88	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf148	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf167	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf185	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Mam16	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Mam27	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Mam37	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	A	.	.	T
Maf1	O_{3+4}	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Maf5	O_{3+4}	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.
Maf11	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	T	.	.	.
Maf20	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Maf34	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Maf78	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Bam6	O_{3+4+8}	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Bam13	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Bam32	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf14	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf21	O_{3+4+8}	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf53	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf159	O_{3+4+8}	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Baf166	O_{3+4+8}	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.
Baf196	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Mam25	O_{3+4+8}	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	A	.	.	.	.	.
Mam28	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Mam31	O_{3+4+8}	.	.	.	.	A	.	.	.	C	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Maf13	O_{3+4+8}	G	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Maf17	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.

Supplementary Tables S2

larp		1613	1690*	1711*	1787	1823	1865*	1868	1895	1898	1922	1964	1990*	2003	2009	2012
Bam73	O_{ST}	A	T	A	G	G	G	G	C	T	G	G	G	C	C	A
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
Baf100	O_{ST}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	A	.	C	.	.	.	.	.	.
Maf37	O_{ST}	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Maf84	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85.2	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	G	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Bam56	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf23	O_{3+4}	.	.	.	.	.	.	.	.	C	.	A	.	.	T	.
Baf32	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf46	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf69	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf81	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf88	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf148	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf167	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf185	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Mam16	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Mam27	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Mam37	O_{3+4}	.	.	.	.	.	.	A	.	C	.	.	.	.	.	.
Maf1	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.
Maf5	O_{3+4}	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.
Maf11	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Maf20	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Maf34	O_{3+4}	.	.	.	.	.	.	A	T	C	.	.	.	.	.	.
Maf78	O_{3+4}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Bam6	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Bam13	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Bam32	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf14	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf21	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf53	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf159	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	G
Baf166	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Baf196	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Mam25	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Mam28	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Mam31	O_{3+4+8}	.	.	G	.	.	.	.	.	C	.	.	.	.	.	.
Maf13	O_{3+4+8}	.	G	.	A	.	.	.	.	C	.	.	.	.	.	.
Maf17	O_{3+4+8}	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.

Supplementary Tables S2

reg_*larp*		13	52	72	114	121	123	125	131	132	133	140	142	145	156	166	168	180	193	198	204	207	248	256	258	274	364	401	421	429	436	447	461	462	475	492
Bam73	O_{ST}	T	T	G	T	T	T	A	T	A	G	C	T	A	T	C	A	C	G	C	G	A	T	C	C	C	C	C	C	C	C	A	T	A	A	C
Baf58	O_{ST}	.	.	.	.	G	.	G	A	T	T	.	C	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	G	.	G	A	T	T	.	C	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ba145	O_{ST}	.	.	.	.	.	A	.	.	T	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	A	G	.	G	A	T	T	.	C	.	.	.	C	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
Maf25	O_{ST}	.	.	.	.	G	.	G	.	T	.	.	.	.	.	.	C	.	T	A	.	C	.	.	.	.	.	.	T	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Bam16	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Bam56	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T	.
Baf23	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf32	O_{3+4}	A	.	.	.	G	.	G	A	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf46	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
Baf69	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	.	T	G	.	.	.	.
Baf81	O_{3+4}	A	.	.	.	G	.	G	.	T	.	.	.	C	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf148	O_{3+4}	A	.	.	.	G	.	G	.	T	.	.	.	C	.	T	C	.	.	A	.	.	.	.	G	.	.	.	.	.	.	G	G	.	.	.
Baf167	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	T	C	.	.	A	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
Baf185	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Maf1	O_{3+4}	A	.	.	.	G	.	G	A	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Maf11	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	.	C	T	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf34	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	G	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam16	O_{3+4}	A	.	.	.	G	.	G	A	T	.	.	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Mam37	O_{3+4}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	.	C	T	.	A	.	.	A	.	.	.	.	.	.	.	.	G	.	.	.	.
Bam6	O_{3+4+8}	A	.	.	.	G	.	G	A	T	.	G	.	.	.	T	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam13	O_{3+4+8}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	T	C	.	.	A	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
Bam32	O_{3+4+8}	A	A	T	.	G	.	G	.	T	.	.	.	.	.	.	C	.	T	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf14	O_{3+4+8}	.	.	.	.	G	.	G	.	T	.	.	.	.	.	.	C	.	T	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	.	G	.	G	A	T	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf21	O_{3+4+8}	A	.	.	.	G	.	G	.	T	.	.	.	.	C	T	C	.	.	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf159	O_{3+4+8}	A	.	.	.	G	.	G	.	T	.	.	.	.	.	.	C	.	T	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf166	O_{3+4+8}	.	.	.	.	G	.	G	.	T	.	.	.	.	.	.	C	.	T	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf17	O_{3+4+8}	A	.	.	.	G	.	G	.	T	.	.	.	.	.	.	C	.	T	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam28	O_{3+4+8}	A	.	.	.	G	.	G	.	T	.	.	.	.	.	.	C	.	T	A	.	C	.	.	.	.	.	A	.	T	.	.	.	.	.	.

Supplementary Tables S2

reg_*larp*		496	511	519	521	536	537	540	551	555	560	568	578	579	606	622	678	695	703	740	741	772	780	799	800	801	802	803	804	805	823	857	888	895	896	898
Bam73	O_{ST}	G	C	T	T	G	G	C	G	G	C	G	C	C	A	C	C	C	A	G	T	G	T	T	A	T	C	T	T	T	T	C	G	G	G	A
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ba145	O_{ST}	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf25	O_{ST}	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	A	G	.	.	.	.	A	.	A	G	A	.	A	.	.	.	.	.	.
Bam16	O_{3+4}	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam56	O_{3+4}	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf23	O_{3+4}	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf32	O_{3+4}	.	.	.	G	T	A	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf46	O_{3+4}	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Baf69	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf81	O_{3+4}	.	.	.	.	T	A	.	.	.	A	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf148	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf167	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf185	O_{3+4}	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf1	O_{3+4}	T	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.
Maf11	O_{3+4}	.	A	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.
Maf34	O_{3+4}	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam16	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam37	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.
Bam6	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam13	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
Bam32	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf14	O_{3+4+8}	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf21	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	G	C	.	.	A	.	.	.	.	.	.	.	.	G	.	.	.	.
Baf159	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
Baf166	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.
Maf17	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam28	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Supplementary Tables S2

reg_*larp*		900	920	922	929	933	935	940	1007	1009	1015	1026	1031	1033	1034	1036	1041	1043	1044	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1064
Bam73	O_{ST}	A	G	T	C	A	T	T	T	C	G	G	G	T	A	A	T	A	C	C	A	A	T	A	G	A	T	C	A	G	T	A	G	A
Baf58	O_{ST}	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ba145	O_{ST}	.	.	.	T	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	G	G	.	C	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf25	O_{ST}	.	.	.	.	.	.	.	.	T	.	.	.	C	.	.	C	.	T	A	T	C	A	G	T	T	G	A	.	A	A	G	C	T
Maf85	O_{ST}	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	.	.	G	G	.	C	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Bam56	O_{3+4}	G	.	G	G	.	C	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf23	O_{3+4}	.	.	G	G	.	C	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf32	O_{3+4}	.	.	G	G	.	G	.	A	T	C	T	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf46	O_{3+4}	.	.	G	G	.	C	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf69	O_{3+4}	.	.	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf81	O_{3+4}	.	A	G	G	G	.	.	A	A	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf148	O_{3+4}	.	.	.	G	.	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf167	O_{3+4}	.	.	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf185	O_{3+4}	.	.	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Maf1	O_{3+4}	.	.	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Maf11	O_{3+4}	.	.	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Maf34	O_{3+4}	.	.	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Mam16	O_{3+4}	.	.	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Mam37	O_{3+4}	.	.	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Bam6	O_{3+4+8}	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	C	.	T	A	T	C	A	G	T	T	G	A	.	A	A	G	C	T
Bam13	O_{3+4+8}	.	.	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Bam32	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	T	A	T	C	A	G	T	T	G	A	.	A	A	G	C	T
Baf14	O_{3+4+8}	.	A	G	G	G	.	.	A	T	C	.	A	C	T	.	.	C	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	C	.	T	A	T	C	A	G	T	T	G	A	.	A	A	G	C	T
Baf21	O_{3+4+8}	.	.	.	.	G	.	.	.	.	.	.	.	C	.	.	C	.	T	A	T	C	A	G	T	T	G	A	.	A	A	G	C	T
Baf159	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	C	.	T	A	T	C	A	G	T	T	G	A	.	A	A	G	C	T
Baf166	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	T	A	T	C	A	G	T	T	G	A	.	A	A	G	C	T
Maf17	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	T	A	T	C	A	G	T	T	G	A	.	A	A	G	C	T
Mam28	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	T	A	T	C	A	G	T	T	G	A	.	A	A	G	C	T

Supplementary Tables S2

reg_*larp*		1065	1067	1068	1070	1074	1078	1080	1120	1121	1123	1134	1162	1191	1192	1194	1220	1235	1251	1293	1295	1297	1307	1317	1368	1369	1410	1490	1596	1501	1503	1504	1505	1511
Bam73	O_{ST}	T	A	T	A	C	T	G	T	A	G	T	C	C	C	A	T	T	G	A	A	A	G	C	A	C	C	G	C	A	A	C	G	T
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	C	T	G	.	G	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	A	.	.	A	.	.	.	.	.	.	.	.	G	.	.	.	.	.	G	.	.	T	.	.	.	T	.	.	.
Ba145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	G	T	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	C	T	G	.	G	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	C
Baf200	O_{ST}	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	G	.	C	T	G	.	G	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	C	T	G	.	G	.	.	.	.	.	.	.	.	.	.
Maf25	O_{ST}	C	T	A	.	A	A	.	A	.	.	A	.	.	.	.	.	G	.	.	T	G	.	G	.	T	.	A	.	.	.	.	.	C
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	C	T	G	.	G	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	.	C	.	G	.	G	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	C
Bam56	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	C
Baf23	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	G	.	C
Baf32	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	G	G	.	.	G	.	.	.	G	T	G	.	.	.	.	.	.	.	.	.	C
Baf46	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	A	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	G	.	C
Baf69	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	T	G	.	G	.	.	.	.	.	.	.	.	.	.
Baf81	O_{3+4}	.	.	.	G	.	A	.	A	.	A	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	C
Baf148	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	C
Baf167	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	T	C
Baf185	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	C
Maf1	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	C	.	G	.	G	.	.	.	.	T	.	.	.	.	C
Maf11	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	C
Maf34	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	T	G	.	G	.	.	.	.	.	.	.	.	.	C
Mam16	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	T	G	.	G	.	.	.	.	.	.	.	.	.	C
Mam37	O_{3+4}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	C	.	G	.	G	.	.	.	.	.	.	.	.	.	C
Bam6	O_{3+4+8}	C	T	A	.	A	A	.	A	.	.	A	.	.	.	C	.	G	.	.	.	G	.	G	.	T	.	A	.	.	.	.	.	C
Bam13	O_{3+4+8}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	T	C
Bam32	O_{3+4+8}	C	T	A	.	A	A	.	A	.	.	A	.	.	.	.	.	G	.	.	T	G	.	G	.	T	.	A	.	.	.	.	.	C
Baf14	O_{3+4+8}	.	.	.	G	.	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	G	.	.	.	C
Baf20	O_{3+4+8}	C	T	A	.	A	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	C
Baf21	O_{3+4+8}	C	T	A	.	A	A	.	A	.	.	A	.	.	.	C	.	G	A	.	T	G	.	G	.	T	.	A	.	.	.	.	.	C
Baf159	O_{3+4+8}	C	T	A	.	A	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	.	T	.	.	.	.	C
Baf166	O_{3+4+8}	C	T	A	.	A	A	.	A	.	.	A	.	.	.	.	.	G	.	.	.	G	.	G	.	T	.	A	T	.	.	.	.	C
Maf17	O_{3+4+8}	C	T	A	.	A	A	.	A	.	.	A	.	.	.	.	.	G	.	.	T	G	.	G	.	T	.	A	.	.	.	.	.	C
Mam28	O_{3+4+8}	C	T	A	.	A	A	A	A	.	.	A	.	.	.	C	.	G	.	.	.	G	.	G	.	T	.	A	.	.	.	.	.	C

Supplementary Tables S2

reg_*larp*		1542	1581	1587	1611	1629	1631	1673	1676	1677	1694	1697	1701	1709	1711	1714	1716	1720	1726
Bam73	O_{ST}	A	T	A	A	A	G	T	G	C	C	A	G	T	T	A	T	T	G
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
Baf75	O_{ST}	T	.	.	.	T	C	G	T	.	A	.	.	.	.	.	.	.	.
Ba145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
Maf25	O_{ST}	.	.	.	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Maf85	O_{ST}	.	.	.	.	.	.	G	T	.	A	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	.	.	.	G	.	.	.	.	A	.	G	.	C	A	.	G	.	A
Bam56	O_{3+4}	.	.	.	G	.	.	.	.	A	.	G	A	.	.	.	.	.	.
Baf23	O_{3+4}	.	C	T	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Baf32	O_{3+4}	.	C	T	G	.	.	.	.	A	.	G	.	.	.	.	G	.	A
Baf46	O_{3+4}	.	C	T	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Baf69	O_{3+4}	.	C	T	G	.	.	.	.	A	A	.	.	.	.	.	.	.	A
Baf81	O_{3+4}	.	.	.	G	.	.	.	.	A	.	G	.	C	A	.	G	.	A
Baf148	O_{3+4}	.	C	T	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Baf167	O_{3+4}	.	.	.	G	.	.	.	.	A	.	G	.	.	.	.	G	.	A
Baf185	O_{3+4}	.	.	.	G	.	.	.	.	A	.	G	.	.	.	.	G	.	A
Maf1	O_{3+4}	.	C	T	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Maf11	O_{3+4}	.	.	.	G	.	.	.	.	A	.	G	.	.	.	.	.	.	.
Maf34	O_{3+4}	.	.	.	G	.	.	.	.	A	.	G	.	C	A	.	G	.	A
Mam16	O_{3+4}	.	C	T	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Mam37	O_{3+4}	.	.	.	G	.	.	.	.	A	.	G	A	.	.	.	.	.	.
Bam6	O_{3+4+8}	.	C	T	G	.	.	.	.	A	.	G	.	C	A	G	G	.	A
Bam13	O_{3+4+8}	.	.	.	G	.	.	.	.	A	.	G	.	.	.	.	G	.	A
Bam32	O_{3+4+8}	.	.	.	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Baf14	O_{3+4+8}	.	C	T	G	.	.	.	.	A	.	G	.	.	.	.	G	.	A
Baf20	O_{3+4+8}	.	.	.	G	.	.	.	.	A	.	G	.	.	.	.	G	.	A
Baf21	O_{3+4+8}	.	.	.	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Baf159	O_{3+4+8}	.	C	T	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Baf166	O_{3+4+8}	.	C	T	G	.	.	.	.	A	.	G	.	C	.	.	G	C	A
Maf17	O_{3+4+8}	.	.	.	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A
Mam28	O_{3+4+8}	.	.	.	G	.	.	.	.	A	.	G	.	C	.	.	G	.	A

Supplementary Tables S2

CG5961		29*	32	34	46	58	83	122	130	142	151	188*	286	295	298	299	336*	367	391	406	412	496	505	556	562*	584*
Bam73	O_{ST}	A	C	T	T	A	C	C	G	A	C	A	C	A	G	C	C	T	T	G	G	G	C	T	G	G
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.	T	.	.	.	.	.	.	.	.	.
Maf37	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Bam16	O_{3+4}	.	.	G	A	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Bam56	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	.	.
Baf23	O_{3+4}	.	.	G	A	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Baf32	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	.	.
Baf46	O_{3+4}	.	A	G	.	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Baf69	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	C	.	.
Baf81	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	.	.
Baf148	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Baf167	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	.	.
Baf185	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	C	.
Mam16	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	.	.
Mam27	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	.	.
Mam37	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Maf1	O_{3+4}	.	.	G	A	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	.	.
Maf5	O_{3+4}	.	.	G	A	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Maf20	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	.	.
Maf22	O_{3+4}	.	A	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	A	.	.	.	.	.	.
Maf34	O_{3+4}	.	.	G	.	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Maf78	O_{3+4}	.	.	G	A	G	.	A	A	G	.	.	.	T	T	T	.	.	C	.	.	.	.	.	.	.
Bam6	O_{3+4+8}	.	.	G	A	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Bam13	O_{3+4+8}	.	.	G	A	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Bam32	O_{3+4+8}	.	.	G	.	G	T	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Baf14	O_{3+4+8}	.	.	G	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Baf21	O_{3+4+8}	.	A	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	A	.	.	.	.	.
Baf53	O_{3+4+8}	G	.	G	.	G	.	A	A	G	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.
Baf159	O_{3+4+8}	.	.	G	A	G	.	A	A	G	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf166	O_{3+4+8}	.	.	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	.	.	.	.	.	.
Baf196	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam28	O_{3+4+8}	.	A	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	.	A	.	.	.	.	.
Maf17	O_{3+4+8}	.	A	G	.	G	.	A	A	G	.	.	.	T	T	T	.	.	.	A	.	.	.	.	.	.
Maf27	O_{3+4+8}	.	A	G	.	G	.	A	A	G	.	.	.	.	T	T	.	.	.	A	.	.	.	.	.	.

Supplementary Tables S2

trus		28*	53	65	86	128	146	161	167	173*	179*	186*	190*	218*	284	294*	335	371	380	413	414*	425	428	440	485	518	521	530	585*	587	588*	593	602	675*	722	797
Bam73	O_{ST}	G	G	C	T	G	C	G	A	T	G	G	A	A	C	C	T	T	A	C	G	C	T	G	A	A	T	T	G	C	A	C	C	A	A	G
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf169	O_{ST}	.	.	T	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf37	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	A
Bam56	O_{3+4}	C	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Baf23	O_{3+4}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	T	T	G	T	.	.	.	.
Baf32	O_{3+4}	C	.	.	.	.	.	.	.	A	.	.	C	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Baf46	O_{3+4}	C	.	.	.	.	.	.	.	A	.	.	C	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Baf69	O_{3+4}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	T	T	G	T	.	.	.	.
Baf81	O_{3+4}	.	.	.	C	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Baf148	O_{3+4}	.	.	.	C	.	.	.	.	A	.	.	C	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Baf167	O_{3+4}	C	.	.	.	.	.	.	.	A	.	A	C	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Baf185	O_{3+4}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Mam16	O_{3+4}	.	.	.	C	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	T	T	G	T	.	.	.	.
Mam27	O_{3+4}	C	.	.	.	.	.	.	.	A	.	.	C	T	T	.	A	.	C	T	T	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Mam37	O_{3+4}	.	.	.	.	.	.	.	.	A	C	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	.	.	G	.	.
Maf1	O_{3+4}	C	.	.	.	.	.	.	.	A	.	.	C	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Maf5	O_{3+4}	C	.	.	.	.	.	A	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Maf20	O_{3+4}	.	A	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Maf22	O_{3+4}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	T	.	.	.
Maf34	O_{3+4}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	T	G	C	.	.	.	.	T	.	.	.	.
Maf78	O_{3+4}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	.	G	T	T	G	T	.	.	.	.
Bam6	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	T	.	.	A
Bam13	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	T	.	.	.
Bam32	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Baf14	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	T	.	.	.
Baf21	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Baf53	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	A
Baf159	O_{3+4+8}	.	.	.	.	.	G	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	T	.	.	.
Baf166	O_{3+4+8}	C	.	.	.	.	.	.	.	A	.	A	C	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Baf196	O_{3+4+8}	.	.	T	.	.	.	.	.	.	.	.	.	.	.	G	.	A	C	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
Mam28	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	.	.	.	.
Maf17	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	T	.	.	.
Maf27	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	T	T	.	A	.	C	T	.	T	C	.	.	G	C	.	.	.	.	T	T	.	.	.

Supplementary Tables S2

trus		816	826	843	851	864	875	966*	1002*	1003*	1010	1034	1040	1043	1052	1055	1064	1115	1121	1124	1159	1199	1200	1272	1279	1292	1328	1340	1366	1373	1381
Bam73	O_{ST}	A	G	G	C	G	C	C	C	A	C	C	A	A	T	C	T	C	T	G	A	T	A	A	T	A	T	C	G	G	G
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	T
Baf75	O_{ST}	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	A	T
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	T	.	A	T
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	C	.	T
Baf169	O_{ST}	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	A	T
Baf179	O_{ST}	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	C	.	T
Baf200	O_{ST}	T	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	A	T
Baf203	O_{ST}	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	C	.	T
Maf37	O_{ST}	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	C	.	T
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	T
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	G	G	A	.	.	.	.	.	.	C	.	.	.	G	.	.	C	.	.
Bam16	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	A	.	.	.	.	.	.	G	.	.	C	.	.
Bam56	O_{3+4}	.	.	.	.	.	.	.	.	.	T	.	G	G	.	T	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Baf23	O_{3+4}	.	.	.	.	T	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	T	.	G	.	.	C	.	.
Baf32	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	A	.	.	.	.	.	.	G	.	.	C	.	.
Baf46	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Baf69	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	T	.	.	.	.	G	.	.	C	.	.
Baf81	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	A	.	.	.	.	.	.	G	.	.	C	.	.
Baf148	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Baf167	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	A	.	.	.	.	.	G	.	.	C	.	.
Baf185	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Mam16	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Mam27	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Mam37	O_{3+4}	.	.	.	.	.	.	.	.	.	.	T	T	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Maf1	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	A	.	C	.	.
Maf5	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Maf20	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Maf22	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	G	.	.	.	.	.	.	.	G	.	.	C	.	.
Maf34	O_{3+4}	.	.	.	.	.	.	A	.	.	.	.	T	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Maf78	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Bam6	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	G	.	.	.	.	.	.	.	G	.	.	C	.	.
Bam13	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	G	.	.	.	.	G	.	.	G	.	.	C	.	.
Bam32	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Baf14	O_{3+4+8}	.	A	.	.	.	T	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Baf21	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	C	G	.	.	C	.	.
Baf53	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Baf159	O_{3+4+8}	.	.	.	.	.	.	.	A	C	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Baf166	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	A	.	.	.	.	.	G	.	.	C	.	.
Baf196	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.
Mam28	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	.	G	.	.	C	.	.
Maf17	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	G	.	.	.	.	.	.	.	G	.	.	C	.	.
Maf27	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	G	.	.	.	.	.	.	.	G	.	.	C	.	.

Supplementary Tables S2

Fmr1		10	11	12	42	69	101	111	117	148	194	211	234	286	289	308	311	312	318	350	356	387	389	400	403	414	438	499	552	614	619	622	623	625	636	640
Bam73	O_{ST}	T	T	T	C	C	C	T	C	G	G	G	T	G	A	T	G	G	C	C	G	C	C	G	G	G	T	G	T	G	A	C	G	T	A	A
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf36	O_{ST}	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf37	O_{ST}	G	.	G	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	.	.
Bam16	O_{3+4}	.	.	.	.	.	.	.	.	.	A	.	.	A	.	.	.	.	T	T	.	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Bam56	O_{3+4}	.	.	.	.	.	.	.	.	.	A	.	G	A	.	A	.	.	.	T	A	.	.	T	.	.	G	.	.	A	.	A	.	A	C	.
Baf23	O_{3+4}	.	.	.	.	.	.	.	.	.	A	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf32	O_{3+4}	.	A	.	.	.	A	C	.	.	.	.	.	A	.	.	.	.	T	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf46	O_{3+4}	.	A	.	.	.	.	.	.	.	A	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf69	O_{3+4}	.	.	.	.	T	A	.	.	.	.	.	.	A	.	.	.	.	.	T	A	.	G	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf81	O_{3+4}	.	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	T	T	.	.	.	.	A	.	A	.	A	C	.
Baf148	O_{3+4}	.	.	.	.	.	.	.	.	C	A	.	.	A	.	.	.	.	T	T	.	.	.	T	.	.	.	.	.	A	.	A	.	A	C	C
Baf167	O_{3+4}	.	.	.	.	.	.	.	.	C	A	.	.	A	.	.	.	.	.	T	A	T	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf185	O_{3+4}	.	.	.	.	.	A	C	.	.	.	.	.	A	.	.	.	.	T	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Mam16	O_{3+4}	.	.	.	.	.	A	.	.	.	A	.	.	A	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Mam27	O_{3+4}	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Maf1	O_{3+4}	.	.	.	.	.	.	.	.	.	.	C	.	A	.	.	A	T	.	T	A	.	.	T	.	.	.	T	.	A	.	A	.	A	C	.
Maf5	O_{3+4}	.	.	.	.	.	.	.	.	.	.	C	.	A	.	.	A	T	.	T	A	.	.	T	.	.	.	T	.	A	.	A	.	A	C	.
Maf11	O_{3+4}	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Maf20	O_{3+4}	.	.	.	.	.	A	.	.	.	.	.	.	A	.	.	.	.	.	T	A	.	.	T	.	A	.	.	.	A	.	A	.	A	C	.
Maf22	O_{3+4}	.	.	.	.	.	.	.	.	C	A	.	.	A	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Maf34	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Maf78	O_{3+4}	.	.	.	.	.	.	.	.	C	A	.	.	A	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Bam6	O_{3+4+8}	.	A	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Bam13	O_{3+4+8}	.	.	.	.	.	.	.	.	A	A	.	.	A	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	G	A	T	A	C	.
Bam32	O_{3+4+8}	.	.	.	.	.	A	.	.	.	A	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf14	O_{3+4+8}	.	A	.	.	.	.	.	T	.	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	G	A	.	A	C	.
Baf20	O_{3+4+8}	.	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf53	O_{3+4+8}	.	.	.	.	.	.	.	.	C	A	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf159	O_{3+4+8}	.	.	.	.	.	A	.	.	.	A	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf166	O_{3+4+8}	.	.	.	.	.	A	.	.	.	A	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Baf196	O_{3+4+8}	.	.	.	.	.	A	.	.	.	A	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Mam28	O_{3+4+8}	.	.	.	.	.	A	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.
Maf17	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A	.	A	.	A	C	.

Supplementary Tables S2

Fmr1		699	733	752	767	772	786	787	795	809	810	814	815	816	817	821	823	826	827	828	829	830	831	832	843	849	851	852	856	857	858	863	867	871	874	875
Bam73	O_{ST}	G	G	A	A	C	G	G	C	C	A	C	G	A	A	T	G	T	T	G	T	G	A	T	T	T	A	A	A	C	T	C	T	A	C	G
Baf58	O_{ST}	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf36	O_{ST}	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf37	O_{ST}	.	.	.	.	.	.	.	T	.	G	G	A	G	C	A	.	A	A	A	A	.	T	.	G	G	G	.	T	T	A	A	A	T	.	T
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Bam56	O_{3+4}	T	.	.	.	T	.	.	G	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	T	T
Baf23	O_{3+4}	C	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf32	O_{3+4}	C	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf46	O_{3+4}	T	.	.	.	.	.	.	G	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf69	O_{3+4}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	T	T
Baf81	O_{3+4}	T	.	.	.	.	C	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf148	O_{3+4}	T	.	.	.	.	.	.	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf167	O_{3+4}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	T	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf185	O_{3+4}	C	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Mam16	O_{3+4}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Mam27	O_{3+4}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	.	T
Maf1	O_{3+4}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	T	T
Maf5	O_{3+4}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	T	T
Maf11	O_{3+4}	.	.	.	.	.	.	.	G	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	T	T
Maf20	O_{3+4}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Maf22	O_{3+4}	T	.	.	.	.	.	.	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Maf34	O_{3+4}	T	.	.	.	.	.	.	G	T	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	T	T
Maf78	O_{3+4}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Bam6	O_{3+4+8}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	T	T
Bam13	O_{3+4+8}	T	.	T	.	.	.	.	.	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	G	G	T	T	A	.	G	G	.	T
Bam32	O_{3+4+8}	T	.	T	.	.	.	.	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf14	O_{3+4+8}	T	.	.	.	.	.	A	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf20	O_{3+4+8}	T	A	.	.	.	A	.	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf53	O_{3+4+8}	T	.	.	.	T	.	.	G	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf159	O_{3+4+8}	T	.	T	.	.	.	.	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf166	O_{3+4+8}	.	.	.	.	.	.	.	G	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Baf196	O_{3+4+8}	T	.	T	.	.	.	.	.	T	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T
Mam28	O_{3+4+8}	T	.	.	.	.	.	.	G	.	G	G	A	G	C	.	.	A	A	A	G	.	T	G	G	G	.	.	T	T	A	.	G	G	T	T
Maf17	O_{3+4+8}	T	.	.	.	.	.	.	.	.	G	G	A	G	C	.	.	A	A	A	G	.	.	G	G	G	G	G	T	T	A	.	G	G	.	T

Supplementary Tables S2

Fmr1		876	880	883	885	886	887	888	890	892	894	895	896	898	900	901	905	909	910	922	934	939	941	942	946	951	954	955	957	958	966	975	981	983	990	991
Bam73	O_{ST}	G	C	A	G	T	A	G	C	A	G	G	G	A	G	C	T	A	T	A	C	A	T	G	T	C	G	T	T	C	C	G	C	T	T	G
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	T	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf36	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Maf37	O_{ST}	C	T	.	A	G	T	T	A	T	C	C	A	G	A	A	C	G	C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Bam56	O_{3+4}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	T
Baf23	O_{3+4}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Baf32	O_{3+4}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Baf46	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	A	T	G	G	.	.	.	.	.	.	.
Baf69	O_{3+4}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Baf81	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Baf148	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Baf167	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Baf185	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Mam16	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	T	.	.	.	.
Mam27	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Maf1	O_{3+4}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	C	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Maf5	O_{3+4}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	C	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Maf11	O_{3+4}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Maf20	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	C	.	.
Maf22	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	A	.
Maf34	O_{3+4}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Maf78	O_{3+4}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Bam6	O_{3+4+8}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	C	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Bam13	O_{3+4+8}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Bam32	O_{3+4+8}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	T	.	.	.	.	.
Baf14	O_{3+4+8}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Baf20	O_{3+4+8}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	C	A	T	G	C	.	.	T	G	G	.	T	.	.	.	.	.
Baf53	O_{3+4+8}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Baf159	O_{3+4+8}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	T	.	.	.	.	.
Baf166	O_{3+4+8}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	C	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Baf196	O_{3+4+8}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	T	.	.	.	.	.
Mam28	O_{3+4+8}	C	T	.	A	.	.	T	.	T	C	.	A	G	A	A	C	G	C	C	A	T	G	C	.	.	T	G	G	.	.	.	.	.	.	.
Maf17	O_{3+4+8}	C	T	.	A	G	.	T	.	T	C	.	A	G	A	A	C	G	C	.	A	T	G	C	.	.	T	G	G	.	T	.	.	.	.	.

Supplementary Tables S2

Fmr1		1031	1068	1106	1126	1133	1134	1171	1205	1212	1225	1267	1351	1359	1362	1370	1402	1409	1410	1429	1453	1459	1474	1479*	1510	1522	1546	1573	1651	1705	1762	1782	1787	1791	1793
Bam73	O_{ST}	A	G	T	C	G	T	A	G	G	T	T	C	C	A	T	C	C	A	T	G	G	G	T	G	T	G	G	G	C	C	C	A	C	A
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Baf75	O_{ST}	.	.	.	T	.	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C
Baf145	O_{ST}	.	.	.	.	T	.	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C
Baf147	O_{ST}	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	T	.
Baf179	O_{ST}	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C
Baf200	O_{ST}	.	.	.	.	.	.	.	T	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf203	O_{ST}	.	.	.	.	T	.	.	.	.	.	G	.	G	.	.	A	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.
Maf36	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.
Maf37	O_{ST}	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	T	.
Maf81	O_{ST}	.	.	.	T	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	T	.	.	.	.	T	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	T	C
Bam16	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Bam56	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	C	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Baf23	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.
Baf32	O_{3+4}	.	.	.	T	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	T	.	.	C	.	A	.	.	.	.	C	.	.
Baf46	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Baf69	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	C	.	.	.	C	.	A	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Baf81	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	.	C	A	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Baf148	O_{3+4}	G	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	A	.	.	.	C	.	.	.	.	.	T	.	.	.
Baf167	O_{3+4}	.	.	.	G	.	.	.	.	.	G	.	G	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Baf185	O_{3+4}	.	.	.	T	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.
Mam16	O_{3+4}	.	T	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
Mam27	O_{3+4}	.	.	.	T	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	A	.	.	.	A	.	A	.	.	.	.	.	.	.	.
Maf1	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	A	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Maf5	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.
Maf11	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Maf20	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Maf22	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Maf34	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	A	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Maf78	O_{3+4}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	.	C	.	.	.	C	A	.	.	.	.	.	.	.	.	.	.
Bam6	O_{3+4+8}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.
Bam13	O_{3+4+8}	.	.	.	G	.	C	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Bam32	O_{3+4+8}	.	.	.	T	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.
Baf14	O_{3+4+8}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	A	.	A	.	.	.	.	.	C	.	.
Baf53	O_{3+4+8}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	A	.	.	.	.	.	.	.
Baf159	O_{3+4+8}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Baf166	O_{3+4+8}	.	.	.	T	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.
Baf196	O_{3+4+8}	.	.	.	T	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	A	.	.	.	.	.	.	.	.
Mam28	O_{3+4+8}	.	.	.	G	.	.	T	.	.	G	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.
Maf17	O_{3+4+8}	.	.	.	T	.	.	T	.	.	G	.	.	G	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.

Supplementary Tables S2

Fmr1		1805	1822	1826	1836	1881	1920	1944	2001	2002	2003	2005	2006	2026	2032	2033	2034	2039	2046	2048	2089
Bam73	O_{ST}	T	A	T	A	C	A	G	C	T	T	C	A	G	C	A	A	A	T	C	A
Baf58	O_{ST}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	G	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf200	O_{ST}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
Baf203	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf36	O_{ST}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf37	O_{ST}	.	.	G	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	G	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf85	O_{ST}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	.	G	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	T
Bam56	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	T
Baf23	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	T
Baf32	O_{3+4}	.	.	.	G	.	.	.	T	C	C	.	T	A	.	.	.	.	.	A	.
Baf46	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Baf69	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Baf81	O_{3+4}	.	.	.	G	T	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Baf148	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Baf167	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Baf185	O_{3+4}	.	.	.	G	.	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.
Mam16	O_{3+4}	.	G	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Mam27	O_{3+4}	.	G	.	G	.	.	.	T	C	C	.	T	A	.	.	.	.	.	A	.
Maf1	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Maf5	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Maf11	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Maf20	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	G	T	.	.	.
Maf22	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Maf34	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Maf78	O_{3+4}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Bam6	O_{3+4+8}	.	.	.	G	.	.	.	.	.	.	A	.	.	T	T	G	T	.	.	.
Bam13	O_{3+4+8}	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
Bam32	O_{3+4+8}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
Baf14	O_{3+4+8}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	G	T	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.
Baf53	O_{3+4+8}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Baf159	O_{3+4+8}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Baf166	O_{3+4+8}	.	.	.	G	.	.	A	.	.	.	.	.	.	.	T	.	T	.	.	.
Baf196	O_{3+4+8}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Mam28	O_{3+4+8}	.	.	.	G	T	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.
Maf17	O_{3+4+8}	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.

Supplementary Tables S2

reg_*Fmr1*		58	99	107	140	141	146	152	160	161	169	171	174	177	183	189	191	193	204	209	277	289	308	322	323	324	325	340	357	374	386	407	425
Bam73	O_{ST}	C	G	G	A	C	G	G	T	G	G	T	G	A	G	T	C	T	G	G	A	G	G	G	T	C	G	C	G	A	A	A	T
Baf58	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf145	O_{ST}	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Baf200	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Baf203	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Mam7	O_{ST}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	A	G	.	.	.	C	.	.	.	.	.	.	.	.
Maf16	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Maf37	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Maf85	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Mf85.2	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	T	.	.	.	G	.	.	.	.	.	.	A	.	.	G	T	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.
Bam56	O_{3+4}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	G	.	.	.
Baf23	O_{3+4}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.
Baf32	O_{3+4}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Baf46	O_{3+4}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	A
Baf69	O_{3+4}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.
Baf81	O_{3+4}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.
Baf148	O_{3+4}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Baf167	O_{3+4}	.	.	T	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Baf185	O_{3+4}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Mam16	O_{3+4}	T	.	.	.	G	.	.	.	T	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Mam27	O_{3+4}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Mam37	O_{3+4}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Maf1	O_{3+4}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	C	.	.	.	.	.	.	.	.
Maf5	O_{3+4}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	C	.	.	.	.	.	.	.	.
Maf11	O_{3+4}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Maf20	O_{3+4}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	.	A	.	C	.	.	.	.	.	.	.	.
Maf22	O_{3+4}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Maf34	O_{3+4}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	T	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Maf78	O_{3+4}	T	.	.	.	G	T	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.
Bam6	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	.	.	C	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Bam13	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.
Bam32	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	G	.	.
Baf14	O_{3+4+8}	T	.	.	.	G	.	.	A	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.	C	.	T	.	.	.	.	.	.
Baf21	O_{3+4+8}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	T	.	.	.	.	.
Baf53	O_{3+4+8}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Baf159	O_{3+4+8}	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	T	.	.	.	.	.
Baf166	O_{3+4+8}	.	A	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Baf196	O_{3+4+8}	.	.	.	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam12	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.
Mam25	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Mam28	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.
Mam31	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.
Maf13	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	T	.
Maf17	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	A	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Mf56	O_{3+4+8}	T	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C	G	.	T	.	.	.	.	.

Supplementary Tables S2

reg_*Fmr1*		451	465	470	493	497	502	522	527	528	539	545	559	583	589	594	604	608	609	612	634	641	651	676	683	687	693	708	711	720	761	763	766
Bam73	O_{ST}	G	A	G	T	T	A	C	C	T	A	C	T	G	G	A	C	T	C	C	A	C	C	G	C	C	G	G	G	A	A	T	C
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Baf145	O_{ST}	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Baf203	O_{ST}	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam7	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf16	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Maf37	O_{ST}	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Maf85	O_{ST}	.	.	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mf85.2	O_{ST}	.	.	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam16	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bam56	O_{3+4}	.	.	.	.	.	.	.	T	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.
Baf23	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Baf32	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	T	.	.	.	.	.	.	T	.	.	A	.	C	.	.	.	.	.
Baf46	O_{3+4}	.	.	.	A	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Baf69	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Baf81	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Baf148	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Baf167	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Baf185	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	T	.	.	.	.	.	.	T	.	.	A	.	C	.	.	T	.	.
Mam16	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Mam27	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Mam37	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.
Maf1	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	A	.
Maf5	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	A	.
Maf11	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	A	.
Maf20	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	C	.	.	.	.	.
Maf22	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Maf34	O_{3+4}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Maf78	O_{3+4}	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	C	T	A	.
Bam6	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.
Bam13	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Bam32	O_{3+4+8}	T	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf14	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf20	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	.	.	A	.	.	.	C	T	.	T	.	.	.	.	.	.	.	.	.	.	.	.
Baf21	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	C	.	.	.	.	.
Baf53	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.
Baf159	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	A	A	.	C	.	.	.	.	.
Baf166	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Baf196	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam12	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	A	.
Mam25	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mam28	O_{3+4+8}	.	.	.	.	A	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Mam31	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	A	.
Maf13	O_{3+4+8}	.	.	A	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
Maf17	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mf56	O_{3+4+8}	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.

Supplementary Tables S2

reg_*Fmr1*		769	772	773	775	786	787	788	806	811	818	826	832	836	848	858	861
Bam73	O_{ST}	T	C	T	A	C	C	C	A	G	A	A	C	C	T	G	C
Baf58	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A
Baf75	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Baf145	O_{ST}	.	A	.	G	.	.	.	.	.	.	.	.	.	A	.	.
Baf147	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Baf169	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Baf179	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Baf200	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Baf203	O_{ST}	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Mam7	O_{ST}	C	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Maf16	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Maf37	O_{ST}	.	A	.	G	.	.	.	.	.	.	.	.	.	A	.	.
Maf81	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Maf85	O_{ST}	.	A	.	.	A	.	.	.	.	.	.	.	.	A	.	.
Mf85.2	O_{ST}	.	A	.	.	A	T	.	.	.	.	.	.	.	A	.	.
Maf116	O_{ST}	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Bam16	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Bam56	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf23	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf32	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf46	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf69	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf81	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf148	O_{3+4}	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf167	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf185	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Mam16	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Mam27	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Mam37	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	T	.
Maf1	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Maf5	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Maf11	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Maf20	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Maf22	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Maf34	O_{3+4}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Maf78	O_{3+4}	.	.	G	.	.	.	A	.	A	.	.	.	A	A	.	.
Bam6	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Bam13	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Bam32	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf14	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf20	O_{3+4+8}	.	.	.	.	.	.	.	.	A	.	G	.	.	A	.	.
Baf21	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf53	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf159	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf166	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Baf196	O_{3+4+8}	.	A	.	.	.	.	.	.	.	.	.	A	.	A	.	.
Mam12	O_{3+4+8}	.	.	G	.	.	A	.	.	A	.	.	.	.	A	.	.
Mam25	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Mam28	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Mam31	O_{3+4+8}	.	.	G	.	.	A	.	T	A	.	.	.	.	A	.	.
Maf13	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Maf17	O_{3+4+8}	.	.	G	.	.	.	.	.	A	.	.	.	.	A	.	.
Mf56	O_{3+4+8}	.	.	G	.	.	.	.	.	A	G	.	.	.	A	.	.

Supplementary Table S3: Genetic differentiation between chromosomal arrangements in coding (exons) and noncoding (introns and regulatory regions) regions of the genes.

		Acph-1		*Ast*		*larp*			CG5961	*trus*		*Fmr1*			Concatenated data[a]	
		exons	introns	exons	introns	exons	introns	reg.	exons	exons	introns	exons	introns	reg.	exons	noncod.
	Size, bp	1080	785	363	1477	1779	133	1555	598	1062	318	582	1390	850	5454	6601
O_{ST} vs. O_{3+4}	D_{XY}	0.022	0.02	0.019	0.014	0.004	0.019	0.024	0.015	0.015	0.009	0.004	0.046	0.013	0.012	0.025
	F_{ST}	0.388	0.449	0.669	0.423	0.365	0.130	0.494	0.818	0.763	0.334	0.112	0.787	0.435	0.611	0.650
	P of Snn	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000 ***	0.000***	0.000***	0.008**	0.000***	0.000***	0.000***	0.000***
O_{ST} vs. O_{3+4+8}	D_{XY}	0.023	0.019	0.02	0.014	0.004	0.019	0.028	0.014	0.014	0.009	0.004	0.044	0.013	0.012	0.025
	F_{ST}	0.464	0.317	0.67	0.487	0.387	0.123	0.433	0.623	0.72	0.291	0.083	0.787	0.382	0.613	0.612
	P of Snn	0.000***	0.000***	0.000***	0.000***	0.000***	0.002**	0.000***	0.000***	0.000***	0.000***	0.124 ns	0.000***	0.000***	0.000***	0.000***
O_{3+4} vs. O_{3+4+8}	D_{XY}	0.015	0.014	0.008	0.011	0.002	0.017	0.025	0.005	0.005	0.002	0.004	0.009	0.008	0.006	0.014
	F_{ST}	0.136	0.175	0.071	0.159	0.049	0.123	0.520	0.041	0.082	-0.013	0.009	0.06	0.021	0.112	0.31
	P of Snn	0.001**	0.000***	0.718 ns	0.018 *	0.002**	0.008**	0.000***	0.022*	0.137 ns	0.617 ns	0.388 ns	0.018*	0.001**	0.004**	0.000***

D_{XY}, average number of nucleotide differences per site between arrangements; F_{ST}, proportion of nucleotide diversity attributable to variation among arrangements. P-value of Snn after permutation test (probability obtained by the permutation test with 1000 replicates): ns, not significant; *, 0.01<P<0.05; **, 0.001<P<0.01; ***, P<0.001.

[a]Analysis performed in the concatenated dataset that includes all eight regions and only 25 sequences in total (see materials and methods).

Reg., regulatory regions; noncod., noncoding regions in concatenated dataset, that includes introns of all genes and regulatory regions of *larp* and *Fmr1*.

Supplementary table S4: Significant associations between informative polymorphic nucleotide sites and *ZnS* statistic values in populations of Barcelona, Málaga and pooling them together.

Population	Inversion	Parameter	*Acph-1*	*Ast*	*larp*	reg_*larp*	CG5961	*trus*	*Fmr-1*	reg_*Fmr-1*	Concat.
Barcelona	O_{ST}	%LD[a]	10.591	6.410	7.143	5.426	0	20	3.333	0	3.797
		%LD[b]	0	0	0	0	0	0	0	0	0
		ZnS	0.275	0.169	0.162	0.217	N/A	0.208	0.150	0.179	0.182
	O_{3+4}	%LD[a]	5.974	3.896	0	3.268	0	14.286	12.554	16.364	2.691
		%LD[b]	0	0	0	0	0	0	0	0	0
		ZnS	0.175	0.206	0.141	0.153	0.357	0.242	0.267	0.321	0.184
	O_{3+4+8}	%LD[a]	4.183	20	13.333	44.934	25	0	8.095	8.333	0
		%LD[b]	0	0	0	0	0	0	0	0	0
		ZnS	0.149	0.271	0.322	0.547	0.332	0.163	0.217	0.247	0.295
	O_{ST}/O_{3+4}	%LD[a]	15.614	14.512	11.538	39.675	58.333	44.444	45.977	19.048	29.790
		%LD[b]	0.724	0	2.564	8.961	41.667	25.071	0	1.429	0
		ZnS	0.157	0.194	0.143	0.317	0.526	0.384	0.460	0.177	0.341
	O_{ST}/O_{3+4+8}	%LD[a]	17.545	30.769	14.103	27.251	52.778	42.461	49.045	29.412	29.345
		%LD[b]	0.483	8.974	2.564	0.482	30.556	12	0	5.882	0
		ZnS	0.158	0.293	0.135	0.240	0.442	0.380	0.497	0.221	0.322
	O_{3+4}/O_{3+4+8}	%LD[a]	6.332	11.492	19.048	35.382	19.444	9.091	7.386	13.636	6.951
		%LD[b]	0.181	1.210	4.762	0	0	0	0	3.030	0
		ZnS	0.094	0.143	0.179	0.361	0.265	0.122	0.103	0.147	0.143
Málaga	O_{ST}	%LD[a]	8.095	0	0	0	0	0	0	14.256	0
		%LD[b]	0	0	0	0	0	0	0	0	0
		ZnS	0.269	1.000	0.571	N/A	N/A	0.467	0.272	0.327	N/A
	O_{3+4}	%LD[a]	3.427	15.810	28.581	0	0	20	10.952	28.571	0
		%LD[b]	0	0	0	0	0	0	0	0	0
		ZnS	0.157	0.263	0.475	0.37	0.000	0.314	0.219	0.372	N/A
	O_{3+4+8}	%LD[a]	0	0	0	0	0	0	0	33.333	-
		%LD[b]	0	0	0	0	0	0	0	0	-
		ZnS	0.267	N/A	0.735	N/A	N/A	N/A	N/A	0.480	-
	O_{ST}/O_{3+4}	%LD[a]	18.744	16.785	27.619	19.359	58.333	30	47.387	20.261	0
		%LD[b]	0	0	2.857	0	58.333	0	0	1.307	0
		ZnS	0.190	0.220	0.281	0.426	0.685	0.378	0.443	0.187	0.780
	O_{ST}/O_{3+4+8}	%LD[a]	9.012	31.954	13.333	0	62.222	28.655	31.328	18.333	-
		%LD[b]	0	0	0	0	0	0	0	0	-
		ZnS	0.191	0.612	0.285	0.926	0.814	0.522	0.554	0.253	-
	O_{3+4}/O_{3+4+8}	%LD[a]	4.123	12.834	14.545	50.196	16.667	17.857	4.710	21.818	-
		%LD[b]	0	0	0	0	0	0	0	3.636	-
		ZnS	0.099	0.196	0.219	0.602	0.160	0.216	0.151	0.196	-
All populations	O_{ST}	%LD[a]	2.128	5.714	10.909	6.650	0	12.727	0.004	7.619	2.398
		%LD[b]	0	0	1.818	0	N/A	0	0	0	0
		ZnS	0.108	0.13	0.189	0.163	N/A	0.164	0.089	0.136	0.136
	O_{3+4}	%LD[a]	7.547	8.266	24.444	3.667	0	14.286	11.561	16.190	1.419
		%LD[b]	0	0	2.222	0.333	0	10.714	0	1.905	0
		ZnS	0.105	0.127	0.262	0.097	0.031	0.174	0.135	0.147	0.120
	O_{3+4+8}	%LD[a]	7.112	18.462	10.714	44.934	12.727	0	3.953	15.151	0
		%LD[b]	0	0	7.143	0	0	0	0	0	0
		ZnS	0.124	0.217	0.239	0.547	0.227	0.175	0.188	0.208	0.295
	O_{ST}/O_{3+4}	%LD[a]	19.074	18.047	21.637	31.429	60	47.293	40.159	13.333	31.015
		%LD[b]	2.314	2.626	2.924	6.364	46.667	25.926	23.936	2.5	0
		ZnS	0.097	0.117	0.129	0.185	0.435	0.302	0.309	0.123	0.270
	O_{ST}/O_{3+4+8}	%LD[a]	16.404	26.105	15.686	28.696	46.97	47.863	47.252	20	29.345
		%LD[b]	1.19	8.673	3.268	5.119	22.727	21.652	27.505	3.077	0
		ZnS	0.099	0.210	0.125	0.245	0.331	0.320	0.386	0.113	0.322
	O_{3+4}/O_{3+4+8}	%LD[a]	9.895	8.708	16.483	31.329	16.364	8.791	8.985	15.833	6.951
		%LD[b]	0.421	0.508	3.297	16.457	0	5.494	0.317	3.333	0
		ZnS	0.063	0.084	0.138	0.269	0.191	0.092	0.073	0.088	0.143

[a] percentage of pairwise comparisons with significant LD by the Fisher test (P=0.005)

[b] percentage of pairwise comparisons with significant LD after Bonferroni correction (WEIR 1996)

*Zn*S, overall pairwise comparisons using only informative sites

Concat., concatenated dataset

Supplementary Table S5a: Neutrality tests and test of population expansion for eight regions in Barcelona population using Ramos-Onsins and Rozas' R_2. Significant values are in bold. The significance of R_2 was calculated by coalescent simulations with estimated levels (ρ) of recombination.

Barcelona		*N*	Tajima's *D*						Fu and Li's *D*			R_2	*Rho* (ρ)	*P*
			Whole sequence	Exons	Introns	Synonymous	Nonsynonymous	Silent	Whole sequence	Introns	Exons			
Acph-1	O_{ST}	8	-0,333	-0,047	-0,496	0,252	-1,175	-0,221	0,056	0,375	-0,143	0.117	22.471	0.078
	O_{3+4}	8	0,119	0,244	-0,454	0,442	-1,030	0,235	0,712	0,504	0,737	0.123	91.659	0.150
	O_{3+4+8}	8	-0,640	-0,655	-0,843	-0,615	-0,812	-0,620	-0,411	-0,819	0,089	0.128	>100	0.217
Ast	O_{ST}	8	-0,307	0,015	-0,616	0,336	-1,055	-0,251	-0,575	-0,445	-0,879	**0.104**	65.141	**0.026**
	O_{3+4}	6	0,127	0,974	0,153	0,974	n.a.	0,127	-0,015	-0,136	0,756	0.128	>100	0.178
	O_{3+4+8}	8	-0,784	-0,222	-0,916	-0,222	n.a.	-0,784	-0,676	-0,683	-0,397	0.142	2.812	0.291
Iarp	O_{ST}	7	-0,855	-1,278	0,753	-0,931	-1,486	-0,4753	-1,057	0,636	-1,541	**0.095**	>100	**0.012**
	O_{3+4}	11	-1,553	-1,726	-0,640	-1,548	-1,650	-1,336	**-2,245***	1,128	**-3,173****	**0.088**	4.481	**0.003**
	O_{3+4+8}	10	-0,470	-1,116	0,477	-0,943	-1,034	-0,188	-0,215	0,592	-0,527	0.138	5.160	0.194
reg_Iarp	O_{ST}	7	-0,204	-	-0,204	-	-	-0,204	-0,691	-0,691	-	0.140	>100	0.401
	O_{3+4}	9	-0,850	-	-0,850	-	-	-0,850	-0,850	-0,850	-	0.099	62.588	0.010
	O_{3+4+8}	8	-0,773	-	-0,773	-	-	-0,773	0,015	0,015	-	0.130	>100	0.232
trus	O_{ST}	9	-0,871	-1,474	-0,270	-1,422	-1,088	-0,812	-0,443	-0,960	-0,144	**0.088**	29.289	**0.002**
	O_{3+4}	10	-0,649	-0,153	-1,667	-0,669	0,477	-1,200	-0,701	-1,021	-0,410	0.114	12.798	0.055
	O_{3+4+8}	7	-0,733	-0,345	-1,535	0,132	-1,237	-0,503	-0,645	-1,677	-0,350	0.166	0.015	0.425
CG5961	O_{ST}	9	-1,412	-1,412	-	-1,486	-1,468	-1,486	**-2,519****	**-**	**-2,519****	0.151	>100	0.294
	O_{3+4}	9	-0,526	-0,526	-	-0,229	-1,088	-0,229	-0,951	-	-0,951	0.133	>100	0.088
	O_{3+4+8}	7	-0,905	-0,905	-	-0,690	-1,237	-0,690	-0,615	-	-0,615	0.125	15.068	0.095
Fmr1	O_{ST}	8	-0,256	-0,555	-0,255	-0,525	n.a.	-0,256	0,261	-0,125	-1,380	0.133	>100	0.285
	O_{3+4}	10	-1,024	-1,106	-0,877	-1,106	n.a.	-1,024	-1,075	-0,659	-1,738	**0.088**	28.755	**0.002**
	O_{3+4+8}	9	-0,206	-0,703	-0,130	-0,703	n.a.	-0,206	-0,300	-0,166	-0,566	0.133	33.538	0.256
reg_Fmr1	O_{ST}	9	-0,797	-	-0,797	-	-	-0,797	-0,507	-0,507	-	**0.110**	18.924	**0.034**
	O_{3+4}	10	-0,269	-	-0,269	-	-	-0,269	-0,621	-0,621	-	0.132	5.372	0.185
	O_{3+4+8}	8	-0,862	-	-0,862	-	-	-0,862	-1,432	-1,432	-	**0.092**	2.487	**0.007**

Supplementary Table S5b: Neutrality tests and test of population expansion for eight regions in Málaga population using Ramos-Onsins and Rozas' R_2. Significant values are in bold. The significance of R_2 was calculated by coalescent simulations with estimated levels (ρ) of recombination.

Málaga		*N*	Tajima's *D*						Fu and Li's *D*			R_2	*Rho* (ρ)	*P*
			Whole sequence	Exons	Introns	Synonymous	Nonsynonymous	Silent		Introns	Exons			
Acph-1	O_{ST}	5	0,243	0,150	0,286	0,138	0,243	0,239	0,276	0,536	0,131	0.165	20.034	0.583
	O_{3+4}	9	-0,551	-0,410	-1,328	-0,221	-1,398	-0,444	-0,178	0,168	-0,348	0.126	85.550	0.173
	O_{3+4+8}	4	-0,486	-0,510	-0,843	-0,447	-0,780	-0,452	-0,191	0,378	-0,351	**0.096**	94.496	**0.040**
Ast	O_{3+4}	8	-0,545	0,015	-0,526	0,336	-1,055	-0,518	-0,362	-0,594	0,854	0.157	0.198	0.099
	O_{3+4+8}	3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0,594	0,469	0,642	0.105	16.372	0.020
Iarp	O_{ST}	6	-0,183	-0,692	0,878	-0,060	-0,06042	-0,246	-0,620	0,845	-1,470	0.133	1.278	0.051
	O_{3+4}	9	-1,084	-1,710	0,358	-1,630	-1,59105	-0,708	-1,677	0,138	**-2,343****	**0.101**	1.179	**0.007**
	O_{3+4+8}	5	-0,097	-0,609	1,641	-0,668	-0,52640	0,298	-0,332	1,579	-0,896	0.162	0.618	0.133
reg_Iarp	O_{ST}	2	n.a.	-	n.a.	-	-	n.a.	-0,850	-0,850	-	0.220	n.a.	n.a.
	O_{3+4}	4	-0,071	-	-0,071	-	-	-0,071	0,322	0,322	-	0.169	31.047	0.501
trus	O_{ST}	4	1,198	0,895	1,633	0,895	n.a.	1,198	1,160	1,095	0,992	0.227	>100	0.932
	O_{3+4}	8	-0,991	-0,942	-1,055	-1,283	-0,4137	-1,336	-1,251	-1,262	-1,135	**0.074**	0.969	**0.000**
	O_{3+4+8}	3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**1,932***	n.a.	**1,932***	0.471	2.335	n.a.
CG5961	O_{3+4}	8	0,585	0,585	-	0,585	n.a.	0,585	1,219	-	1,219	0.167	47.748	0.346
	O_{3+4+8}	3	n.a.	n.a.	-	n.a.	n.a.	n.a.	1,146	-	1,146	0.272	n.a.	n.a.
Fmr1	O_{ST}	4	-0,228	-0,754	-0,071	-0,754	n.a.	-0,228	-0,624	-0,512	-1,380	0.266	0.059	0.815
	O_{3+4}	9	-0,479	-0,170	-0,714	0,062	-1,08823	-0,445	-0,056	-0,013	-0,252	0.123	18.384	0.129
reg_Fmr1	O_{ST}	7	-0,580	-	-0,580	-	-	-0,580	-0,709	-0,709	-	0.125	3.830	0.054
	O_{3+4}	10	-0,578	-	-0,578	-	-	-0,578	-1,036	-1,036	-	0.122	6.494	0.095
	O_{3+4+8}	7	-0,275	-	-0,275	-	-	-0,275	-0,595	-0,595	-	**0.128**	0.583	**0.048**

Supplementary Table S6a: Standard MKT for coding regions in Barcelona and Málaga populations.

Population	Gene	Arrangement	*N*	Polymorphism		Divergence		*NI*	*α*	X^2	*P*
				P_s	P_n	D_s	D_n				
Barcelona	*Acph1*	O_{ST}	8	26	7	140.89	53.87	0.704	0.295	0.599	0.438
		O_{3+4}	8	33	5	130.45	52.79	0.374	**0.625**	3.995	**0.045**
		O_{3+4+8}	9	39	4	130.44	53.87	0.248	**0.751**	7.299	**0.006**
	Ast	O_{ST}	9	5	1	34.01	5.06	1.343	-0.343	0.061	0.804
		O_{3+4}	7	5	0	29.31	3.02	0.000	1.000	0.508	0.475
		O_{3+4+8}	8	4	0	29.31	3.02	0.000	1.000	0.407	0.523
	larn	O_{ST}	9	10	6	113.80	86.72	0.787	0.212	0.199	0.654
		O_{3+4}	11	7	7	118.09	86.72	1.361	-0.361	0.313	0.575
		O_{3+4+8}	10	3	3	1285.77	975.37	1.318	-0.318	0.114	0.734
	CG5961	O_{ST}	9	6	2	65.07	11.18	1.939	-0.939	0.585	0.444
		O_{3+4}	10	4	1	74.67	11.18	1.669	-0.669	0.197	0.656
		O_{3+4+8}	7	7	2	70.75	11.18	1.807	-0.807	0.480	0.488
	trus	O_{ST}	9	10	1	102.26	22.40	0.456	0.543	0.558	0.454
		O_{3+4}	10	5	4	102.22	20.33	4.021	**-3.021**	4.314	**0.037**
		O_{3+4+8}	7	5	2	102.25	20.33	2.011	-1.011	0.666	0.414
	Fmr1	O_{ST}	8	4	0	19.80	2.00	0.000	1.000	0.398	0.527
		O_{3+4}	10	9	0	17.40	2.00	0.000	1.000	1.000	0.317
		O_{3+4+8}	9	9	0	18.60	2.00	0.000	1.000	0.939	0.332
			N	Polymorphism		Divergence		*NI*	*α*	X^2	*P*
				P_s	P_n	D_s	D_n				
Málaga	*Acph1*	O_{ST}	5	30	5	133.89	53.87	0.414	0.585	3.148	0.075
		O_{3+4}	9	42	7	125.40	52.79	0.395	**0.604**	4.664	**0.030**
		O_{3+4+8}	4	28	5	132.14	52.79	0.446	0.553	2.578	0.108
	Ast	O_{ST}	4	1	0	34.00	5.06	0.000	1.000	0.148	0.700
		O_{3+4}	9	8	3	27.81	3.02	3.451	-2.451	2.008	0.156
		O_{3+4+8}	3	2	2	30.84	3.02	10.203	**-9.203**	5.246	**0.021**
	larn	O_{ST}	6	4	4	116.65	85.63	1.362	-0.362	0.184	0.667
		O_{3+4}	8	5	1	29.31	3.02	1.939	-0.939	0.288	0.591
		O_{3+4+8}	5	6	9	116.64	85.63	2.043	-1.043	1.772	0.183
	CG5961	O_{ST}	4	2	0	65.07	11.18	0.000	1.000	0.342	0.558
		O_{3+4}	8	3	0	72.69	11.18	0.000	1.000	0.459	0.498
		O_{3+4+8}	3	3	0	76.75	11.18	0.000	1.000	0.435	0.509
	trus	O_{ST}	4	7	0	104.03	22.40	0.000	1.000	1.490	0.222
		O_{3+4}	8	9	7	98.73	20.33	3.776	**-2.776**	6.215	**0.012**
		O_{3+4+8}	3	2	0	102.26	20.33	0.000	1.000	0.396	0.528
	Fmr1	O_{ST}	4	3	0	19.80	2.00	0.000	1.000	0.300	0.583
		O_{3+4}	9	7	1	17.40	2.00	1.239	-0.239	0.027	0.868
		O_{3+4+8}	2	1	0	19.80	2.00	0.000	1.000	0.100	0.750

N, sample size; P_s, neutral polymorphic sites; P_n, non-neutral polymorphism; D_s, neutral divergence; D_n, non-neutral divergence; *NI*, neutrality index; *α*, proportion of adaptive substitutions. The significant values of *α* with $p<0.05$ are in bold. The estimates were computed with the divergence corrected by JUKES and CANTOR (1969).

Supplementary Table S6b: Generalized MKT for noncoding regions in two populations.

Population	Region		*N*	Polymorphism P_s	Polymorphism P_n	Divergence D_s	Divergence D_n	*NI*	*α*	X^2	*P*
Barcelona	*Acph1*	O_{ST}	8	15	11	77.37	99.83	0.568	0.431	1.799	0.179
		O_{3+4}	10	23	16	69.40	95.58	0.505	0.494	3.639	0.056
		O_{3+4+8}	9	21	20	69.38	94.18	0.701	0.298	1.029	0.310
	Ast	O_{ST}	8	5	22	15.20	154.90	0.431	0.569	2.325	0.127
		O_{3+4}	7	5	39	12.52	153.74	0.635	0.364	0.669	0.413
		O_{3+4+8}	9	4	30	12.52	149.12	0.629	0.370	0.586	0.433
	Iarp	O_{ST}	9	7	5	66.19	31.79	1.487	-0.487	0.408	0.522
		O_{3+4}	11	5	4	67.51	30.40	1.776	-0.776	0.677	0.410
		O_{3+4+8}	10	3	5	68.82	31.79	3.607	-2.607	3.158	0.075
	regIarp	O_{ST}	7	7	37	66.19	400.54	0.873	0.126	0.097	0.754
		O_{3+4}	9	5	48	67.51	405.70	1.597	-0.597	0.936	0.333
		O_{3+4+8}	6	3	32	68.82	407.12	1.803	-0.803	0.935	0.333
	trus	O_{ST}	9	6	5	63.62	48.50	1.093	-0.093	0.019	0.888
		O_{3+4}	10	4	4	61.86	48.50	1.275	-0.275	0.110	0.739
		O_{3+4+8}	8	4	2	60.21	47.22	0.637	0.362	0.261	0.609
	Fmr1	O_{ST}	8	3	22	14.36	348.16	0.302	0.697	3.531	0.060
		O_{3+4}	10	6	45	13.14	355.22	0.277	**0.722**	6.905	**0.008**
		O_{3+4+8}	9	6	31	13.14	360.90	0.188	**0.811**	12.230	**0.000**
	regFmr1	O_{ST}	9	3	17	14.36	54.26	1.499	-0.499	0.345	0.556
		O_{3+4}	10	6	21	13.14	55.35	0.831	0.168	0.110	0.739
		O_{3+4+8}	8	6	17	13.14	73.12	0.509	0.490	1.478	0.223
			N	Polymorphism P_s	Polymorphism P_n	Divergence D_s	Divergence D_n	*NI*	*α*	X^2	*P*
Málaga	*Acph1*	O_{ST}	5	16	6	72.54	96.99	0.280	**0.719**	7.021	**0.008**
		O_{3+4}	9	21	15	66.32	92.78	0.510	0.489	3.290	0.069
		O_{3+4+8}	4	14	12	70.92	95.58	0.636	0.363	1.154	0.282
	Ast	O_{ST}	3	1	9	15.20	158.39	0.863	0.136	0.018	0.892
		O_{3+4}	9	7	43	11.23	151.43	0.455	0.544	2.454	0.117
		O_{3+4+8}	3	2	19	13.84	158.39	0.830	0.169	0.054	0.814
	Iarp	O_{ST}	6	3	5	67.50	31.79	3.538	-2.538	3.052	0.080
		O_{3+4}	9	6	6	70.15	31.79	2.206	-1.206	1.714	0.190
		O_{3+4+8}	5	4	4	68.82	31.79	2.164	-1.164	1.135	0.286
	regIarp	O_{ST}	2	3	23	67.50	412.56	1.254	-0.254	0.130	0.717
		O_{3+4}	4	6	20	70.15	392.91	0.595	0.404	1.176	0.278
		O_{3+4+8}	2	4	5	68.82	390.12	0.220	**0.779**	5.824	**0.015**
	trus	O_{ST}	4	4	1	63.62	48.50	0.327	0.672	1.060	0.303
		O_{3+4}	8	5	1	60.18	49.78	0.241	0.758	1.891	0.169
		O_{3+4+8}	3	2	0	61.91	49.78	0.000	1.000	1.585	0.207
	Fmr1	O_{ST}	5	2	55	14.36	332.80	1.186	-0.186	0.049	0.823
		O_{3+4}	9	5	35	13.14	356.63	0.258	**0.741**	6.823	**0.008**
		O_{3+4+8}	2	1	16	14.36	362.32	0.634	0.365	0.185	0.666
	regFmr1	O_{ST}	7	2	13	14.36	55.35	1.686	-0.686	0.418	0.517
		O_{3+4}	10	5	18	13.14	53.18	0.889	0.110	0.038	0.843
		O_{3+4+8}	7	1	12	14.36	57.61	2.990	-1.990	1.117	0.290

N, sample size; P_s, neutral polymorphic sites; P_n, non-neutral polymorphism; D_s, neutral divergence; D_n, non-neutral divergence; *NI*, neutrality index; *α*, proportion of adaptive substitutions. The significant values of *α* with $p<0.05$ and nearly significant are in bold. The estimates were computed with the divergence corrected by JUKES and CANTOR (1969).

Appendix I

Dolgova et al. BMC Evolutionary Biology 2010, 10:363
http://www.biomedcentral.com/1471-2148/10/363

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Genetic constraints for thermal coadaptation in *Drosophila subobscura*

Olga Dolgova[1], Carla Rego[2], Gemma Calabria[3,4], Joan Balanyà[3,4], Marta Pascual[3,4], Enrico L Rezende[1], Mauro Santos[1*]

Abstract

Background: Behaviour has been traditionally viewed as a driver of subsequent evolution because behavioural adjustments expose organisms to novel environments, which may result in a correlated evolution on other traits. In *Drosophila subobscura*, thermal preference and heat tolerance are linked to chromosomal inversion polymorphisms that show parallel latitudinal clines worldwide, such that "cold-climate" ("warm-climate") chromosome arrangements collectively favour a coherent response to colder (warmer) settings as flies carrying them prefer colder (warmer) conditions and have lower (higher) knock out temperatures. Yet, it is not clear whether a genetic correlation between thermal preference and heat tolerance can partially underlie such response.

Results: We have analyzed the genetic basis of thermal preference and heat tolerance using isochromosomal lines in *D. subobscura*. Chromosome arrangements on the O chromosome were known to have a biometrical effect on thermal preference in a laboratory temperature gradient, and also harbour several genes involved in the heat shock response; in particular, the genes *Hsp68* and *Hsp70*. Our results corroborate that arrangements on chromosome O affect adult thermal preference in a laboratory temperature gradient, with cold-climate O_{st} carriers displaying a lower thermal preference than their warm-climate O_{3+4} and O_{3+4+8} counterparts. However, these chromosome arrangements did not have any effect on adult heat tolerance and, hence, we putatively discard a genetic covariance between both traits arising from linkage disequilibrium between genes affecting thermal preference and candidate genes for heat shock resistance. Nonetheless, a possible association of juvenile thermal preference and heat resistance warrants further analysis.

Conclusions: Thermal preference and heat tolerance in the isochromosomal lines of *D. subobscura* appear to be genetically independent, which might potentially prevent a coherent response of behaviour and physiology (i.e., coadaptation) to thermal selection. If this pattern is general to all chromosomes, then any correlation between thermal preference and heat resistance across latitudinal gradients would likely reflect a pattern of correlated selection rather than genetic correlation.

Background

Ectotherms exhibit a suite of behavioural and physiological strategies to cope with spatiotemporal variation in ambient temperature [1]. For instance, behavioural adjustments (e.g. modifying daily activity patterns and selecting favourable microclimates; [2]) can buffer the impact of sub-optimal temperatures, and are the main means of thermoregulation in small insects [3-5]. Although such adjustments can enable ectotherms to maintain relatively constant body temperatures (T_b) at different seasons and/or latitudes [2,6], the observation of cyclical seasonal changes in genetic markers putatively related to thermal adaptation [7,8] and the clinal variation in thermal stress tolerance in some *Drosophila* species [9-11] suggest that behavioural thermoregulation may be insufficient to fully compensate shifts in environmental temperature [12].

If behavioural thermoregulation is not fully compensatory and climate variation influences the actual T_b and physiological performance of organisms distributed over broad latitudinal ranges (i.e., performance falls below its optimum during cooling and warming), then temperature

* Correspondence: mauro.santos@uab.es
[1]Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva (GBE), Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
Full list of author information is available at the end of the article

BioMed Central
© 2010 Dolgova et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

is more than just a key environmental factor that affects development, growth, and survival of individuals [13,14]: it is likely the main selective agent that drives - directly or indirectly - the evolution of clinal patterns in genetic, phenotypic, and life history traits. Furthermore, the divergence of thermal optima in the different subpopulations according to the T_b experienced by the organism is expected to bolster a covariance between behavioural shifts (thermal preference) and performance [6,15]. This is related to the idea of "coadaptation" [16], where natural selection is supposed to favour the harmonious adjustment among the suite of (co-)evolving traits [7,17]. Parallel clines on different continents or along independent temperature gradients can thus offer an invaluable opportunity to study thermal coadaptation since the role of temperature in driving those clines is quite compelling.

Some widespread latitudinal clines in *Drosophila* also provide an additional advantage for studies of thermal coadaptation: there is a relatively well-known historical record following the invasion of a new geographical region (e.g. [18,19]). Perhaps the best example is that of *Drosophila subobscura*, a native Palaearctic species that invaded the Americas about 30 years ago, and spread rapidly on both South and North America. Clinal patterns for phenotypic traits and genetic polymorphisms emerged very rapidly during these two independent colonization events [20-22]. For instance, North American populations soon evolved decreased desiccation resistance with increasing latitude as expected, which matches the pattern found in Old World populations and suggests that strong selection for thermal-related traits along latitudinal gradients is taking place. On the other hand, in South America this trait shows the opposite pattern: higher desiccation tolerance is observed in colder areas [23]. Contrasting outcomes were also observed for other clinally varying traits - wing cell size and cell number [24], and wing shape [20,25] - where the role of temperature remains elusive, which apparently suggests that selective pressures vary in the different clines. An alternative explanation, however, is that evolution can sometimes be constrained by antagonistic genetic correlations (i.e., genetic correlations among traits that are not in accord with the direction of selection [26,27]) arising from linkage disequilibrium between alleles at different loci, and patterns of linkage disequilibrium can vary among populations or seasons [28,29]. In this context, we now know that contrasting wing shape clines in *D. subobscura* came out as a correlated response of the world-wide parallel inversion clines [21] because inversion-shape relationships in native and colonizing populations are opposite (presumably due to the different associations between inversions and particular alleles which influence the trait), probably as a result of the bottleneck effect that occurred during the colonization of America [30]. Besides, different patterns of linkage disequilibrium could result from variability in migration rates between genetically differentiated populations in the various latitudinal clines [31]. In summary, conflicting outcomes between old and rapidly evolving new clines should probably not be viewed as a nuisance, but as reminder that an appropriate knowledge of the underlying genetic architecture is required to further understand why (or why not) these inconsistencies arise. More specifically, if behaviour "drives" the subsequent parallel evolution in morphology and physiology as predicted ([6]; but see [32]), it is essential to analyze the genetic basis of thermal preference and temperature-related traits to see whether or not thermal coadaptation can happen along a cline.

We have recently undertaken a within-population large-scale study to analyze the association between chromosomal inversion polymorphisms that show parallel latitudinal clines in native and colonizing populations of *D. subobscura*, with the thermal preferences (T_p: the preferred body temperature in a laboratory thermal gradient, which we expect to correlate with the thermal optimum for performance; [33]) and knock out temperatures (T_{ko}: the temperature required to knock out a fly in a water-bath) of their carriers [34]. The main results can be summarized as follows: (*i*) flies carrying "cold-adapted" or "cold-climate" chromosome arrangements (i.e., those chromosome arrangements in all five major acrocentric chromosomes that show a negative correlation coefficient with maximum temperatures along the cline, or a positive correlation coefficient with latitude in Palaearctic populations; [35,36]) prefer a lower T_p and had a lower T_{ko}, in accordance with the natural patterns; (*ii*) different chromosomes were responsible for the bulk of the genetic variation in T_p (chromosomes A and O) and T_{ko} (chromosome E); and (*iii*) T_p and T_{ko} were phenotypically uncorrelated, which agrees with the observation that different independently segregating chromosomes were mainly responsible for the corresponding associations. Taken at a face value, behavioural thermoregulation and performance were indeed "coadapted" in the sense that cold-climate (warm-climate) chromosome arrangements collectively favour a coherent response to colder (warmer) environments, but this was not due to a genetic covariance of behaviour and physiology. There were, however, two potential limitations in the study. First, each individual fly was scored for only one chromosome of its diploid set and, hence, dominance effects (if any) where hidden in the analysis. Second, both intra- and interchromosomal contributions were mixed because the assayed flies had the genetic background from the sampled wild population. Although it might be argued that this protocol is somehow closer to what happens in nature, these uncontrolled factors might have precluded

a better characterization of the underlying genetic effects. Accordingly, although the amount of genetic variation on T_p and T_{ko} explained by the combined effect of all chromosomes carrying at least one cold-climate gene arrangement was statistically significant, it only accounted for 1% of the total phenotypic variation [34].

Here we examine if T_p and T_{ko} are genetically correlated and might evolve in a coherent fashion in response to selection; i.e., whether behaviour and physiology are coadapted at the genetic level. We take advantage of the fact that the polymorphic inversions on chromosome O appear to be associated with behavioural thermoregulation in *D. subobscura* [34], and that this is the only chromosome that can be used to measure the expression of associated traits in replicated inbred and outbred genotypes. Namely, chromosome O is the only one for which a balancer stock (*Va*/*Ba*: *Varicose*/*Bare*; [37]) is available (a balancer is a specially constructed chromosome that carries a dominant morphological marker that is homozygous lethal and multiple inversions to suppress recombination). This is the longest chromosome in *D. subobscura* (190 cM which correspond to approximately 31 Mb [38]), and is homologous to arm 3R in *D. melanogaster* [39,40]. Some chromosome arrangements (O_{st} and O_{3+4}) show conspicuous northwest-southwest latitudinal clines in Palaearctic populations (Figure 1a). Chromosome O harbours several genes involved in the heat shock response [41]; in particular, gene *Hsp*68 (located in section O(89A) [42,43] and relatively close to the proximal breakpoint of inversion O_8 [44]), and gene *Hsp*70 (located in section O(94A) [42,43] and included inside the warm-climate chromosomal arrangement O_{3+4}, and close to the distal breakpoint of inversion O_8 [44]) (Figure 1b). Hsp70 appears to be the primary protein involved in thermotolerance in *D. melanogaster* [45] - though apparently not in other *Drosophila* species [46] -, and *Hsp70* allele frequencies show latitudinal clines and change in response to thermal evolution in the laboratory [47]. In addition, correlated responses to selection for knock down resistance at 39°C have also been found for *Hsp68* in *D. melanogaster* [48].

Previous work also showed that *D. subobscura* flies carrying O chromosomes derived from replicated thermal lines [49,50] that had evolved in the laboratory at warm temperatures (22°C) had a higher total net fitness than its cold-adapted (13°C) counterparts; that is, a significant shift in thermal optima was observed [51]. All in all, it seems that there is indeed room for the coevolution of behaviour and physiological tolerance in *D. subobscura*. However, we show here that thermal preference and heat tolerance appear to be genetically independent. Therefore, any latitudinal correlation between both traits would likely reflect a pattern of correlated selection across populations rather than within-population genetic correlations.

Experimental settings

In south-western European populations, the most frequent chromosome arrangements for chromosome O are O_{st}, O_{3+4}, O_{3+4+7}, and O_{3+4+8} [52]. The first two arrangements show a clear contrasting clinal pattern in original Palaearctic populations, with O_{st} increasing and O_{3+4} decreasing in frequency with increasing latitude [35,56] (Figure 1a). Arrangement O_{3+4+8} is also interesting because in historical times it was mainly restricted to the Mediterranean region, being the most abundant chromosomal arrangement in northern Africa [53]. However, in the last decades its distribution has changed dramatically and recent surveys revealed frequencies as high as 22.6% in Groningen, Netherlands, where it was previously absent [22,54]. Six independent isochromosomal lines for each of these three arrangements (i.e., $O_j^1,\dots,O_j^6$; j = st, 3 + 4, 3 + 4 + 8) were used in the present experiments. Extensive genetic differentiation of up to 4 Mb (i.e., about 15% of the euchromatic portion) has been detected among these arrangements [55]. In other words, there are compelling reasons to think that the chromosome arrangements used in this work are genetically differentiated for *Hsp*70, and probably also for *Hsp68* since inversion effects can extend as far as 1000 kilobases outside from breakpoints [56,57].

Following Santos et al. [58] the experimental flies were obtained from 54 crosses, which will be referred to as inbred (isogenic: $O_j^1 \times O_j^1, O_j^2 \times O_j^2, \dots, O_j^6 \times O_j^6$ with 18 crosses in total), or as outbred including both structural homokaryotypes ($O_j^1 \times O_j^2, O_j^2 \times O_j^3, \dots, O_j^6 \times O_j^1$ with 18 cyclically permuted reciprocal crosses in total) and heterokaryotypes ($O_j^1 \times O_k^1, O_j^2 \times O_k^2, \dots, O_j^6 \times O_k^6$; $j \neq k$; with 18 reciprocal crosses in total). Two developmental temperatures were used in the experiment to study potentially important effects of phenotypic plasticity: 18°C and 22°C. The reason for this was the huge difference (about 7°C-8°C) between our previous estimate of T_p (pooled average 16.6°C; [34]) in *D. subobscura* flies raised at 18°C, and that obtained by Huey and Pascual (23.7°C; [12]) where flies were raised at 22°C. Even though the flies assayed came from different sources - south-western Europe in Rego et al. [34], and North America in Huey and Pascual [12] -, which could account for the observed difference because thermal responses can vary between populations [59], it remains to be seen whether developmental plasticity can affect estimates of thermal preference and heat tolerance.

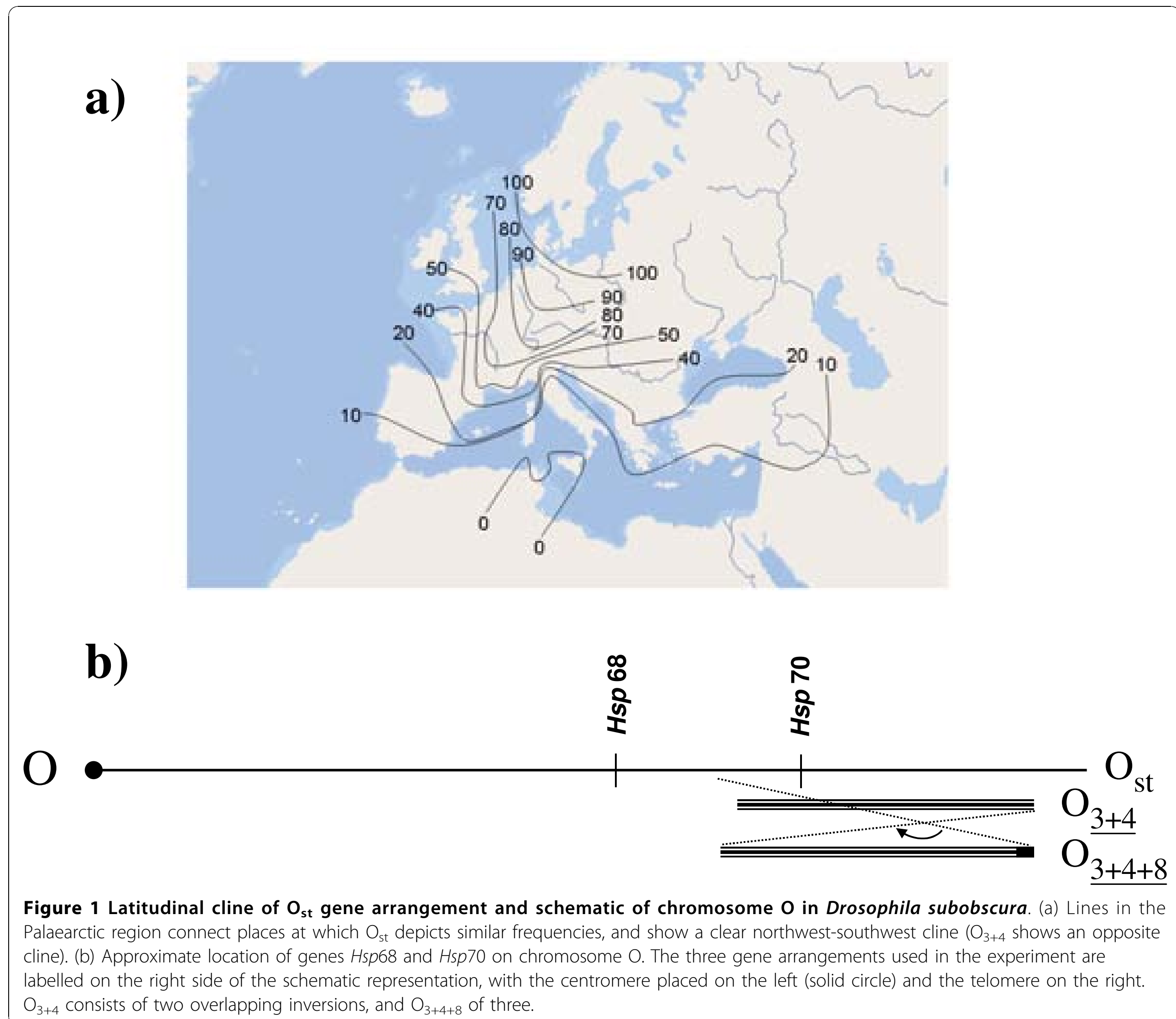


Figure 1 Latitudinal cline of O_{st} gene arrangement and schematic of chromosome O in *Drosophila subobscura*. (a) Lines in the Palaearctic region connect places at which O_{st} depicts similar frequencies, and show a clear northwest-southwest cline (O_{3+4} shows an opposite cline). (b) Approximate location of genes *Hsp*68 and *Hsp*70 on chromosome O. The three gene arrangements used in the experiment are labelled on the right side of the schematic representation, with the centromere placed on the left (solid circle) and the telomere on the right. O_{3+4} consists of two overlapping inversions, and O_{3+4+8} of three.

Results

Association between thermal preference and knock out temperature

The phenotypic correlation between T_{p} and T_{ko} was assessed from their partial correlation coefficient, holding constant the variables developmental temperature, sex, plate hour, and water bath (see Methods). In no case were the partial correlations statistically significant: inbred crosses $r_{T_{\mathrm{p}} \cdot T_{\mathrm{ko}}} = 0.065$, t = 1.21, df = 347, P = 0.226; outbred crosses $r_{T_{\mathrm{p}} \cdot T_{\mathrm{ko}}} = -0.030$, t = 0.79, df = 701, P = 0.429. Furthermore, as expected from the low values of the phenotypic correlation, the genetic (karyotypic) correlation for the outbred flies was also close to zero (r_k = -0.068, P = 0.914). The conclusion is that both traits are nearly orthogonal to each other (pooled $r_{T_{\mathrm{p}} \cdot T_{\mathrm{ko}}} = 1.2 \times 10^{-4}$, t = 0.004, df = 1054, P = 0.997) and, hence, they will be analyzed separately in what follows.

Consanguinity and developmental effects

a) Thermal preference

Inbreeding and developmental temperature effects on T_{p} were simultaneously analyzed by contrasting isogenic *vs*. outbred homokaryotypic flies reared at both experimental temperatures (Figure 2). The factorial analysis of covariance (ANCOVA) only detected statistically significant differences for karyotypes, karyotype × inbreeding interaction, and karyotype × developmental temperature interaction effects (Table 1). Average (± SD) T_{p} was not different between rearing temperatures (flies reared at 18°C: 18.7°C ± 4.1°C; flies reared at 22°C: 18.8°C ± 3.1°C) or

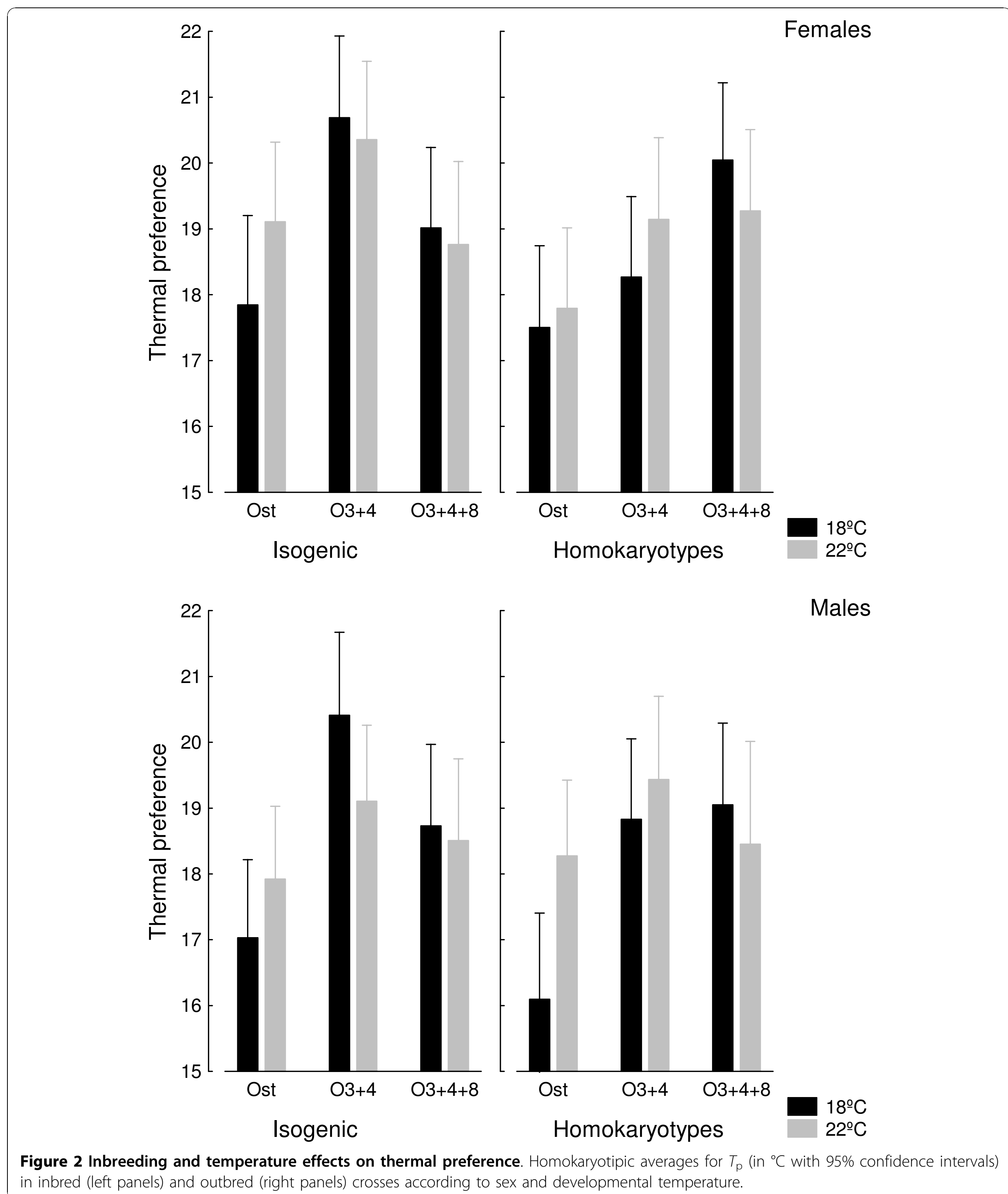


Figure 2 Inbreeding and temperature effects on thermal preference. Homokaryotipic averages for T_p (in °C with 95% confidence intervals) in inbred (left panels) and outbred (right panels) crosses according to sex and developmental temperature.

sexes (females: 19.0°C ± 3.6°C; males: 18.5°C ± 3.6°C), although in this last case the effect was marginally non-significant (P = 0.053). Permutation tests (see Methods) corroborated that the three assayed karyotypes differ in T_p (P = 0.001).

Scheffé post hoc tests using the mean square of the nested "cross" effect as the error term showed that the thermal preference of O_{st}/O_{st} flies was significantly lower when compared to those of O_{3+4}/O_{3+4} and O_{3+4+8}/O_{3+4+8} homokaryotypes, which did not differ between

Table 1 Inbreeding and temperature effects on thermal preference

Source of variation	d.f.	Mean Square	*F*	*P*
Covariate (plate hour)	1	25.502	2.07	0.151
Karyotype (κ)	2	231.515	18.29	<0.001
Cross $\subset \kappa$	15	12.676	1.03	0.425
Inbreeding (ι)	1	30.514	2.47	0.116
Temperature (τ)	1	4.119	0.33	0.564
Sex (ς)	1	46.227	3.74	0.053
$\kappa \times \iota$	2	40.337	3.27	0.039
$\kappa \times \tau$	2	40.031	3.24	0.040
$\kappa \times \varsigma$	2	6.195	0.50	0.606
$\iota \times \tau$	1	11.063	0.90	0.344
$\iota \times \varsigma$	1	6.257	0.51	0.477
$\tau \times \varsigma$	1	0.408	0.03	0.856
$\kappa \times \iota \times \tau$	2	17.477	1.42	0.243
$\kappa \times \iota \times \varsigma$	2	11.532	0.93	0.393
$\kappa \times \tau \times \varsigma$	2	7.600	0.62	0.541
$\iota \times \tau \times \varsigma$	1	12.123	0.98	0.322
$\kappa \times \iota \times \tau \times \varsigma$	2	4.245	0.34	0.709
Error	717	12.346		

Flies raised from inbred (isogenic) and outbred crosses of *Drosophila subobscura* reared at 18°C and 22°C. Karyotypes being compared are O_{st}/O_{st}, O_{3+4}/O_{3+4}, and O_{3+4+8}/O_{3+4+8}. ($\subset$ means "nested in".)

them. The difference is consistent for both isogenic and outbred flies (Figure 2). From the present data we can conclude that the preferred temperature ranges or "set point" (T_{set}) ranges (central 50% of preferred body temperatures; [60]) are bounded by 15.1°C - 20.5°C for O_{st}/O_{st} karyotypes, and 16.6°C - 22.2°C for the other two karyotypes.

The karyotype × inbreeding interaction arises from the somewhat different behaviour between O_{st}/O_{st} and O_{3+4}/O_{3+4} karyotypes on one side, and O_{3+4+8}/O_{3+4+8} on the other: for the first two karyotypes T_p was slightly higher in inbred crosses when compared to their outbred counterparts, whereas the opposite was true for the O_{3+4+8}/O_{3+4+8} karyotype. Average T_p was, however, almost identical for inbred (18.9°C ± 3.6°C) and outbred (18.5°C ± 3.6°C) flies. On the other hand, O_{st}/O_{st} flies raised at 22°C had a higher T_p than those raised at 18°C, but no clear trend was observed for O_{3+4}/O_{3+4} and O_{3+4+8}/O_{3+4+8} karyotypes.

b) Knock out temperature

Knock out temperatures are plotted in Figure 3. The ANCOVA (Table 2) detected statistically significant differences for the effects of rearing temperature and sex. Flies reared at 18°C had a higher T_{ko} than flies reared at 22°C (mean ± SD: 33.3°C ± 2.1°C *vs*. 32.6°C ± 2.3°C), and females had a higher T_{ko} than males (33.4°C ± 1.9°C *vs*. 32.5°C ± 2.4°C). Even though T_{ko} was slightly lower for the isogenic lines when compared to their outbred counterparts (32.8°C ± 2.2°C *vs*. 33.1°C ± 2.2°C), inbreeding effects were clearly non-significant (P = 0.136).

Gene arrangement effects in the outbred lines

a) Thermal preference

The genetic and environmental (developmental temperature) contributions of chromosome O to T_p (and T_{ko}; below) was assessed from the outbred crosses including all possible karyotypes. Outbred crosses are obviously more relevant to the real situation because inbred genotypes are homozygous for deleterious alleles, and also for alleles that might display heterozygote advantage in the original outbred population. The only statistically significant effects detected by the ANCOVA model (Table 3) were those arising from genetic differences among karyotypes (permutation tests corroborated that the three assayed karyotypes differ in T_p; $P = 0.0018$) and sexes, with females having a higher T_p (mean ± SD: 18.7°C ± 3.6°C) than males (18.0°C ± 3.6°C). As above, average T_p was slightly lower for flies reared at 18°C (18.1°C ± 4.0°C) than at 22°C (18.6°C ± 3.2°C), but the difference was marginally non-significant ($P = 0.069$).

The linear contrast between the two O_{st}/O^*_{3+4} heterokaryotypes (O^*_{3+4} pools into a single class the arrangements that share O_{3+4}; see Methods) reveals that O_{st}/O_{3+4} and O_{st}/O_{3+4+8} flies displayed a similar average T_p (18.5°C ± 3.8°C *vs*. 18.0°C ± 3.7°C, respectively). However, some differences were detected among the three O^*_{3+4}/O^*_{3+4} karyotypes, which can be attributed to some under-dominance because average T_p for O_{3+4}/O_{3+4+8} flies (18.1°C ± 3.4°C) was lower than that for the corresponding homokaryotypes (O_{3+4}/O_{3+4}: 18.9°C ± 3.5°C; O_{3+4+8}/O_{3+4+8}: 19.3°C ± 3.6°C). In any case, the main difference was between O_{st} and O^*_{3+4} carriers, with mainly additive genetic effects (Figure 4). As already indicated, O_{st}/O_{st} flies clearly preferred lower temperatures than O_{3+4}/O_{3+4} or O_{3+4+8}/O_{3+4+8} flies.

b) Knock out temperature

The ANCOVA for T_{ko} (Table 4) did not detect any difference among karyotypes, in accordance with the previous findings for the inbred crosses. Similarly, the main differences arose between developmental temperature (flies reared at 18°C: 33.6°C ± 1.9°C; flies reared at 22°C: 32.8°C ± 2.3°C) and sex (females: 33.7°C ± 1.8°C; males: 32.7°C ± 2.4°C).

The genetic correlation between T_p and T_{ko} after pooling O_{3+4} and $O_{3+4+8}$was r_p = -0.130 (P = 0.917). Again, the conclusion is that these two traits are uncorrelated. Figure 4 plots the genotypic values in the

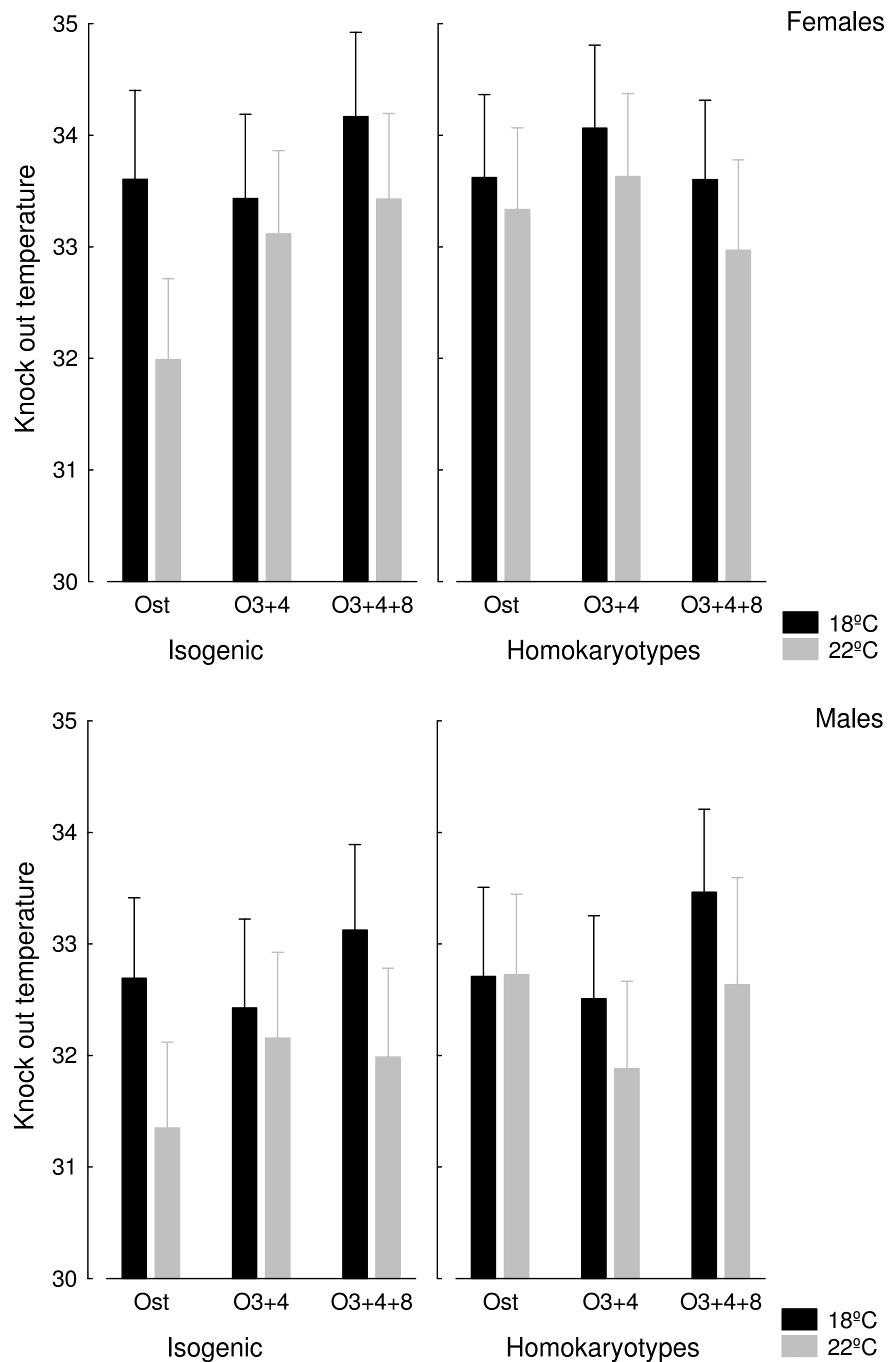


Figure 3 Inbreeding and temperature effects on knock out temperature. Homokaryotipic averages for T_{ko} (in °C with 95% confidence intervals) in inbred (left panels) and outbred (right panels) crosses according to sex and developmental temperature.

additive-dominance scales for T_p and T_{ko}, together with their statistical significance obtained from the appropriate contrasts (Table 3, 4).

c) Average effects on thermal preference

Our experiment only provides an estimation of the gene (chromosome O) action on T_p and does not allow inferences to the base population. It is possible, however, to obtain estimates of the average effects, or "statistically additive effects", by taking into account the gene action and allelic (chromosome arrangement) frequencies in the natural populations [61]. Assuming that the chromosome arrangement effects are roughly the same along the cline (for a measure of climatic temperatures along the Palaearctic cline see Figure 1 in [62]), Table 5 gives

Table 2 Inbreeding and temperature effects on knockout temperature

Source of variation	d.f.	Mean Square	*F*	*P*
Covariate (water bath)	1	103.117	24.04	<0.001
Karyotype (κ)	2	3.878	0.36	0.704
Cross ⊂ κ	15	11.027	2.57	0.001
Inbreeding (ι)	1	9.538	2.22	0.136
Temperature (τ)	1	77.034	17.96	<0.001
Sex (ς)	1	154.979	36.13	<0.001
$\kappa \times \iota$	2	4.176	0.97	0.378
$\kappa \times \tau$	2	1.999	0.47	0.628
$\kappa \times \varsigma$	2	8.106	1.89	0.152
$\iota \times \tau$	1	1.047	0.24	0.621
$\iota \times \varsigma$	1	0.435	0.10	0.750
$\tau \times \varsigma$	1	0.022	0.01	0.943
$\kappa \times \iota \times \tau$	2	7.798	1.82	0.163
$\kappa \times \iota \times \varsigma$	2	8.926	2.08	0.126
$\kappa \times \tau \times \varsigma$	2	1.693	0.39	0.674
$\iota \times \tau \times \varsigma$	1	0.241	0.06	0.813
$\kappa \times \iota \times \tau \times \varsigma$	2	0.159	0.04	0.964
Error	668	4.289		

Flies raised from inbred (isogenic) and outbred crosses of *Drosophila subobscura* reared at 18°C and 22°C. Karyotypes being compared are O_{st}/O_{st}, O_{3+4}/O_{3+4}, and O_{3+4+8}/O_{3+4+8}. (⊂ means "nested in".)

the average effects (females and males pooled) estimated from the frequencies of the different arrangements in European populations spanning about 17° latitude [52,54]. The interpretation is that flies inheriting a O_{st} chromosome will choose a temperature ranging from around 0.31°C - 0.45°C below the average temperature chosen by the population (conversely, flies carrying warm-climate chromosome arrangements will choose a temperature ranging from around 0.03°C - 0.52°C above the average).

Combined with our previous results with chromosome A (which is the sex chromosome and additive values can be estimated using males' T_p; [34]), where gene arrangement A_{st} exhibits a similar latitudinal pattern than O_{st} and flies carrying A_{st} also display a laboratory thermal preference towards colder temperature, the conclusion is that flies inheriting simultaneously A_{st} and O_{st} will choose temperatures ranging from approximately 0.5°C - 1.0°C below the average (these estimates assume perfect additivity).

Discussion

The present results with isogenic lines and their crosses corroborate and extend our previous work with wild flies from south-western Europe [34]. They confirm that arrangements on chromosome O have a biometrical effect on thermal preference in a laboratory temperature gradient, with cold-climate O_{st} carriers displaying a lower T_p than their warm-climate O_{3+4} and O_{3+4+8} counterparts. In addition, T_p and T_{ko} were again found to be uncorrelated, and we can now discard a potential genetic covariance between both traits arising from linkage disequilibrium between genes affecting thermal preference and candidate genes for heat shock resistance (i.e., *Hsp68* and *Hsp70*; [42,43]) located inside of, or close to, the chromosome regions covered by the inversions analyzed here (see Background). In other words, we conclude that variation on O chromosome arrangements does not have any effect on knock out temperature (but see below). Note, however, that this does not imply that genes on chromosome O have no effect on

Table 3 Karyotype and temperature effects on thermal preference

Source of variation	d.f.	Mean Square	*F*	*P*
Covariate (plate hour)	1	147.947	11.84	<0.001
Karyotype (κ)	5	60.774	4.97	0.002
O_{st}/O^*_{3+4}	1	0.853	0.07	0.793
O^*_{3+4}/O^*_{3+4}	2	42.884	3.51	0.043
O_{st}/O_{st}, O_{st}/O^*_{3+4}, O^*_{3+4}/O^*_{3+4}	2	106.330	8.70	0.001
additive effect	1	205.854	16.85	<0.001
dominance effect	1	3.532	0.29	0.595
Cross ⊂ κ	30	12.220	0.98	0.502
Temperature (τ)	1	41.328	3.31	0.069
Sex (ς)	1	91.221	7.30	0.007
$\kappa \times \tau$	5	19.791	1.58	0.162
$\kappa \times \varsigma$	5	10.805	0.86	0.505
$\tau \times \varsigma$	1	4.948	0.40	0.529
$\kappa \times \tau \times \varsigma$	5	8.863	0.71	0.617
Error	691	12.498		

Flies raised from outbred crosses of *Drosophila subobscura* reared at 18°C and 22°C. Karyotypes being compared are O_{st}/O_{st}, O_{3+4}/O_{3+4}, O_{3+4+8}/O_{3+4+8}, O_{st}/O_{3+4}, O_{st}/O_{3+4+8} and O_{3+4}/O_{3+4+8}. O^*_{3+4} stands for O_{3+4} + O_{3+4+8}. (⊂ means "nested in".)

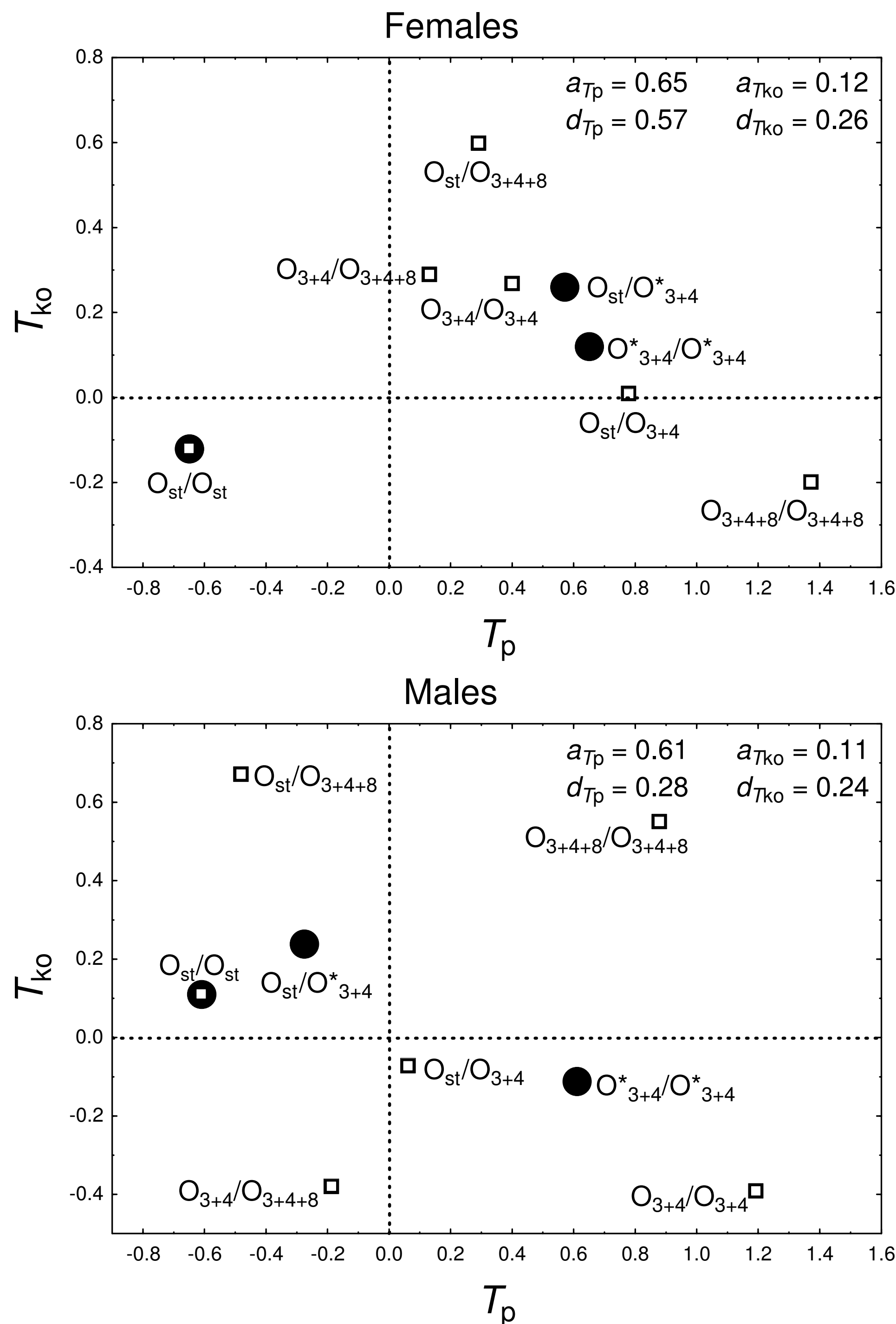


Figure 4 Karyotypic values in the additive-dominance scale. Deviation values for thermal preference (T_p) and knockout temperature (T_{ko}) were measured after pooling arrangements O_{3+4} and O_{3+4+8} into a single class (O^*_{3+4}), and the coordinate point (0, 0) was taken as the midparent (i.e., the average of T_p and T_{ko} for the two karyotypes O_{st}/O_{st} and O^*_{3+4}/O^*_{3+4}). Females (upper panel) and males (lower panel) are plotted separately because the interaction karyotype × sex was statistically significant for T_{ko} (Table 4). In the original scale the (0, 0) point corresponds to an average T_p of 18.31°C for females and 17.91°C for males, and an average T_{ko} of 33.58°C for females and 32.61°C for males. Open squares give the values for all six karyotypes to appreciate their dispersion from the midparent, as well as their dispersion from the pooled O_{st}/O^*_{3+4} and O^*_{3+4}/O^*_{3+4} karyotypes (black circles). Statistical significance for additive ($a_{T_p}, a_{T_{ko}}$) and dominance ($d_{T_p}, d_{T_{ko}}$) effects are given in Tables 3 and 4. Note also that the phenotypic ($r_{T_p \cdot T_{ko}} = -0.030$) and genetic ($r_k$ = -0.068, r_p = -0.130; see Methods) correlations were non-significantly different from zero (see text for details).

Table 4 Karyotype and temperature effects on knockout temperature

Source of variation	d.f.	Mean Square	*F*	*P*
Covariate (water bath)	1	101.377	25.87	<0.001
Karyotype (κ)	5	4.295	0.57	0.724
O_{st}/O^*_{3+4}	1	11.598	1.52	0.228
O^*_{3+4}/O^*_{3+4}	2	0.016	0.002	0.998
O_{st}/O_{st}, O_{st}/O^*_{3+4}, O^*_{3+4}/O^*_{3+4}	2	4.872	0.64	0.536
additive effect	1	0.015	0.001	0.965
dominance effect	1	8.632	1.13	0.296
Cross $\subset \kappa$	30	7.641	1.95	0.002
Temperature (τ)	1	107.075	27.33	<0.001
Sex (ς)	1	180.874	46.16	<0.001
$\kappa \times \tau$	5	7.576	1.93	0.087
$\kappa \times \varsigma$	5	8.777	2.24	0.049
$\tau \times \varsigma$	1	1.650	0.42	0.517
$\kappa \times \tau \times \varsigma$	5	2.329	0.59	0.704
Error	654	3.918		

Flies raised from outbred crosses of *Drosophila subobscura* reared at 18°C and 22°C. Karyotypes being compared are O_{st}/O_{st}, O_{3+4}/O_{3+4}, O_{3+4+8}/O_{3+4+8}, O_{st}/O_{3+4+8}, and O_{3+4}/O_{3+4+8}. O^*_{3+4} stands for O_{3+4} + O_{3+4+8}. ($\subset$ means "nested in".)

T_{ko} (actually, statistically significant differences were detected among crosses within karyotypes; Table 4); it simply indicates that any allelic variation of putative genes influencing this trait is not in linkage disequilibrium with inversions on this chromosome.

The new findings were: (*i*) a lack of inbreeding depression for both T_p and T_{ko}; (*ii*) a lack of phenotypic plasticity for T_p according to the temperature at which the flies were raised (18°C and 22°C); and (*iii*) a substantial effect of developmental temperature on T_{ko}. The absence of inbreeding depression for T_p agrees with the genetic analysis from outbred flies, where a dominance effect after pooling chromosome arrangements O_{3+4} and O_{3+4+8} into a single class (O^*_{3+4}) was absent (Table 3; note that the differences detected among the three O^*_{3+4}/O^*_{3+4} karyotypes, and attributed to some underdominance, could not be appreciated in the inbreeding analysis because it only included inbred and outbred homokaryotypes). On the other hand, the lack of inbreeding depression for T_{ko} is expected and does not mean anything here, simply because no "gene" effects linked to chromosomal arrangements on chromosome O were detected. At first sight this might be surprising because a well-characterized cellular defence mechanism once environmental temperature approaches the upper

Table 5 Average effect of chromosome O on thermal preferences (°C)

		Frequency		Average effect	
Population	**Coordinates**	O_{st}	O^*_{3+4}	O_{st}	**Rest**
Málaga (Spain)	36°43′N–4°25′W	0.080	0.407	-0.4506	0.0392
Punta Umbría (Spain)	37°10′N–6°57′W	0.066	0.410	-0.4494	0.0318
Calviá (Spain)	39°33′N–2°29′E	0.057	0.590	-0.4485	0.0271
Riba-roja (Spain)	39°33′N–0°34′W	0.148	0.324	-0.4530	0.0787
Queralbs (Spain)	42°13′N–2°10′E	0.290	0.493	-0.4395	0.1795
Lagrasse (France)	43°05′N–2°37′E	0.330	0.590	-0.4312	0.2124
Montpellier (France)	43°36′N–3°53′E	0.362	0.557	-0.4232	0.2401
Villars (France)	45°26′N–0°44′E	0.389	0.581	-0.4155	0.2645
Leuk (Switzerland)	46°19′N–7°39′E	0.595	0.365	-0.3267	0.4800
Vienna (Austria)	48°13′N–16°22′E	0.625	0.270	-0.3095	0.5158
Tübingen (Germany)	48°32′N–9°04′E	0.606	0.351	-0.3205	0.4930
Louvain-la-Neuve (Belgique)	50°43′N–4°37′E	0.397	0.540	-0.4130	0.2719
Groningen (The Netherlands)	53°13′N–6°35′E	0.502	0.405	-0.3733	0.3763

O^*_{3+4} pools gene arragements O_{3+4} and O_{3+4+8} used in the present work. Together with O_{st}, their combined frequency is ≥ 0.90 in central European populations and drops to approximately 0.50 in south-western Europe, where arrangement O_{3+4+7} is also frequent. However, from our previous data [34] no difference in T_p is detected between O_{3+4+7} and O^*_{3+4}, which justifies their pooling and allows estimating average effects assuming two gene arrangements: O_{st} and the rest. Gene arrangement frequencies where taken from the "new collections" in Solé et al. [52] and Balanyà et al. [54].

thermal limits is the heat shock response, and in *D. melanogaster* the major inducible heat shock protein Hsp70 appears to be the primary protein involved in thermotolerance [45,63]. Recent work, however, questions the pervasive role of Hsp70 in the mediation of the heat stress response and suggests that it may be life-stage specific, being important in larvae but not in adults [64]. Our results are apparently consistent with the lack of association between Hsp70 and adult heat resistance (but see further discussion below), although also raise a caveat to the conclusion that there is no covariance between T_p and T_{ko}. Thus, it could be the case that Hsp70 variation across karyotypes is associated with juvenile tolerance to heat stress, an important trait in *Drosophila* particularly in summer when larval feeding patches can become lethally hot [65]. This possibility warrants further analysis.

An important concern here is that Hsp70 production might not be inducible in the dynamic experimental protocol we used to estimate upper thermal tolerance, where temperature increased 0.1°C min^{-1}. One apparently compelling reason for this is that the estimated maximum thermal limits that *D. melanogaster* can tolerate decrease from approximately 39.9°C with heating rate 0.5°C min^{-1} to 38.7°C with heating rate 0.1°C min^{-1} [66], a puzzling result because slower heating rates should allow individuals to acclimatize to new temperatures and also because slow heating rates pre-exposes individuals to non-lethal high temperatures ("hardening"), which increases heat shock resistance [10]. We have recently discussed why these conflicting outcomes arise, and suggest that the contribution of other stressors (e.g. higher desiccation in long thermal tolerance assays associated with slow warming rates) can potentially overshadow thermal acclimation effects in dynamic assays with varying heating rates [67]. In other words, we challenge the idea that induced thermotolerance does not occur in dynamic assays with slow heating rates. At this stage this is just speculative because Hsp70 production was not measured in our flies, but the problem is important because *Drosophila* adults are likely to experience slow heating rates in nature of 0.06 - 0.1°C min^{-1} [66,68] and further empirical studies are required to explain the apparently inconsistent findings.

The pooled average T_p here was (mean ± SD) 18.4°C ± 3.6°C (T_{set}: 15.4°C - 21.2°C; these figures include only outbred lines) and about the same at both rearing temperatures. The difference with our previous estimate for wild-flies from Adraga (16.6°C, T_{set}:12.4°C - 20.4°C; [34]) does not seem to be overreached, and could be partially explained by the fact that the present flies were genetically homogeneous for all chromosomes from the *ch-cu* marker strain but chromosome O (recall that the sex chromosome A also had a significant effect on T_p; [34]). This strain has a long history of maintenance at 18°C in the laboratory. In any case, our estimates remain substantially lower than that from Huey and Pascual (23.7°C, T_{set}: 21.2°C - 25.9°C; [12]), and the difference cannot be accounted by flies' rearing temperature. No reasonable explanation for the discrepancy can be offered at this moment, but the additional result that developmental temperature substantially affected T_{ko} makes us confidently conclude that our estimates are indeed closer to the actual T_p of the species. Flies reared at 22°C showed lower heat resistance than their counterparts reared at 18°C (32.8°C *vs*. 33.6°C; outbred lines), which could be a consequence of their smaller size due to the inverse relationship between body size and developmental temperature [69,70]. However, resistance to heat does not seem to be associated with body size [71] - we have also analyzed the association between T_{ko} and wing size from our previous experiment where both traits were recorded [30,34] and found no relationship whatsoever (results not shown). Most likely, 22°C was a suboptimal and potentially stressful temperature for our flies, making them to be weaker and less resistant to the heat shock. Note, however, that this conclusion might not be extrapolated to wild flies that harbour higher levels of genetic variability than our chromosomal lines.

To interpret the interplay between thermal preference and heat stress resistance, an understanding of the environmental temperatures experienced by *D. subobscura* along climatic gradients is required. As far as we are aware, the only data available on T_b for active flies along a latitudinal gradient (spanning 12°) come from recent work by Huey and Pascual [12] in western North America. They found that mean T_b varies by as much as 21°C (from 8°C to 29°C), and that the temporal activity of flies during the day did not match predictions from optimal temperature regulation or desiccation avoidance. Temperatures of maximum activity in summer (Figure 2 in [12]) - when wild flies are smaller probably due to their higher developmental temperatures and/or crowding conditions [72]; and crowding is known to affect adult thermal stress resistance in *Drosophila* [73] - are dangerously close to the T_{ko} obtained here for the outbred flies raised at 22°C. This suggests that active *D. subobscura* flies can experience extreme conditions in the wild, and one would expect flies' activity to be correlated with heat resistance under these conditions if behaviour and physiology were coadapted. Some evidence indicates that diurnal activity patterns in summer can vary according to inversion polymorphism, and chromosome arrangements on the O chromosome seem to behave as expected from our data: O_{st} is more frequent towards the evening while chromosomes carrying gene arrangement O_{3+4} are most frequently sampled at midday [74]. This behavioural thermoregulation,

however, would not confer less susceptibility to high temperatures because the genetic basis of both traits does not seem to allow for the building up of "coadaptation". It is well known from basic population genetics theory that genetic covariance between traits can arise when alleles at different loci are associated (linkage disequilibrium), and this critically depends on relatively low recombination rates [75]. The lack of association between T_p and T_{ko} in *D. subobscura* is fully consistent with their genetic basis as independently segregating chromosomes are involved [34]. Yet, a correlation between these traits can be expected at the interpopulational level due to patterns of correlated selection (rather than genetic correlations) across a latitudinal gradient because of the congruent latitudinal clinal variation for chromosome arrangements on the E (which influences T_{ko} [34]), and on the A and O chromosomes (which influence T_p [[34], this work]).

We now speculate that the mismatch between T_p and T_{ko} could apparently generate an interesting dynamics in the population frequencies of different chromosome arrangements on chromosome O. Suppose the daily activity of flies in the warmest months of the year follows the previously described pattern; that is, flies carrying gene arrangement O_{3+4} are more active at midday and, therefore, have a higher risk of a heat shock than O_{st} and are selected against. On the other hand, assuming T_p corresponds closely with temperatures that maximize fitness O_{3+4} flies likely enjoy a fitness advantage in summer. The net effect would be a compromise between "behaviour unresponsiveness" and general performance, which means that chromosome arrangements on chromosome O may or may not cycle seasonally according to average environmental temperature (i.e., O_{3+4} could be expected to increase in frequency in summer and decrease in winter if general performance is what matters). Interestingly, both patterns have been detected: consistent seasonal cycling at a north-western population in Spain [8] and apparently no seasonal variation at a north-eastern population also in Spain [76]. The point here is that parallel seasonal changes should also be detected for chromosome A since it also affects T_p [34]. In accordance with this prediction, no seasonal cycling was detected for chromosome A in the north-eastern population, but unfortunately no information is available for the other population because chromosome O was the only chromosome scored. It would be very interesting to see what happens for chromosome A in the cycling population.

Conclusions

For ectotherms facing spatiotemporal variation in environmental temperature theory predicts that a coevolution between thermal preference and physiological performance can occur [1]. In the widespread species *D. subobscura* behavioural thermoregulation and heat tolerance are "coadapted" in the sense that flies carrying cold-climate (warm-climate) chromosome arrangements tend to choose colder (warmer) temperatures and have lower (higher) heat stress tolerance [34]. We have analyzed the genetic basis of these thermal traits using isochromosomal lines for the O chromosome. This chromosome was known to affect thermal preference [34], and also harbours several genes involved in the heat shock response (*Hsp68* and *Hsp70*) [42,43]. These genes are located inside of, or close to, the chromosome regions covered by inversions that show conspicuous northwest-southwest latitudinal clines in Palaearctic populations, as well as seasonal fluctuations that are in agreement with the latitudinal patterns [22]. Our results corroborate that arrangements on chromosome O affect adult thermal preference: flies inheriting the cold-climate O_{st} chromosome are predicted to choose a temperature around 0.31°C - 0.45°C below the average temperature chosen by the population and, conversely, flies inheriting the warm-climate O_{3+4} and O_{3+4+8} chromosomes are expected to choose a temperature ranging from around 0.03°C - 0.52°C above the average. However, these chromosome arrangements did not have any differential effect on adult heat tolerance. We conclude that thermal preference and heat tolerance in *D. subobscura* appear to be genetically independent and, therefore, any latitudinal correlation between both traits would likely reflect a pattern of correlated selection across populations rather than within-population genetic correlations.

Methods

Origin of flies and experimental procedures

D. subobscura wild flies were collected near Barcelona (41°43'N, 2°13'E) in October 2007. More than 200 isofemale lines were derived and used to obtain isochromosomal lines for the O chromosome in an otherwise homogeneous genetic background following standard protocols [77]. Briefly, one offspring male from each isofemale line was crossed to three or four virgin females from the *ch-cu* marker strain, which is homozygous for the morphological recessive markers on the O chromosome *cherry* eyes (*ch*) and *curled* wings (*cu*) and the chromosomal arrangement O_{3+4}. A single wild-type male from each cross was repeatedly backcrossed to three or four *ch-cu* females for at least five generations in order to homogenize the genetic background, and the chromosomal arrangement carried by the wild chromosome was identified after the second backcross. To derive the isochromosomal lines, wild-type males from each line were crossed with the *Va*/*Ba* balancer stock [78], which has the same genetic background as the

ch-cu strain. Once obtained, the isochromosomal lines were genotyped for 13 microsatellite loci located on the O chromosome to check that no recombination events occurred during the different crosses. The 18 independent isochromosomal lines used in this study (see Experimental settings) were found to be homozygous for all the loci. The lines were kept at 18°C (12:12 light/dark cycle) in 130-mL bottles with low adult density (around 20 pairs/bottle) to standardize the rearing conditions before egg collections.

To obtain the experimental flies, all 54 crosses (inbred and outbred) were performed at 18°C by mating 4 days-old virgin males and females from the corresponding isochromosomal lines. After six days the males were discarded and the females (an equal number from each reciprocal cross in the outbred combinations) were transferred to egg-laying chambers containing fresh food and charcoal colouring. Eggs were placed in vials (45 eggs/vial containing 6 mL of food) at two rearing temperatures: 18°C and 22°C. Non-anaesthetized emerging flies were stored in bottles at low adult density and used to evaluate laboratory thermal preference (T_p) and knock out temperature (T_{ko}) for each cross (see below). All fly handling was done at room temperature using CO_2 anaesthesia only to sort virgin flies and to place females in the egg-laying chambers.

Thermal preference behaviour in a laboratory gradient and heat resistance

Laboratory T_p was measured as previously described [34]. Briefly, adult flies (about 7 days old) were individually placed in separate lanes on an aluminium base plate where a thermal gradient with temperatures ranging from 11°C to 29°C was generated. Adults were given approximately 1 h to adjust, and afterwards their positions were recorded four times every 10 min. We used the median of the four measurements to estimate T_p of each fly. Measurements were performed in a room with constant temperature (22°C - 23°C), and the flies were assayed under white light illumination. This protocol renders a repeatable assessment of flies' thermal preferences [34]. After the thermal preference assay, each fly was gently removed from the lane and individually placed in a vial with fresh food for the subsequent assay of heat stress tolerance.

One day after measurements of thermal preference flies were assayed for heat resistance also as previously described [34]. Adults were individually placed in sealed empty vials and immersed in water-baths at T_{min} = 24°C. Every 10 min individuals were scored for mobility (fly active or knocked out) and the temperature of the water was increased by $\Delta T = +1$°C. The procedure was repeated until the water-baths reached T_{max}, defined as the temperature when the last active fly was knocked out (T_{max} = 38°C was the upper limit in the assays; median T_{max} = 33°C). For each fly T_{ko} was estimated as the temperature taken to knock it out (defined as the onset of muscle spasms; [79]).

Statistical methods

The experimental setup was devised to assay one male and one female from each cross and temperature per day (five blocks) for both T_p and T_{ko}, amounting to 1,080 flies in total. Some mishaps (e.g. individuals flew away or just died during the assays) were, however, unavoidable and the final data set contains a few more than or a few less than 10 flies in several crosses (the harmonic means of flies per cross and temperature were: T_p assay, 5.04 females and 4.80 males; T_{ko} assay, 4.89 females and 4.37 males). Statistical analysis with and without block design qualitatively yielded the same results. Therefore, to simplify matters blocks were not considered in the linear models below.

a) Consanguinity and temperature effects

Inbreeding and temperature effects were simultaneously analyzed by contrasting isogenic *vs.* outbred homokaryotypic flies reared at both developmental temperatures. The linear model used was:

$$T_{p(ijklmn)} = \mu + \kappa_i + C_{j(i)} + \tau_k + \iota_l + \varsigma_m + \kappa\tau_{ik} + \kappa\iota_{il} + \kappa\varsigma_{im} + \tau\iota_{kl} + \tau\varsigma_{km} + \dots + \iota\varsigma_{lm} + \kappa\tau\iota_{ikl} + \kappa\tau\varsigma_{ikm} + \kappa\iota\varsigma_{ilm} + \tau\iota\varsigma_{klm} + \kappa\tau\iota\varsigma_{iklm} + \varepsilon_{ijklmn}, \quad (1)$$

where μ is the overall grand mean, κ_i is the fixed effect of the karyotype (i =1, 2, 3), $C_{j(i)}$ is the random effect of the jth cross (j = 1, 2, ..., 6) within karyotype i, τ_k is the fixed effect of the developmental temperature (18°C or 22°C), ι_l is the fixed effect of inbreeding (isogenic or outbred homokaryotypic flies), ς_m is the fixed effect of sex, and ε_{ijklmn} is the residual error associated with the thermal preference (T_p) of the nth fly from the mth sex with the ith karyotype from the jth cross that was derived from the ιth group of crosses and assayed at the kth temperature. The covariate plate-hour was also introduced in the model to control for differences in circadian activity since several trials were conducted during each day. A similar linear model was used for knock out temperature, also introducing water-bath as a covariate since T_{ko} was assessed in different water-baths.

Notice that for the main effect "karyotype" the linear model (1) can be conveniently reduced to the following two-level nested ANOVA model:

$$T_{p(ijk)} = \mu + \kappa_i + C_{j(i)} + e_{ijk}, \quad (2)$$

where the sum of squares for the error term e_{ijk} is simply the sum of the sum of squares for the remainder terms in (1). The usefulness of this model reduction is to efficiently perform randomization tests to test the

null hypothesis about karyotype effects in a randomized (i.e., random assignment) experiment [80]. Permutation tests are far less sensitive to the presence of outliers than parametric tests. The null hypothesis of no karyotype effect was tested here after performing random permutations among replicate and selection temperature for the among selection temperature F-statistics. Each test used 10,000 random permutations.

b) Karyotype variation

To asses the effect of O chromosome karyotypes on T_p and T_{ko} we have focused in the outbred crosses, including both structural homo- and heterokaryotypes. The linear model used was similar to (1) including the fixed effect of karyotype (κ_i; i = 1, 2, ..., 6), the random effect of cross within karyotypes ($C_{j(i)}$; j = 1, 2, ..., 6), the fixed effect of developmental temperature, and the fixed effect of sex. The covariate plate-hour was also introduced in the model. As above, a similar linear model was used for knock out temperature, also introducing water-bath as a covariate.

In the original Palaearctic populations chromosome arrangements O_{3+4} and O_{3+4+8} have a higher frequency at lower latitudes than arrangement O_{st}, and the converse is true a higher latitudes [35,36]. For this reason, the variation explained by the six karyotypes was further decomposed after pooling the first two arrangements into a single class (O^*_{3+4}) as follows: between the two O_{st}/O^*_{3+4} heterokaryotypes; among the three O^*_{3+4}/O^*_{3+4} karyotypes; and among O_{st}/O_{st}, O_{st}/O^*_{3+4}, O^*_{3+4}/O^*_{3+4}. The karyotypic values for T_p and T_{ko} were also estimated in the additive-dominance scale [81,82] after pooling the two chromosome arrangements that share O_{3+4} (each comparison or contrast between two means has one degree of freedom).

The genetic correlation between T_p and T_{ko} can be approached as indicated in Betrán et al. [26]. Assuming that the components of the between karyotypes sums of squares and cross-products (SSCP) hypothesis matrix ($\mathbf{H}_k$) are entirely genetic in origin, the correlation coefficient between the means of all six outbred karyotypes is given by:

$$r_k = \frac{\mathbf{H}_k(1,2)}{\sqrt{\mathbf{H}_k(1,1)\mathbf{H}_k(2,2)}}, \tag{3}$$

where $\mathbf{H}_k$ (1, 2) is the off-diagonal element (sum of products of karyotype averages), and $\mathbf{H}_k$ (i, i) is a diagonal element (sum of squares of karyotypes averages) for the ith variable. This correlation coefficient is obviously an approximation to the genetic correlation because the $\mathbf{H}_k$ matrix also contains a fraction of the variation among the isogenic lines used to obtain the outbred flies (see Experimental settings). The correlation coefficient can be tested as:

$$t = r\sqrt{\frac{k-2}{1-r^2}}, \tag{4}$$

where k is the number of karyotypes [83]. After pooling the arrangements that share arrangement O_{3+4} into a single class, we can now obtain the new hypothesis matrix $\mathbf{H}_p$. The correlation coefficient between the pooled averages can be estimated as:

$$r_p = \frac{\mathbf{H}_p(1,2)}{\sqrt{\mathbf{H}_k(1,1)\mathbf{H}_k(2,2)}}. \tag{5}$$

The square of this correlation can be interpreted as that fraction of the total variation among karyotypes that is explained by O_{st}/O_{st}, O_{st}/O^*_{3+4}, O^*_{3+4}/O^*_{3+4}.

c) Computer software for statistical analysis

The computer programs used for statistical data analyses were MATLAB algebra program environment (ver. 7.0.4 [84]) together with the collection of tools supplied by the Statistics Toolbox. The statistical software packages STATISTICA version 9 [85] and SPSS version 15 [86] were also used.

Acknowledgements

We thank three anonymous reviewers for comments on the manuscript. OD is supported by a pre-doctoral fellowship (BES-2007-17438) from the Ministerio de Educación y Ciencia (Spain). CR is supported by a post-doctoral fellowship (SFRH/BPD/39998/2007) from Fundação para a Ciência e a Tecnologia (Portugal). GC is supported by a pre-doctoral fellowship (BES-2007-15096) from the Ministerio de Educación y Ciencia (Spain). ELR is supported by a Ramón y Cajal contract from the Ministerio de Ciencia e Innovación (Spain). This research was supported by grants CGL2006-13423-C01 and CGL2006-13423-C02-02 from the Ministerio de Ciencia y Tecnología (Spain); CGL2009-12912-C03-01, CGL2010-15395 and BFU2009-07564 from the Ministerio de Ciencia e Innovación; and 2009SGR 636 from Generalitat de Catalunya to the Grup de Biologia Evolutiva.

Author details

[1]Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva (GBE), Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain. [2]Departamento de Ciências Agrárias, Azorean Biodiversity Group-CITAA, Universidade dos Açores, Terra-Chã, 9701-851 Angra do Heroísmo, Portugal. [3]Departament de Genètica, Grup de Biologia Evolutiva (GBE), Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain. [4]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, 08071 Barcelona, Spain.

Authors' contributions

OD, GC, JB, MP and MS sampled flies from the natural population and setup isofemale lines. OD and GC obtained the isochromosomal lines. JB helped in gene arrangement identification. GC and MP assayed the isochromosomal lines for microsatellites. OD, CR and GC carried out experimental crosses and egg collections. OD, GC, JB, and MS assayed the flies for thermal preference and knock out temperature. OD, CR, ELR and MS conceived the study. OD, CR, and MS carried out statistical analyses and drafted the manuscript. All authors read and approved the final manuscript.

Received: 12 July 2010 Accepted: 25 November 2010
Published: 25 November 2010

References

1. Angilletta MJ: *Thermal Adaptation. A Theoretical and Empirical Synthesis* Oxford: Oxford Univ Press; 2009.
2. Stevenson RD: **The relative importance of behavioral and physiological adjustments controlling body temperature in terrestrial ectotherms.** *Am Nat* 1985, **126**:362-386.
3. Casey TM, Hegel JR: **Caterpillar setae: insulation for an ectotherm.** *Science* 1981, **214**:1131-1133.
4. Kingsolver JG, Watt WB: **Thermoregulatory strategies in *Colias* butterflies: thermal stress and limits to adaptation in temporally varying environments.** *Am Nat* 1983, **121**:32-35.
5. Willmer PG, Unwin S: **Field analyses of insect heat budget: reflectance, size and heating rates.** *Oecologia* 1981, **50**:250-255.
6. Huey RB, Hertz PE, Sinervo B: **Behavioral drive versus behavioral inertia in evolution: a null model approach.** *Am Nat* 2003, **161**:357-366.
7. Dobzhansky Th: *Genetics of the Evolutionary Process* New York: Columbia Univ Press; 1970.
8. Rodríguez-Trelles F, Alvarez G, Zapata C: **Time-series analysis of seasonal changes of the O inversion polymorphism of *Drosophila subobscura*.** *Genetics* 1996, **142**:179-187.
9. Coyne JA, Bundgaard J, Prout T: **Geographic variation of tolerance to environmental stress in *Drosophila pseudoobscura*.** *Am Nat* 1983, **122**:474-488.
10. Hoffmann AA, Sørensen JG, Loeschcke V: **Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches.** *J Therm Biol* 2003, **28**:175-216.
11. Hoffmann AA: **Physiological climatic limits in *Drosophila*: patterns and implications.** *J Exp Biol* 2010, **213**:870-880.
12. Huey RB, Pascual M: **Partial thermoregulatory compensation along a latitudinal cline in a rapidly evolving invasive species. *Drosophila subobscura* in North America.** *Ecology* 2009, **90**:1715-1720.
13. Sinclair BJ, Vernon P, Klok CJ, Chown SL: **Insects at low temperatures: an ecological perspective.** *Trends Ecol Evol* 2003, **18**:257-262.
14. Hofmann GE, Todgham AE: **Living in the now: physiological mechanisms to tolerate a rapidly changing environment.** *Annu Rev Physiol* 2010, **72**:127-145.
15. Huey RB, Bennett AF: **Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards.** *Evolution* 1987, **41**:1098-1115.
16. Angilletta MJ, Bennett AF, Guderley H, Navas CA, Seebacher F, Wilson RS: **Coadaptation: a unified principle in evolutionary thermal biology.** *Physiol Biochem Zool* 2006, **79**:282-294.
17. Mayr E: *Animal Species and Evolution* Cambridge, Harvard Univ Press; 1963.
18. Ayala FJ, Serra L, Prevosti A: **A grand experiment in evolution *Drosophila subobscura* colonization of the Americas.** *Genome* 1989, **31**:246-255.
19. Hoffmann AA, Weeks AR: **Climatic selection on genes and traits after 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia.** *Genetica* 2007, **129**:133-147.
20. Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L: **Rapid evolution of a geographical cline in size in an introduced fly.** *Science* 2000, **287**:308-309.
21. Balanyà J, Serra L, Gilchrist GW, Huey RB, Pascual M, Mestres F, Solé E: **Evolutionary pace of chromosomal polymorphism in colonizing populations of *Drosophila subobscura*: an evolutionary time series.** *Evolution* 2003, **57**:1837-1845.
22. Rezende EL, Balanyà J, Rodríguez-Trelles F, Rego C, Fragata I, Matos M, Serra L, Santos M: **Climate change and chromosomal inversions in *Drosophila subobscura*.** *Clim Res* 2010, **43**:103-114.
23. Gilchrist G, Jeffers L, West B, Folk D, Suess J, Huey RB: **Clinal patterns of desiccation and starvation resistance in ancestral and invading populations of *Drosophila subobscura*.** *Evol Appl* 2008, **1**:513-23.
24. Calboli FCF, Gilchrist GW, Partridge L: **Different cell size and cell number contribution in two newly established and one ancient body size cline of *Drosophila subobscura*.** *Evolution* 2003, **57**:566-573.
25. Gilchrist GW, Huey RB, Serra L: **Rapid evolution of wing size clines in *Drosophila subobscura*.** *Genetica* 2001, **112-113**:273-286.
26. Betrán E, Santos M, Ruiz A: **Antagonistic pleiotropic effect of second-chromosome inversions on body size and early life-history traits in *Drosophila buzzatii*.** *Evolution* 1998, **52**:144-154.
27. Etterson JR, Shaw RG: **Constraint to adaptive evolution in response to global warming.** *Science* 2001, **294**:151-154.
28. Fontdevila A, Zapata C, Alvarez G, Sanchez L, Méndez J, Enriquez I: **Genetic coadaptation in the chromosomal polymorphism of *Drosophila subobscura*. I. Seasonal changes of gametic disequilibrium in a natural population.** *Genetics* 1983, **105**:935-955.
29. Rodríguez-Trelles F: **Seasonal cycles of allozyme-by-chromosomal-inversion gametic disequilibrium in *Drosophila subobscura*.** *Evolution* 2003, **57**:839-848.
30. Fragata I, Balanyà J, Rego C, Matos M, Rezende EL, Santos M: **Contrasting patterns of phenotypic variation linked to chromosomal inversions in native and colonizing populations of *Drosophila subobscura*.** *J Evol Biol* 2010, **23**:112-123.
31. Nosil P, Crespi BJ, Sandoval CP, Kirkpatrick M: **Migration and the genetic covariance between habitat preference and performance.** *Am Nat* 2006, **167**:E66-E78.
32. Duckworth RE: **The role of behavior in evolution: a search for mechanism.** *Evol Ecol* 2009, **23**:513-531.
33. Dillon ME, Wang G, Garrity PA, Huey RB: **Thermal preference in *Drosophila*.** *J Therm Biol* 2009, **34**:109-119.
34. Rego C, Balanyà J, Fragata I, Matos M, Rezende EL, Santos M: **Clinal patterns of chromosomal inversion polymorphisms in *Drosophila subobscura* are partly associated with thermal preferences and heat stress resistance.** *Evolution* 2010, **64**:385-397.
35. Menozzi P, Krimbas CB: **The inversion polymorphism of *Drosophila subobscura* revisited: synthetic maps of gene arrangement frequencies and their interpretation.** *J Evol Biol* 1992, **5**:625-641.
36. Krimbas CB: **The inversion polymorphism of *Drosophila subobscura*.** In *Drosophila Inversion Polymorphism.* Edited by: Krimbas CB, Powell JR. Boca Raton: CRC Press; 1992:127-220.
37. Sperlich D, Feuerbach-Mravlag H, Lange P, Michaelidis A, Pentzos-Daponte A: **Genetic load and viability distribution in central and marginal populations of *Drosophila subobscura*.** *Genetics* 1977, **86**:835-848.
38. Pegueroles C, Araúz PA, Pascual M, Mestres F: **A recombination survey using microsatellites: the O chromosome of *Drosophila subobscura*.** *Genetica* 2010, **138**:795-804.
39. Powell JR: *Progress and Prospects in Evolutionary Biology. The Drosophila Model* New York: Oxford Univ Press; 1997.
40. Santos J, Serra L, Solé E, Pascual M: **FISH mapping of microsatellite loci from *Drosophila subobscura* and its comparison to related species.** *Chromosome Res* 2010, **18**:213-226.
41. Laayouni H, García-Franco F, Chávez-Sandoval BE, Trotta V, Beltran S, Corominas M, Santos M: **Thermal evolution of gene expression profiles in *Drosophila subobscura*.** *BMC Evol Biol* 2007, **7**:42.
42. Moltó MD, Pascual L, Martínez-Sebastián MJ, de Frutos R: **Genetic analysis of the heat shock response in three *Drosophila* species of the *obscura* group.** *Genome* 1992, **35**:870-880.
43. Cuenca JB, Galindo MI, Saura AO, Sorsa V, de Frutos R: **Ultrastructure of regions containing homologous loci in polytene chromosomes of *Drosophila melanogaster* and *Drosophila subobscura*.** *Chromosoma* 1998, **107**:113-126.
44. Kunze-Mühl E, Müller E: **Weitere Untersuchungen über die chromosomale Struktur und natürlichen Strukturtypen von *D. subobscura*.** *Chromosoma* 1958, **9**:559-570.
45. Parsell DA, Taulien J, Lindquist S: **The role of heat-shock proteins in thermotolerance.** *Phil Trans Roy Soc London Series B* 1993, **339**:279-286.
46. Krebs RA: **A comparison of Hsp70 expression and thermotolerance in adults and larvae of three *Drosophila* species.** *Cell Stress Chaperones* 1999, **4**:243-249.
47. Bettencourt BR, Kim I, Hoffmann AA, Feder ME: **Response to natural and laboratory selection at the *Drosophila hsp70* genes.** *Evolution* 2002, **56**:1796-1801.
48. McColl G, Hoffmann AA, McKechnie SW: **Response of two heat shock genes to selection for knockdown heat resistance in *Drosophila melanogaster*.** *Genetics* 1996, **143**:1615-1627.
49. Santos M, Iriarte PF, Céspedes W, Balanyà J, Fontdevila A, Serra L: **Swift laboratory thermal evolution of wing shape (but not size) in *Drosophila***

subobscura and its relationship with chromosomal inversion polymorphism. *J Evol Biol* 2004, **17**:841-855.
50. Santos M, Céspedes W, Balanyà J, Trotta V, Calboli FCF, Fontdevila A, Serra L: **Temperature-related genetic changes in laboratory populations of *Drosophila subobscura*: evidence against simple climatic-based explanations for latitudinal clines.** *Am Nat* 2005, **165**:258-273.
51. Santos M: **Evolution of total net fitness in thermal lines: *Drosophila subobscura* likes it "warm".** *J Evol Biol* 2007, **20**:2361-2370.
52. Solé E, Balanyà J, Sperlich D, Serra L: **Long-term changes in the chromosomal inversion polymorphism of *Drosophila subobscura*. I. Mediterranean populations from southwestern Europe.** *Evolution* 2002, **56**:830-835.
53. Prevosti A: **Chromosomal inversion polymorphism in the southwestern range of *Drosophila subobscura* distribution area.** *Genetica* 1974, **45**:111-124.
54. Balanyà J, Solé E, Oller JM, Sperlich D, Serra L: **Long-term changes in the chromosomal inversion polymorphism of *Drosophila subobscura*. II. European populations.** *J Zool Syst Evol Res* 2004, **42**:191-201.
55. Munté A, Rozas J, Aguadé M, Segarra C: **Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*.** *Genetics* 2005, **69**:1573-1581.
56. Andolfatto P, Depaulis F, Navarro A: **Inversion polymorphism and nucleotide variability in *Drosophila*.** *Genet Res* 2001, **77**:1-8.
57. Laayouni H, Hasson E, Santos M, Fontdevila A: **Inversion polymorphism and nucleotide variability in different regions of the second-chromosome in *Drosophila buzzatii*.** *Mol Biol Evol* 2003, **20**:931-944.
58. Santos M, Iriarte PF, Cespedes W: **Genetics and geometry of canalization and developmental stability in *Drosophila subobscura*.** *BMC Evol Biol* 2005, **5**:7.
59. Yamamoto AH: **Temperature preference of *Drosophila immigrans* and *D. virilis*: Intra- and inter-population genetic variation.** *Jpn J Genet* 1994, **69**:67-76.
60. Hertz PE, Huey RB, Stevenson RD: **Evaluating temperature regulation by field-active ectotherms: the fallacy of the inappropriate question.** *Am Nat* 1993, **142**:796-818.
61. Falconer DS, Mackay TF: *Introduction to Quantitative Genetics.* 4 edition. Essex, England: Longman; 1996.
62. Balanyà J, Oller JM, Huey RB, Gilchrist GW, Serra L: **Global genetic change tracks global climate warming in *Drosophila subobscura*.** *Science* 2006, **313**:1773-1775.
63. Sørensen JG, Kristensen TN, Loeschcke V: **The evolutionary and ecological role of heat shock proteins.** *Ecol Lett* 2003, **6**:1025-1037.
64. Jensen LT, Cockerell FE, Kristensen TN, Rako L, Loeschcke V, McKechnie SW, Hoffmann AA: **Adult heat tolerance variation in *Drosophila melanogaster* is not related with Hsp70 expression.** *J Exp Zool* 2010, **313A**:35-44.
65. Feder ME, Blair N, Figueras H: **Oviposition site selection: unresponsiveness of *Drosophila* to cues of potential thermal stress.** *Anim Behav* 1997, **53**:585-588.
66. Chown SL, Jumbam KR, Sørensen JG, Terblanche JS: **Phenotypic variance, plasticity and heritability estimates of critical thermal limits depend on methodological context.** *Funct Ecol* 2009, **23**:133-140.
67. Rezende EL, Tejedo M, Santos M: **Estimating the adaptive potential of critical thermal limits: methodological problems and evolutionary implications.** *Funct Ecol* 2010.
68. Mitchell KA, Hoffmann AA: **Thermal ramping rate influences evolutionary potential and species differences for upper thermal limits in *Drosophila*.** *Funct Ecol* 2010, **24**:694-700.
69. Moreteau B, Morin J-P, Gibert P, Petavy G, Pla E, David JR: **Evolutionary changes of nonlinear reaction norms according to thermal adaptation: a comparison of two *Drosophila* species.** *C R Acad Sci Paris* 1997, **320**:833-841.
70. Santos M, Brites D, Laayouni L: **Thermal evolution of pre-adult life history traits, geometric size and shape, and developmental stability in *Drosophila subobscura*.** *J Evol Biol* 2006, **19**:2006-2021.
71. Bubli OL, Imasheva AG, Loeschcke V: **Selection for knockdown resistance to heat in *Drosophila melanogaster* at high and low larval densities.** *Evolution* 1998, **52**:619-625.
72. Kari JS, Huey RB: **Size and seasonal temperature in free-ranging *Drosophila subobscura*.** *J Therm Biol* 2000, **25**:267-272.
73. Sørensen JG, Loeschcke V: **Larval crowding in *Drosophila melanogaster* induces Hsp70 expression, and leads to increased adult longevity and adult thermal stress resistance.** *J Insect Physiol* 2001, **47**:1301-1307.
74. Savković V, Stamenkovic-Radak M, Andjelkovic M: **Diurnal variability of gene arrangement frequencies in *Drosophila subobscura* populations from two habitats.** *J Zool Syst Evol Res* 2004, **42**:208-214.
75. Hedick PW: *Genetics of Populations.* 2 edition. Boston: Jones and Bartlett Publ; 2000.
76. de Frutos R, Prevosti A: **Temporal changes of chromosomal polymorphism in natural populations of *Drosophila subobscura*.** *Genetica* 1984, **63**:181-187.
77. Iriarte FP, Céspedes W, Santos M: **Quantitative-genetic analysis of wing form and bilateral asymmetry in isochromosomal lines of *Drosophila subobscura* using the Procrustes methods.** *J Genet* 2003, **82**:95-113.
78. Mestres F, Pegueroles G, Prevosti A, Serra L: **Colonization of America by *Drosophila subobscura*: lethal genes and the problem of the O5 inversion.** *Evolution* 1990, **44**:1823-1836.
79. Lutterschmidt WI, Hutchinson VH: **The critical thermal maximum: data to support the onset of spasms as the definitive end point.** *Can J Zool* 1997, **75**:1553-1560.
80. Edgington ES: *Randomization Tests.* 3 edition. New York: Marcel Dekker; 1995.
81. Mather K, Jinks JL: *Introduction to Biometrical Genetics* London: Chapman and Hall; 1977.
82. Mather K, Jinks JL: *Biometrical Genetics: The Study of Continuous Variation.* 3 edition. London: Chapman and Hall; 1982.
83. Sokal RR, Rohlf FJ: *Biometry.* 3 edition. New York: Freeman; 1995.
84. MathWorks Inc: **Matlab, V. 7.5.0.342 (R2007b).** *The Language of Technical Computing* 2007 [http://www.mathworks.com].
85. StatSoft Inc: **STATISTICA (*data analysis software system*), version 9.** 2009 [http://www.statsoft.com].
86. SPSS Inc: *SPSS for Windows* 2006 [http://www.spss.com].

doi:10.1186/1471-2148-10-363
Cite this article as: Dolgova *et al.*: Genetic constraints for thermal coadaptation in *Drosophila subobscura*. *BMC Evolutionary Biology* 2010 **10**:363.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

Appendix II

JOURNAL OF Evolutionary Biology

doi: 10.1111/j.1420-9101.2012.02463.x

Hsp70 protein levels and thermotolerance in *Drosophila subobscura*: a reassessment of the thermal co-adaptation hypothesis

G. CALABRIA*, O. DOLGOVA†, C. REGO‡, L. E. CASTAÑEDA†§, E. L. REZENDE†, J. BALANYÀ*, M. PASCUAL*, J. G. SØRENSEN¶, V. LOESCHCKE¶ & M. SANTOS†

*Departament de Genètica, Grup de Biologia Evolutiva (GBE)/Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain
†Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva (GBE), Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain
‡Departamento de Ciências Agrárias, Azorean Biodiversity Group-CITAA, Universidade dos Açores, Terra-Chã, Angra do Heroísmo, Portugal
§Instituto de Ecología y Evolución, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile
¶Department of Bioscience, Aarhus University, Aarhus, Denmark

Keywords:

Drosophila;
heat tolerance;
Hsp70 expression;
latitudinal clines;
thermal co-adaptation.

Abstract

Theory predicts that geographic variation in traits and genes associated with climatic adaptation may be initially driven by the correlated evolution of thermal preference and thermal sensitivity. This assumes that an organism's preferred body temperature corresponds with the thermal optimum in which performance is maximized; hence, shifts in thermal preferences affect the subsequent evolution of thermal-related traits. *Drosophila subobscura* evolved worldwide latitudinal clines in several traits including chromosome inversion frequencies, with some polymorphic inversions being apparently associated with thermal preference and thermal tolerance. Here we show that flies carrying the warm-climate chromosome arrangement O_{3+4} have higher basal protein levels of Hsp70 than their cold-climate O_{st} counterparts, but this difference disappears after heat hardening. O_{3+4} carriers are also more heat tolerant, although it is difficult to conclude from our results that this is causally linked to their higher basal levels of Hsp70. The observed patterns are consistent with the thermal co-adaptation hypothesis and suggest that the interplay between behaviour and physiology underlies latitudinal and seasonal shifts in inversion frequencies.

Introduction

Thermal acclimation is a plastic response that occurs within the lifetime of individuals and triggers a variety of physiological adjustments. This includes the alteration in gene expression and increased heat-shock resistance by prior short-term exposition to sublethal high temperatures ('hardening'; Hoffmann *et al.*, 2003). The heat-induced heat-shock proteins Hsp70s have for a long time being considered essential for heat stress survival (e.g. Parsell *et al.*, 1993; Sørensen *et al.*, 2003). Hsp70s are a highly conserved chaperone family that assist in protein folding and mitigate cellular damage during thermal stress, although apparently with a fitness cost to the organism (Krebs & Loeschcke, 1994; Krebs & Feder, 1997a, 1998; Feder *et al.*, 2002). This cost may impose a trade-off to the maximum attained levels of Hsp70, which in *Drosophila* might extend throughout the life cycle because Hsp70 expression seems to be genetically coupled at larval and adult stages (Krebs *et al.*, 1998; but see Sarup *et al.*, 2006).

However, the actual role played by Hsp70 in the heat tolerance of *Drosophila* is proving to be a controversial issue. Some studies where manipulation of Hsp70 levels was achieved by using transgenic *Drosophila melanogaster* larvae have shown that individuals producing more protein improved thermotolerance under laboratory

Correspondence: Gemma Calabria, Departament de Genètica, Grup de Biologia Evolutiva (GBE), Facultat de Biologia, Universitat de Barcelona, Diagonal 643, 08028 Barcelona, Spain.
Tel.: +34 93 493 5801; fax: +34 93 403 4420;
e-mail: gemmacalabria@ub.edu

(Krebs & Feder, 1998; Bettencourt *et al.*, 2008) and field-like conditions (Roberts & Feder, 2000); furthermore, standing natural variation of Hsp70 expression also correlates positively with larval thermotolerance (Krebs & Feder, 1997b). On the other hand, weak or no associations between Hsp70 levels and adult thermotolerance have been detected (Dahlgaard *et al.*, 1998; Krebs *et al.*, 1998; Jensen *et al.*, 2010). These contrasting outcomes are somewhat odd in view of the correlated response for increased Hsp70 levels in *Drosophila* lines selected for adult heat tolerance (Sørensen *et al.*, 1999; Folk *et al.*, 2006), which suggests that these traits are pleiotropically linked.

We have recently uncovered that segregating chromosome arrangements in *Drosophila subobscura* are associated with differential adult thermal preferences and heat tolerances (Rego *et al.*, 2010). *Drosophila subobscura* is one of the species with the richest inversion polymorphism in the genus. Compelling evidence that inversions' latitudinal clines are adaptive in this species is that within a decade after its invasion of South and North America from native European populations, flies evolved latitudinal clines that were generally parallel with clines in the Old World (Prevosti *et al.*, 1985, 1988; Balanyà *et al.*, 2003). Further evidence comes from the long-term shifts in inversion frequencies indicating that the genetic constitution of *D. subobscura* populations worldwide is responding to climate change (Balanyà *et al.*, 2006). These results suggest that directional selection must be strong, and temperature is a likely selective agent. Consistent with this assumption, we found that flies carrying 'cold-climate' chromosome arrangements (i.e. those chromosome arrangements that show a negative correlation coefficient with maximum temperatures along the cline, or a positive correlation coefficient with latitude; Balanyà *et al.*, 2003) prefer lower temperatures and have lower heat tolerances than their 'warm-climate' counterparts (Rego *et al.*, 2010). These findings stimulated our assessment of the thermal co-adaptation hypothesis (Dolgova *et al.*, 2010); that is, the idea that behaviour (thermal preference) 'drives' evolution in new directions and can speed up physiological (thermotolerance) adaptation (Huey & Bennett, 1987; Angilletta, 2009).

To test the thermal co-adaptation hypothesis, we focused on the largest and inversion richest chromosome O of *D. subobscura* (Dolgova *et al.*, 2010) because (i) its complex gene arrangements from more equatorial populations – with overlapping inversions in the distal (segment I) and nonoverlapping inversions in the proximal (segment II) segment to the centromere (Krimbas & Loukas, 1980) – are gradually replaced by the standard gene arrangement (O_{st}) as populations approach high latitudes, (ii) this chromosome is associated with thermal preferences (Rego *et al.*, 2010) and harbours several genes involved in the heat-shock response, including the gene *Hsp70* located inside the region covered by the overlapping inversions 3+4 (Moltó *et al.*, 1992) (Fig. 1), and (iii) warm- (O_{3+4}) and cold-climate (O_{st}) chromosome arrangements are highly differentiated for their genetic content in the 3+4 region (Munté *et al.*, 2005). Therefore, we tested whether these gene arrangements could host a set of allelic combinations for thermal preference and thermotolerance that would generate a positive genetic correlation between these two traits in line with the predictions from the thermal co-adaptation hypothesis. Our expectation was that flies' preferred body temperatures and heat tolerances according to their genetic constitution for chromosome O should agree with the clinal patterns. Although we were able to replicate Rego *et al.*'s (2010) results showing that O_{3+4} carriers display a preference for higher temperatures than their O_{st} counterparts, no difference between these arrangements was detected for heat tolerance (Dolgova *et al.*, 2010). Consequently, we concluded that any correlation between these traits across latitudinal gradients would presumably reflect a pattern of correlated selection rather than genetic correlation (Dolgova *et al.*, 2010).

Considering that inversion polymorphisms in other chromosomes associated with thermotolerance exhibit latitudinal variation in the expected direction (Rego *et al.*, 2010), it is unclear why different arrangements in chromosome O seem to have negligible effects on this trait in our flies. There are at least three possible explanations. First, it might be the case that there is not enough genetic differentiation for *Hsp*70 between O_{3+4} and O_{st} chromosome arrangements. Second, assuming that there is indeed genetic differentiation it might happen that there is no association between Hsp70 protein levels and adult thermotolerance in *D. subobscura*, in accordance with what has been recently suggested to occur in *D. melanogaster* (Jensen *et al.*, 2010). Third, flies carrying different arrangements may differ in their plasticity to heat shock, resulting in contrasting patterns

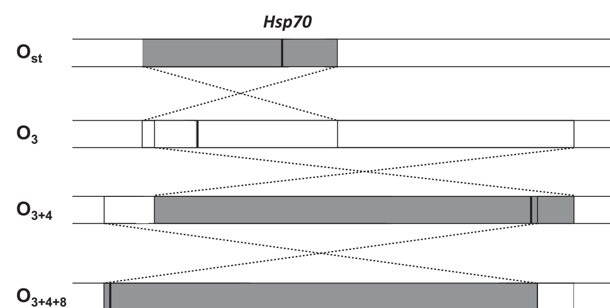


Fig. 1 Schematic representation of segment I of chromosome O in *Drosophila subobscura* showing the relative position of the gene *Hsp70* (thick black line) in the three chromosome arrangements used in this study (O_{st}, O_{3+4} and O_{3+4+8}). O_3 is the now extinct original chromosome arrangement from which O_{st} and O_{3+4} derived. Shaded bars indicate the inverted regions.

of basal and induced heat tolerance (see Rezende *et al.*, 2011). Consequently, the initial assessment of the thermal co-adaptation hypothesis could be somewhat flawed if the potential fitness benefits of heat-induced thermotolerance under extreme field conditions (Loeschcke & Hoffmann, 2007) were not appropriately assessed in the experiments performed by Dolgova *et al.* (2010).

In this study, we focus on the latter two alternatives. We first show that warm- and cold-climate O chromosome arrangements in *D. subobscura* are indeed differentiated for basal Hsp70 protein expression levels. We then estimate basal and induced thermotolerance using two different dynamic assays, because measured heat tolerance is known to be affected by variable heating rates (Terblanche *et al.*, 2007; Chown *et al.*, 2009; for a mechanistic explanation of why this may happen, see Rezende *et al.*, 2011; Santos *et al.*, 2011). We show that thermal preference and thermotolerance are indeed coupled as predicted from the thermal co-adaptation hypothesis and discuss the confounding factors that have influenced the outcome of the heat tolerance assays in our previous experiments (Dolgova *et al.*, 2010). Finally, we discuss our results in connection with the latitudinal clines and seasonal variation in the frequencies of chromosome O gene arrangements.

Materials and methods

Base stocks and fly handling

The origin and establishment of the 18 isochromosomal lines for the O chromosome in an otherwise highly homogeneous genetic background used in this study have been described elsewhere (Dolgova *et al.*, 2010). Briefly, more than 200 isofemale lines were established from wild flies collected near Barcelona (41 43′N, 2 13′E) in October 2007. From each isofemale line, we isolated a single O chromosome by repeatedly backcrossing (for at least five generations) a wild-type male with females from the *ch-cu* strain, which carries the recessive markers on the O chromosome *cherry* eyes (*ch*) and *curled* wings (*cu*) and has a highly homogeneous genetic background. The isochromosomal lines were then obtained using the *Va/Ba* (*Varicose/Bare*) balancer stock, which has the same genetic background than the *ch-cu* strain. The final isochromosomal lines used in the experiments were the same than the ones used by Dolgova *et al.* (2010) and included six independent lines for each of the three chromosome arrangements O_{st}, O_{3+4}, and O_{3+4+8}.

Most frequent chromosome arrangements in Southern Europe carry the inversions O_{3+4} in segment I and are gradually replaced by arrangement O_{st} as we move towards the north: the combined frequency of O_{st} and arrangements carrying 3+4 is generally higher than 70% (Solé *et al.*, 2002; Balanyà *et al.*, 2004). Our main concern here is therefore to test for putative differences in Hsp70 protein expression levels, as well as in basal and inducible heat tolerances between chromosome arrangements O_{3+4} and O_{st}. Arrangement O_{3+4+8} was also included in the study because (i) the gene *Hsp70* changed its chromosome position relative to the centromere when inversion O_8 arose on a O_{3+4} chromosome arrangement (Fig. 1), and (ii) it allows us to asses the extent of among-arrangement genetic differentiation for Hsp70 protein expression levels given the reduced recombination rate in the region covered by these arrangements (Pegueroles *et al.*, 2010).

Because outbred individuals are obviously more representative of field conditions than isochromosomal lines, the homokaryotypic O_{3+4}/O_{3+4}, O_{3+4+8}/O_{3+4+8} and O_{st}/O_{st} experimental flies were derived from six cyclically permuted reciprocal crosses among the six isochromosomal lines per chromosomal type (i.e. 1×2, 2×3, , 6×1). It is however important to notice that the outbred progeny within each cross is genetically identical, and therefore, the within-cross variation only reflects environmental variance. Flies were always reared at 18 C, and all fly handling was carried out at room temperature using CO_2 anaesthesia only to sort virgin flies.

Thermal exposure conditions

Preliminary experiments using two randomly selected outbred crosses were performed to pinpoint the stressful high temperature most appropriate for the heat hardening treatment to induce the heat-shock response. Heat knockdown times were measured in 30 one-week-old nonvirgin females from each cross. The flies were placed individually in 5-mL glass vials and exposed acutely to 38 or 36 C by immersion in preheated water baths with capacity for 60 vials each. Water temperature was controlled by a programmable heating unit (JULABO ED; JULABO Labortechnik GmbH, Seelbach, Germany) that also ensured proper water circulation. To minimize measurement error, the flies were video-recorded during the assays with a digital HD video camera (SONY HDR-CX110E, Tokyo, Japan), and knockdown times were estimated from the videos. The results (Fig. 2) clearly suggest that 38 C is close to the critical thermal maximum (CT_{max}) because all flies were knocked down after 4.8 min (the time required for 50% of the flies to succumb to heat stress was LT_{50} = 2 min). Flies were able to withstand 36 C for up to 16 min (LT_{50} = 9.7 min).

To induce thermotolerance, the experimental flies from each cross were transferred in groups of ten flies (1-week-old nonvirgin females or males) to empty vials (2×8 cm) with moistened stoppers to avoid desiccation. Because pre-exposure at 5–6 C lower than CT_{max} is usually enough to elicit heat hardening in *Drosophila* (see Hoffmann *et al.*, 2003), the vials were allocated and evenly spaced in a rack and immersed in a water bath at 32 C for 1 h (no fly died after this exposure). The flies

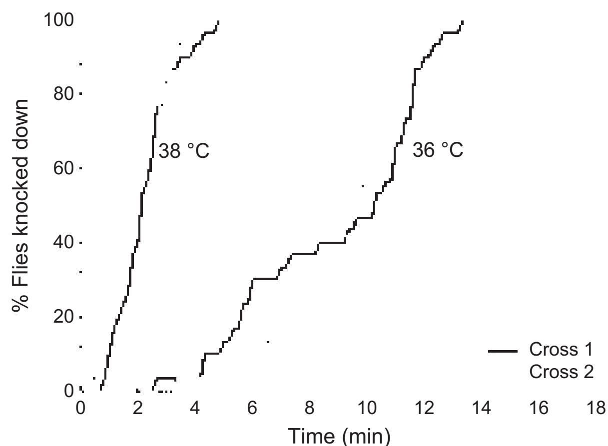


Fig. 2 Knockdown times of genetically homogeneous *Drosophila subobscura* females acutely placed at 36 or 38 C and derived from two crosses using different isochromosomal lines. Each sigmoid-like line was obtained from 30 one-week-old females and also illustrates the stochastic effect associated with the cumulative probability of dying after a finite exposure to a stressful high temperature.

were then allowed to recover normal mobility at 18 C for 2 h to measure Hsp70 protein induction, or for 6 h prior to the heat knockdown assays performed (see below). The heat-shock response requires *de novo* synthesis of proteins, and a recovery period following high-temperature treatment is often necessary to induce rapid heat hardening (Hoffmann *et al.*, 2003; Chown & Nicolson, 2004).

Hsp70 protein expression

Following the heat-shock treatment and the 2-h recovery time, flies were placed in Eppendorf tubes, frozen in liquid nitrogen and stored at −80 C until the level of Hsp70 protein expression was assayed. Control flies from both sexes for each cross kept at 18 C were also frozen to quantify the basal level of Hsp70.

Flies from each replicated group were homogenized in 1.0 mL of homogenizing mix (Sigma-Aldrich Co., St Louis, MO). Lysates were centrifuged for 30 min at 13 226 *g* at 4 C, and the supernatant was transferred to two 0.5-mL microtubes and stored on ice. The total protein concentration in the supernatant was determined from quadruplicate 10-μL samples in a 96-well microplate with a BCA assay (Sigma-Aldrich Co.) according to the manufacturer's instructions and standardized against bovine serum albumin diluted in sterile distilled water.

Hsp70 concentrations from whole animal lysates were quantified with an enzyme-linked immunosorbent assay (ELISA) with an Hsp70-specific monoclonal primary antibody (clone 5A5, dilution 1:1000 PBS; Thermo Scientific Inc., Bremen, Germany) and a HRP-conjugated secondary antibody (anti-mouse IgG; ThermoScientific) following the ELISA protocol described in Sørensen *et al.* (1999). Colour reaction was measured with a spectophotometric microplate reader (PowerWave XS2 Biotek, Bad Friedrichshall, Germany) at 490 nm. Each plate contained 23 samples (heat-induced or control, randomly distributed) and a standard (see below) in four replicates. A blank (without primary antibody) of each sample and standard was included to allow corrections for a nonspecific signal.

The Hsp70 standard was derived from all different crosses after mixing flies to set up a mass population from which two groups of 45 individuals each were hardened as previously described. Data were standardized before statistical analysis. One randomly selected ELISA plate was used as the reference plate, and the standard value of this plate was used to obtain a correction factor for the rest.

Adult heat tolerance assays

Basal and hardened heat tolerances were measured on six females and six males from each of the performed homokaryotype crosses involving the O_{3+4} and O_{st} isochromosomal lines. Because the water baths have capacity for sixty 5-mL glass vials, only five cyclically permuted reciprocal crosses within each chromosomal type were assayed. Three females and three males from each cross were individually placed in sealed empty vials and immersed in two water baths for each treatment.

Heat knockdown temperature was scored as the temperature for individual flies to be knocked down and immobilized in a dynamic assay, where temperature is increased at a constant rate until the end point is observed. Estimates of heat knockdown temperature can however vary dramatically depending on the methodology employed (Terblanche *et al.*, 2007; Chown *et al.*, 2009), which may be partially explained as a result of the varying levels of physiological stress, dehydration and so on experienced by the flies when enclosed in sealed empty vials immersed in a water bath for different periods of time (Rezende *et al.*, 2011; Santos *et al.*, 2011). We therefore used two different experimental protocols to measure basal and hardened heat tolerance: a slow ramping protocol with initial temperature T_0 = 24 C and heating rate ΔT = 0.1 C min^{-1} as in Dolgova *et al.* (2010), and a fast ramping assay with T_0 = 24 C and ΔT = 0.6 C min^{-1}. In both cases, an equilibration time of 10 min was allowed before increasing temperature.

Measurement error was of some concern in the fast ramping assays because the water baths would reach the estimated acute CT_{max} of 38 C (see above) in approximately 23 min. We therefore placed a thermocouple in the water baths, video-recorded the flies and the temperature displayed by the thermocouple (nearest 0.1 C), and estimated the heat knockdown temperature of each fly from the time point at which it collapsed

using the linear interpolation $T_k = T_i + \Delta T \times t_k$. In this expression, T_k is the estimated knockdown temperature, T_i is the temperature of the thermocouple at time t_i just before the fly was knocked down, and t_k is the recorded time when the fly was knocked down. The videos were analysed twice (each time by a different observer), and measurement error was estimated as one minus the intraclass correlation coefficient obtained from the one-way ANOVA (Sokal & Rohlf, 1995, p. 213).

Statistical analyses

Levene's test was used to test for homogeneity of variances. Levels of Hsp70 protein were analysed using the linear model.

$$y_{ijklm} = \mu + K_i + c_{j(i)} + S_k + T_l + KS_{ik} + KT_{il} + ST_{kl} + KST_{ikl} + \varepsilon_{ijklm},$$

Where μ is the overall grand mean, K_i is the fixed effect of the ith karyotype (O_{3+4}/O_{3+4}, O_{3+4+8}/O_{3+4+8} and O_{st}/O_{st}), $c_{j(i)}$ is the random effect of the jth cross ($j = 1,2,...,6$) within karyotype i, S_k is the fixed effect of sex, T_l is the fixed effect of treatment (basal and heat-induced) and ϵ_{ijklm} is the residual error. Hsp70 levels between the warm- (O_{3+4}/O_{3+4}) and cold-climate (O_{st}/O_{st}) homokaryotypes for those gene arrangements that show more significant opposite latitudinal clines (Balanyà *et al.*, 2003) were compared by means of linear contrasts (each comparison or contrast between two means has one degree of freedom).

A similar linear model was used to analyse heat tolerance. Because the variance in knockdown temperature was significantly higher under slow ramping (Levene's $F_{1,472} = 23.40$, $P < 0.001$), the data from the different dynamic assays were analysed separately. An additional reason to perform separate analyses is that the confounding sources of error to reliably estimate heat tolerance are more important in longer assays (i.e. under slow ramping; Rezende *et al.*, 2011; Santos *et al.*, 2011). Analyses were performed using the statistical packages STATISTICA 9.0 (StatSoft, Tulsa, OK, USA) and SPSS 15.0 (SPSS, Chicago, IL, USA).

Results

Induction of Hsp70 protein levels after the heat shock

Global protein levels of Hsp70 increased after the heat-shock treatment as expected, with no significant sex or sex × treatment interaction effects (Table 1). Overall differences among the three homokaryotypes were also detected, and it was clear from the linear contrast that the warm-climate homokaryotype O_{3+4}/O_{3+4} had higher Hsp70 protein levels on average (around 5% higher) than its cold-climate O_{st}/O_{st} counterpart. However, an interesting and somewhat unexpected karyotype × treatment interaction effect was also observed (Table 1, Fig. 3). Thus, the comparatively high basal protein level of Hsp70 in O_{3+4}/O_{3+4} homokaryotypes did not increase after the heat shock, whereas it boosted by around 30% in both O_{3+4+8}/O_{3+4+8} and O_{st}/O_{st} flies. Following heat induction, the three homokaryotypes reached similar Hsp70 protein levels (Fig. 3). The linear contrast comparing O_{3+4}/O_{3+4} vs. O_{st}/O_{st} for the karyotype × treatment interaction effect clearly indicates that these warm- and cold-climate homokaryotypes behave quite differently for Hsp70 protein levels according to basal or heat-induced conditions, whereas O_{3+4+8}/O_{3+4+8} has similar behaviour than O_{st}/O_{st}.

Table 1 ANOVA for the effects of karyotype (O_{3+4}/O_{3+4}, O_{3+4+8}/O_{3+4+8} and O_{st}/O_{st}), sex and treatment (basal and heat-induced) on Hsp70 protein expression levels (data plotted in Fig. 1) (⊂ means 'nested in').

Source of variation	d.f.	Mean square	*F*	*P*
Karyotype (*K*)	2	0.325	5.39	0.010
O_{3+4}/O_{3+4} vs. O_{st}/O_{st}	1	0.323	5.35	0.035
Cross ⊂ *K*	15	0.059	0.95	0.516
Sex (*S*)	1	0.035	0.55	0.460
Treatment (*T*)	1	0.388	6.17	0.015
K × *S*	2	0.024	0.38	0.687
K × *T*	2	0.359	5.72	0.004
O_{3+4}/O_{3+4} vs. O_{st}/O_{st}	1	0.426	6.79	0.010
S × *T*	1	0.018	0.29	0.590
K × *S* × *T*	2	0.046	0.74	0.481
Error	108	0.063		

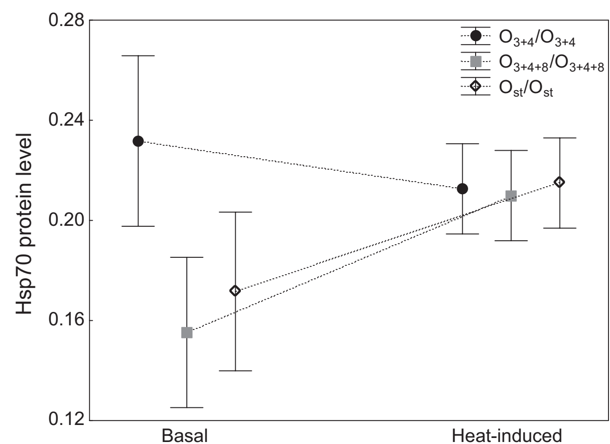


Fig. 3 Hsp70 expression (optical density; standard values used for correction) in adult *Drosophila subobscura* (females and males combined). The error bars are 95% confidence intervals around the mean of outbred flies from the three O chromosome homokaryotypes derived from six independent isochromosomal lines each (see text for details). The heat-induced Hsp70 expression levels were obtained after treating the flies with a heat shock of 32 C for 1 h and recovery time at 18 C for 2 h.

© 2012 THE AUTHORS. *J. EVOL. BIOL.* 25 (2012) 691–700
JOURNAL OF EVOLUTIONARY BIOLOGY © 2012 EUROPEAN SOCIETY FOR EVOLUTIONARY BIOLOGY

Basal and hardened heat tolerance

Results for the adult heat tolerance assays are plotted in Fig. 4. Measurement error in the fast ramping protocol, estimated as one minus the intraclass correlation coefficient between the two knockdown temperatures calculated from the recorded videos, was 18%. The qualitative conclusions remained the same for each estimated knockdown temperature, and we used the average of the two values as our measure of heat tolerance.

The average knockdown temperature with slow ramping (mean ± SD: 35.07 ± 1.57 C) was substantially lower than with fast ramping (37.59 ± 1.19 C). This is in accordance with what had been observed by Chown *et al.* (2009) for *D. melanogaster* and is also expected from our theoretical treatment of ramping effects on heat tolerance (Rezende *et al.*, 2011; Santos *et al.*, 2011). Pearson correlation coefficients between knockdown temperatures across ramping conditions, using the pooled (basal and hardened separately) within-cross average knockdown temperature as the relevant variables (recall that that the outbred progeny within each cross is genetically identical), were not statistically different from zero in any case (Table 2).

An increase in heat tolerance following heat hardening was observed with slow ramping but not with fast ramping (Table 3). Conversely, differences between karyotypes were only detected with fast ramping: O_{3+4}/O_{3+4} flies had a higher knockdown temperature than their O_{st}/O_{st} counterparts (37.77 vs. 37.41 C). Yet, the marginally significant karyotype × sex interaction effect under slow ramping indicates that O_{3+4}/O_{3+4} females also had a higher knockdown temperature than O_{st}/O_{st} females, whereas males had about the same heat tolerance regardless karyotype. Also important is that in no case was the karyotype × treatment (basal and hardened heat tolerance) interaction effect statistically significant.

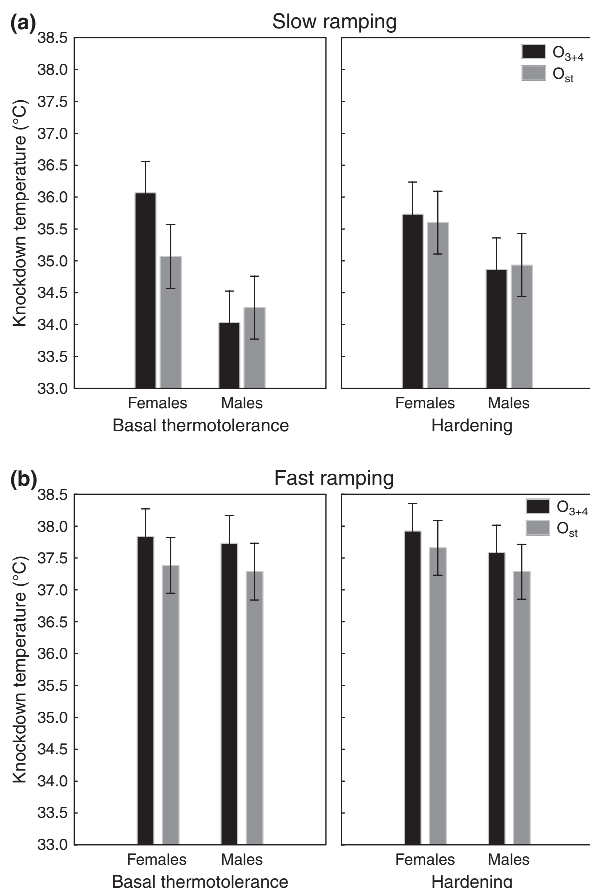


Fig. 4 Knockdown temperatures in nonhardened and hardened flies carrying warm- (O_{3+4}/O_{3+4}) and cold-climate (O_{st}/O_{st}) homokaryotypes in *Drosophila subobscura* under (a) slow ramping (ΔT = 0.1 C min^{-1}) and (b) fast ramping (ΔT = 0.6 C min^{-1}). Error bars are 95% confidence intervals.

Discussion

The striking observations on differences between warm- (O_{3+4}) and cold-climate (O_{st}) chromosome arrangements arising from this study are (i) the sustained higher protein levels of Hsp70 in O_{3+4}/O_{3+4} homokaryotypes under basal conditions, (ii) the higher basal thermotolerance of O_{3+4}/O_{3+4} females under slow ramping and (iii) the slightly higher basal and heat-induced thermotolerance of both O_{3+4}/O_{3+4} females and males when compared to their O_{st}/O_{st} counterparts under fast ramping. The lack of induction of Hsp70 in O_{3+4}/O_{3+4} flies might be due to a higher temperature threshold for induction than that required for O_{st}/O_{st} and O_{3+4+8}/O_{3+4+8} flies (Fig. 3). We however find this explanation unlikely because the stressful temperature used for heat hardening is about 10 C above the upper bound of the 'set point' (T_{set}) range (central 50% of preferred body temperatures; Hertz *et al.*, 1993) for the warm-climate homokaryotype O_{3+4}/O_{3+4} (T_{set} = 16.6–22.2 C; Dolgova *et al.*, 2010) and only 3.2 C below its average knockdown temperature under slow ramping. It seems therefore reasonable to assume that our thermal exposure conditions (see Materials and methods) should have triggered rapid heat hardening induction in all homokaryotypes tested in the experiments.

Overproduction of Hsp70 at normal temperatures has deleterious consequences on *Drosophila* cells (Feder *et al.*, 1992) and adversely impacts larval development (Krebs & Feder, 1997a) and female fecundity (Krebs & Loeschcke, 1994). These studies have led to the idea that the existing copy number of *Hsp70* in *D. melanogaster* (six copies of *Hsp70* per haploid genome; Gong & Golic, 2004) represents a balance between selection for its chaperone function under stress conditions and against its deleterious effects on fitness (Feder & Hofmann, 1999). As stated by Sørensen (2010, p. 705), 'it is not always clear when the level of constitutive and inducible heat-shock protein

Table 2 Pearson correlation coefficients between the estimated knockdown temperatures from the slow ($\Delta T = 0.1$ C min^{-1}) and fast ($\Delta T = 0.6$ C min^{-1}) ramping assays.

	Basal	Hardened
Females	0.6412	0.2994
Males	0.0244	0.5197

The pooled within-cross average knockdown temperatures were used as the relevant variables. In no case was the correlation coefficient statistically different from zero after a sequential Bonferroni correction.

(Hsp) expression should be interpreted as reflecting the capacity or ability to mount a strong defense (i.e. as a benefit) or when it should be interpreted as reflecting the need to mount a strong response as the organism is stressed (i.e. as a cost)'. If this cost-benefit reasoning applies to *D. subobscura,* the question immediately arises: Why are the basal protein levels of Hsp70 in O_{3+4}/O_{3+4} homokaryotypes about the same than those attained by O_{st}/O_{st} flies after heat hardening induction? The seemingly obvious explanation is that Southern European populations, where the frequency of chromosome arrangement O_{3+4} is relatively high (Solé *et al.*, 2002), are more likely exposed to occasionally high temperatures than populations located at higher latitudes, and the benefits of mounting the stress response might be outweighed by the costs. This is in accordance with active flies having much lower body temperature in cooler seasons and at higher latitudes (Huey & Pascual, 2009). Interestingly, the opposite seasonal fluctuations between O_{st} and O_{3+4} carrying flies observed in an extensively studied population, where O_{3+4} consistently increases in summer and decreases in winter (Rodr´guez-Trelles *et al.*, 1996), also seems to favour this interpretation. O_{3+4+8} has traditionally being considered a 'warm-climate' chromosome arrangement although its negative correlation with latitude in Palaearctic populations was mainly due to its high frequency in Northern Africa (Menozzi & Krimbas, 1992). However, no clines were observed for this arrangement in the colonized area (Balanyà *et al.*, 2003) and recent studies show that O_{3+4+8} is dramatically increasing its frequency in Northern European populations (Rezende *et al.*, 2010), which raises some caveats on the putative status of O_{3+4+8} as a warm-climate gene arrangement. Thus, the contrasting pattern of O_{3+4} in front of the other two arrangements could be due to differences in the genetic content of this warm adapted arrangement. Two hypothesis could explain the higher levels of Hsp70 expression in the O_{3+4} arrangement. On the one hand, basal thermotolerance has been found to be positively associated with hsp70 copy number in *D. melanogaster* (Bettencourt *et al.*, 2008); thus, in *D. subobscura* different copy number of the hsp70 genes in the different arrangements, with a higher number in the O_{3+4} arrangement, could explain the results encountered. On the other hand variation in the cis-regulatory region could lead to differences in the genetic expression of this gene as found in *D. melanogaster* (Bettencourt *et al.*, 2002). However, further analyses are needed to determine which hypothesis is more likely.

If the degree of thermotolerance in our flies were entirely dependent on Hsp70 protein levels, we would expect O_{st}/O_{st} flies to have about the same average knockdown temperature than their O_{3+4}/O_{3+4} counterparts after the heat-induced treatment, when Hsp70 levels were about the same across karyotypes (Fig. 3). However, no significant karyotype × treatment interaction was detected in any case (hardening effects were also absent under fast ramping). It is therefore difficult to conclude from our present results that the high basal levels of Hsp70 in the warm-climate O_{3+4}/O_{3+4} karyotypes are causally linked to their higher basal thermotolerance.

The function of Hsp70 heat-shock proteins in *Drosophila* adult heat tolerance remains unclear even after our

Table 3 ANOVA for the effects of karyotype (O_{3+4}/O_{3+4} and O_{st}/O_{st}), sex and treatment (basal and hardened heat tolerance) on knockdown temperature estimated from two dynamic assays. The conditions in the 'slow ramping' assay were $T_0 = 24$ C and $\Delta T = 0.1$ C min^{-1} (data plotted in Fig. 4a) and in the 'fast ramping' assay $T_0 = 24$ C and $\Delta T = 0.6$ C min^{-1} (data plotted in Fig. 4b). ($\subset$ means 'nested in').

	Slow ramping				Fast ramping			
Source of variation	d.f.	MS	*F*	*P*	d.f.	MS	*F*	*P*
Covariate (water bath)	1	0.711	0.38	0.539	1	0.059	0.04	0.840
Karyotype (*K*)	1	2.557	0.33	0.583	1	7.891	9.19	0.016
Cross $\subset$ *K*	8	7.828	4.16	< 0.001	8	0.859	0.60	0.776
Sex (*S*)	1	71.007	37.77	< 0.001	1	3.125	2.19	0.140
Treatment (*T*)	1	10.691	5.69	0.018	1	0.160	0.11	0.738
$K \times S$	1	7.594	4.04	0.046	1	0.002	0.00	0.967
$K \times T$	1	1.800	0.96	0.329	1	0.419	0.29	0.589
$S \times T$	1	6.319	3.36	0.068	1	0.944	0.66	0.417
$K \times S \times T$	1	3.937	2.09	0.149	1	0.008	0.01	0.940
Error	221	1.880			219	1.427		

study (Dahlgaard *et al.*, 1998; Krebs *et al.*, 1998; Sørensen *et al.*, 1999; Folk *et al.*, 2006; Jensen *et al.*, 2010). One suggested reason for this is that the different experimental protocols used to assess CT_{max} seemingly capture different physiological and genetic mechanisms underlying thermotolerance because estimates (knockdown temperatures or times) are poorly correlated (Sgrò *et al.*, 2010). We have also detected a relatively low correlation between knockdown temperatures assessed under slow or fast ramping, but doubt this is because our heat tolerance assays are estimating different things. A more reasonable explanation is that the amount of experimental noise introduced by methodology to reliably estimate heat tolerance is substantial, as suggested by the low repeatability of this trait (about 0.2; Krebs & Loeschcke, 1997) and confirmed by our theoretical treatment, showing that many experimental protocols result in unreliable and highly biased estimates of CT_{max} (Santos *et al.*, 2011).

We have recently recommended that ramping rates in experiments aiming at studying thermal tolerance should be as fast as possible to minimize the contribution of uncontrolled variables (e.g. depletion of resources during assays, short-term acclimatory responses and stochasticity effects; Rezende *et al.*, 2011; Santos *et al.*, 2011). More important in the present context is that slow ramping protocols can grossly underestimate underlying genetic differences in critical thermal limits (Rezende *et al.*, 2011), which might explain why karyotype variation in heat stress resistance was not detected in our previous experiment (Dolgova *et al.*, 2010). We therefore believe that the present results with the fast ramping assay provide a more accurate representation of the genetic differences in heat tolerance between O_{st}/O_{st} and O_{3+4}/O_{3+4} flies.

In summary, our study clearly suggests that warm-climate karyotypes in *D. subobscura* are better able to cope with high stressful temperatures than their cold-climate counterparts. Combined with their consistent differences in thermal preferences (i.e. O_{3+4} carriers display a higher thermal preference than O_{st} ones; Dolgova *et al.*, 2010), we now conclude (Dolgova *et al.*'s claims notwithstanding) that the thermal co-adaptation hypothesis does find some support in this species and might also adaptively explain its diurnal activity patterns in summer, where O_{3+4} carriers are more active at midday (Savković *et al.*, 2004). However, the causal link between the comparatively high basal levels of Hsp70 heat-shock protein and high adult heat tolerance of O_{3+4}/O_{3+4} flies is not yet clear-cut.

Acknowledgments

We thank Doth Andersen for the technical help. Wolf Blanckenhorn and two anonymous reviewers provided constructive comments on an earlier draft. G. Calabria was supported by a predoctoral fellowship (BES-2007-15096) from the Ministerio de Educación y Ciencia (Spain) and by a short visit grant (ref. number 3097) from the ESF Programme on 'Thermal adaptation in ectotherms: linking life history, physiology, behaviour and genetics'. O. Dolgova was supported by a predoctoral fellowship (BES-2007-17438) from the Ministerio de Educación y Ciencia. C. Rego was supported by a post-doctoral fellowship (SFRH/BPD/39998/2007) from Fundação para a Ciência e a Tecnologia (Portugal). L. E. Castañeda is supported by a Juan de la Cierva fellowship (JCI-2010-06156) from the Ministerio de Ciencia e Innovación (Spain). E. L. Rezende is supported by a Ramón y Cajal contract and by grant BFU2009-07564 from the Ministerio de Ciencia e Innovación. M. Pascual is supported by grant CTM2010-22218 from the Ministerio de Ciencia e Innovación. M. Santos is supported by grant CGL2010-15395 from the Ministerio de Ciencia e Innovación and by the ICREA Acadèmia programme. Financial support by grant 2009SGR 636 from Generalitat de Catalunya to the Grup de Biologia Evolutiva is also gratefully acknowledged.

References

Angilletta, M.J. 2009. *Thermal Adaptation. A Theoretical and Empirical Synthesis*. Oxford Univ Press, Oxford, U.K.

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.

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

Balanyà, J., Solé, E., Oller, J.M., Sperlich, D. & 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.

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

Bettencourt, B.R., Hogan, C.C., Nimali, M. & 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 thermotolerance in *Hsp70* null flies. *BMC Biol.* **6**: 5.

Chown, S.L. & Nicolson, S.W. 2004. *Insect Physiological Ecology: Mechanisms and Patterns*. Oxford University Press, Oxford, UK.

Chown, S.L., Jumbam, K.R., Sørensen, J.G. & Terblanche, J.S. 2009. Phenotypic variance, plasticity and heritability estimates of critical thermal limits depend on methodological context. *Funct. Ecol.* **23**: 133–140.

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.

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

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

Feder, J.H., Rossi, J.M., Solomon, J., Solomon, N. & Lindquist, S. 1992. The consequences of expressing *hsp70* in *Drosophila* cells at normal temperatures. *Genes Dev.* **6**: 1402–1413.

Feder, M.E., Bedford, T.B.C., Albright, D.R. & Michalak, P. 2002. Evolvability of Hsp70 expression under artificial selection for inducible thermotolerance in independent populations of *Drosophila melanogaster*. *Physiol. Biochem. Zool.* **75**: 325–334.

Folk, D.G., Zwollo, P., Rand, D.M. & 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.

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

Hertz, P.E., Huey, R.B. & Stevenson, R.D. 1993. Evaluating temperature regulation by field-active ectotherms: the fallacy of the inappropriate question. *Am. Nat.* **142**: 796–818.

Hoffmann, A.A., Sørensen, J.G. & Loeschcke, V. 2003. Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *J. Therm. Biol.* **28**: 175–216.

Huey, R.B. & Bennett, A.F. 1987. Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. *Evolution* **41**: 1098–1115.

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

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

Krebs, R.A. & Feder, M.E. 1997a. Deleterious consequences of Hsp70 overexpression in *Drosophila melanogaster* larvae. *Cell Stress Chaperones* **2**: 60–71.

Krebs, R.A. & Feder, M.E. 1997b. 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. & Feder, M.E. 1998. Hsp70 and larval thermotolerance in *Drosophila melanogaster*: how much is enough and when is more too much? *J. Insect Physiol.* **44**: 1091–1101.

Krebs, R.A. & Loeschcke, V. 1994. Costs and benefits of activation of the heat-shock response in *Drosophila melanogaster*. *Funct. Ecol.* **8**: 730–737.

Krebs, R.A. & Loeschcke, V. 1997. Estimating heritability in a threshold trait: heat-shock tolerance in *Drosophila buzzatii*. *Heredity* **79**: 252–259.

Krebs, R.A., Feder, M.E. & Lee, J. 1998. Heritability of expression of the 70KD heat-shock protein in *Drosophila melanogaster* and its relevance to the evolution of thermotolerance. *Evolution* **52**: 841–847.

Krimbas, C.B. & Loukas, M. 1980. The inversion polymorphism of *Drosophila subobscura*. *Evol. Biol.* **12**: 163–234.

Loeschcke, V. & Hoffmann, A.A. 2007. Consequences of heat hardening on a field fitness component in *Drosophila* depend on environmental temperature. *Am. Nat.* **169**: 175–183.

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

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

Munté, A., Rozas, J., Aguadé, M. & Segarra, C. 2005. Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. *Genetics* **69**: 1573–1581.

Parsell, D.A., Taulien, J. & Lindquist, S. 1993. The role of heat-shock proteins in thermotolerance. *Philos. Trans. R. Soc. Lond. B* **339**: 279–286.

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

Prevosti, A., Serra, L., Ribó, G., Aguadé, M., Sagarra, E., Monclús, M. *et al.* 1985. Colonization of *Drosophila subobscura* in Chile. II. Clines in the chromosomal arrangements. *Evolution* **39**: 838–844.

Prevosti, A., Ribó, G., Serra, L., Aguadé, M., Balañá, J., Monclús, M. *et al.* 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.

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.

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

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

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

Rodr´guez-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.

Santos, M., Castañeda, L.E. & Rezende, E.L. 2011. Making sense of heat tolerance estimates in ectotherms: lessons from *Drosophila*. *Funct. Ecol.* **25**: 1169–1180.

Sarup, P., Sorensen, J., Dimitrov, K., Barker, J. & Loeschcke, V. 2006. Climatic adaptation of *Drosophila buzzatii* populations in southeast Australia. *Heredity* **96**: 479–486.

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

Sgrò, C.M., Overgaard, J., Kristensen, T.N., Mitchell, K.A., Cockerell, F.E. & Hoffmann, A.A. 2010. A comprehensive assessment of geographic variation in heat tolerance and hardening capacity in populations of *Drosophila melanogaster* from eastern Australia. *J. Evol. Biol.* **23**: 2484–2493.

Sokal, R.R. & Rohlf, F.J. 1995. *Biometry*, 3rd edn. Freeman, New York, NY.

Solé, E., Balanyà, J., Sperlich, D. & Serra, L. 2002. Long-term changes in the chromosomal inversion polymorphism of

Drosophila subobscura. I. Mediterranean populations from southwestern Europe. *Evolution* **56**: 830–835.

Sørensen, J.G. 2010. Application of heat shock protein expression for detecting natural adaptation and exposure to stress in natural populations. *Curr. Zool.* **56**: 703–713.

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.

Sørensen, J.G., Kristensen, T.N. & Loeschcke, V. 2003. The evolutionary and ecological role of heat shock proteins. *Ecol. Lett.* **6**: 1025–1037.

Terblanche, J.S., Deere, J.A., Clusella-Trullas, S., Janion, C. & Chown, S.L. 2007. Critical thermal limits depend on methodological context. *Proc. Biol. Sci.* **274**: 2935–2942.

Data deposited at Dryad: doi: 10.5061/dryad.6vn7t861

Received 19 October 2011; revised 30 December 2011; accepted 3 January 2012

"If anybody wants to clap," said Eeyore… ,
"now is the time to do it."

A.A. Milne,
Winnie-the-Pooh