MOLECULAR GENETIC ANALYSIS OF A SEASONAL CHARACTER IN *D. melanogaster* NATURAL POPULATIONS

Thesis submitted for the degree of Doctor of Philosophy

at the University of Leicester

by

Valeria Zonato

University of Leicester

November 2012

ACKNOWLEDGMENTS

First of all I would like to thank Bambos, who took the risk of accepting me in his research group and who always made sure I was happy and was not considering quitting my PhD (again). This thesis, if anything, proves that I stayed until the end. It has not always been easy, but every discussion and meeting with him not only restored full enthusiasm and motivation, but also reassured me that you do not necessarily need to be boring, serious and ordinary to have a very successful scientific career. Thanks for giving me the opportunity to attend many conferences, the Summer School, and the 'on the road' field work trip from Malaga to Barcelona. I really enjoyed these experiences thoroughly, even though I will always regret not making it to Florida.

I am grateful to have been given the opportunity to work with two Erasmus students who helped me with the lab work. These experiences made me realise that student supervision is difficult, energy demanding but equally rewarding. Sabrina contributed to some of the results in Chapter 6, whereas Giorgio helped substantially with Chapter 4 and later proceeded as a PhD student, becoming a unique and irreplaceable friend more than a colleague. Thank you for being the best student one could wish for, I am afraid overall you have been more helpful to me than the other way around. Thanks for coping with my mood swings, for never growing tired of my increasing stress and anger, and for being happy of my happiness. Hope you and Laura will forgive me one day, for not warning you enough about the 'British summer' before you decided to move to Leicester.

Many thanks to all the people in the fly lab, past and present, for creating a friendly, international environment, and for making every round of fly pushing a bit less painful. Special thanks go to Mirko for the many helpful and stimulating scientific discussions and for contributing to the results presented in Chapter 6. Thanks to John, for his witty comments and jokes, for standing the extreme level of 'italianness' of our lunch breaks, and for coming at my rescue when I fainted in the corridor. Thanks to Stephane, the only person who can express friendship, support and understanding even by being low, rude and vulgar, and who thankfully continues to be a friend now, even if he is on the other side of the Ocean. Thank you Ed, for the scientific and moral support provided in the last few weeks.

I am grateful to the people I have been living with at some point in Leicester, I could not have been any luckier! My first memories of Leicester will always and inevitably be linked

to Carlo's friendship and help. Situations and people change over time, but those memories are, at least for me, uncontaminated by current events. With Shanow and Luciana I spent endless nights gossiping and complaining: I'm grateful to them for always trying to cheer me up, not to mention the amazing dinners and brigadeiros they cooked.

Chris does not belong to the above paragraph, he has been much more than a housemate to me. Dear Homie, thank you for standing my (mild?) OCDs, for not talking to me in the morning, for sharing fried chicken balls and liquorice sweets and for always being there to listen and help me with a honesty and straightforwardness which is as helpful as rare. I genuinely hope the Hinckley HM was only the first of many others to be run together.

Thank you to everyone who made the last few months of the writing up process an unexpectedly nice and relatively relaxed experience.

Last but not least a thank you to my family who have always been there, discrete but trustworthy, constantly offering their help, trusting my decisions, enduring my bad temper and sharing my happiness. Thank you.

ABBREVIATIONS

(AND ACRONYMS)

aa	Amino acid
bp	Base pair
CA	Corpus Allatum
CC	Corpora Cardiaca
CNS	Central Nervous System
CPP	Critical Photoperiod
DD	Constant Darkness
DNA	Deoxyribonucleic Acid
Ecd	Ecdysone
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
gr	Grams
h	Hour
Indel	Insertion/deletion
JH	Juvenile Hormone
Kb	Kilo base
km	Kilometers
L	Litre
LD	Light Dark cycle
LD	Linkage Disequilibrium
min	Minute
n.a.	Not available
ns	Not significant
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PNS	Peripheral Nervous System
PPRC	Photoperiodic Response Curve
QTL	Quantitative Trait Loci
RNA	Ribonucleic Acid
s	Second
SAD	Seasonal Affective Disorder
SD	Standard Deviation
SEM	Standard Error of the Mean
SNP	Single Nucleotide Polymorphism
TG	Thoracic Gland
YP	Yolk Protein
ZT	Zeitgeber Time

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

CHAPTER 2: MATHERIALS & METHODS

CHAPTER 3: TIMELESS

CHAPTER 4: COUCH POTATO

Chapter 5: insulin-like receptor

Chapter 6: diapause

CHAPTER 7: FINAL DISCUSSION

CHAPTER 8: APPENDIX

REFERENCES

1. INTRODUCTION

1.1 BACKGROUND

Over thousands of years humans have slowly acquired the capability to modify their surrounding environment (through irrigation and deforestation), and to protect themselves from adverse climatic conditions. Today, due to high tech infrastructures, clothing, insulation and transport there are only a few places on Earth which remain unexplored. But what if we did not have technology at our aid? This is the case for all other animals and plants. In order for other living organisms (and therefore their progeny) to have better chances of survival, they had to adapt to their environment, evolving strategies to best use its resources. To add to the challenge, due to the Earth axial tilt, almost all regions on Earth are affected by cycling conditions throughout the year. The Tropics experience wet/dry seasons, temperate regions are affected by relatively mild changes in photoperiod (day-to-night length ratio) and temperature, whereas Polar regions go through extreme changes in photoperiod (Figure 1.1). Organisms therefore not only have to adapt to a specific environment, but they have to cope with the dramatic changes this environment goes through every time the Earth completes a revolution around the Sun: cycling food and water availability (or accessibility), presence of predator species, accessibility to shelters/nests, capability to camouflage with the changing environment.
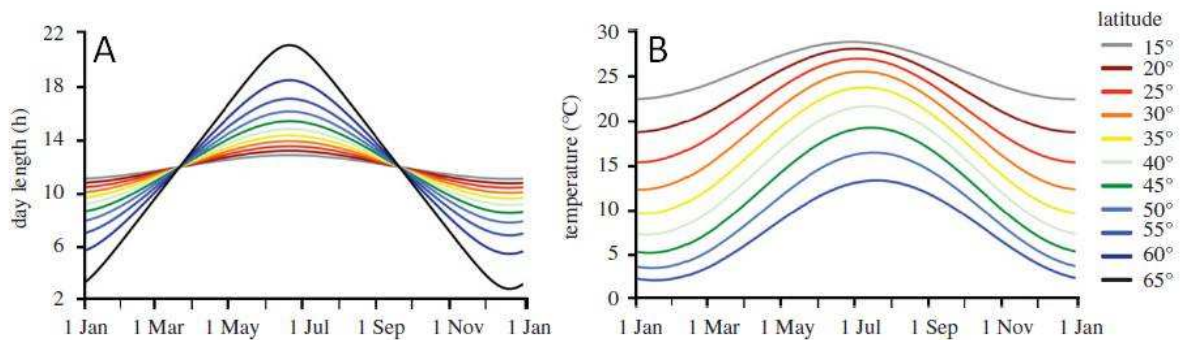
Figure 1.1
Photoperiod (A) and Temperature (B) changes throughout the year, in a range of latitudes. Figure from (Wilczek *et al.*, 2010).

These environmental challenges have always been faced by almost all organisms, and failure to adapt successfully is catastrophic. With such a wide and strong selective pressure it is not surprising that different species have ended up using two completely different approaches to escape harsh environmental conditions:

- In SPACE: allows animals to escape unfavourable climatic conditions by migrating to less challenging environments;

- In TIME: animals "toughen up" and "go on hold", until conditions become more bearable.

Regardless of the strategy adopted, timing is vital. Organisms need to be able to anticipate seasonal changes rather than merely react to them, since they cannot afford to get caught unprepared by potentially lethal conditions. In order to do so, they "measure" photoperiod, and they start getting ready for the disadvantageous conditions long before they manifest themselves. The first definition of photoperiodism appeared almost 100 years ago, and it was used by Garner to refer to the plant responses to day length (Garner & Allard, 1922). He showed that the reproduction of many plants, including carrot, lettuce and tobacco, is regulated by day length. Some plants undergo flowering and fruiting when exposed to long summer photoperiods, whereas others respond to short wintry photoperiods. However some plants are more sensitive than others to exposure from different photoperiods. Considerable

effort has been put into trying to understand how organisms perceive the annually cycling environmental cues, and the physiological/metabolic/phenotypic responses that follow.

1.2 ESCAPING WINTER IN TIME

The most commonly known way to escape harsh seasonal conditions in time, is to undergo hibernation. Animals reach a state of inactivity, slow metabolism and low body temperature, and they manage to deep sleep through winter using the food or the body fat they accumulated during the summer months. It is the strategy of choice of many mammals but also cold blooded animals such as bees, frogs, lizards and snakes hibernate.

Insects are small, ectothermic organisms, therefore they are very vulnerable to temperature changes. Drosophilids have been reported to be able to migrate between favourable environments within short period of times, and release-recapture experiments show that they can cover on average 500 m in 15 h (Coyne *et al.*, 1982). This capability, despite being remarkable, is of no help when one has to migrate to the other hemisphere. It is not surprising then that the insect strategy of choice is the escape from unfavourable conditions in time by entering dormancy, which differs from hibernation because in this case animals do not grow at all while in this state.

1.2.1 KINDS OF DORMANCIES

Two different kinds of insect dormancies (adaptive arrest in development or reproduction) have been described (Koštál, 2006).

- QUIESCENCE: immediate response of the organism to a sudden form of stress (e. g. lack of an essential environmental factor, abrupt change in environmental conditions). All the organism's physiological processes are restored to normality as soon as standard environmental conditions are restored.

- DIAPAUSE: neuroendocrinally controlled interruption of development or reproduction. It involves profound changes in metabolism which precede the beginning of unfavourable conditions and are not necessarily terminated when environmental conditions are restored.

Diapause is the strategy adopted to face predictable seasonal changes, and it can be further classified according to its compulsoriness and to when it occurs, in terms of developmental stage or time of the year.

- COMPULSORINESS:
    - COMPULSORY: animals enter diapause at every generation, regardless of the cue received from the environment. This is the case of the gypsy moth *Lymantria dispar* (Leonard, 1968), whose eggs mostly hatch in spring, regardless of the photoperiod and temperature they have been exposed to.
    - FACULTATIVE: whether to go or not in diapause exclusively depends on the environmental conditions faced by the organism. If cues are not strong enough to trigger a response, individuals might never

undergo diapause in their entire lifespan. This kind of diapause is much more common than the compulsory one.

- TIME OF THE YEAR:

    – SUMMER (aestivation type).

    – WINTER (hibernation type).

    These two kinds of diapause are very similar: they are initiated by environmental cues (above all the lengthening and shortening of photoperiods respectively) and involve cessation of development and increased resistance to stress. Some species can undergo both kinds of diapause within the same generation, in two distinct developmental stages (Masaki, 1980).

- DEVELOPMENTAL STAGE:

    – DEVELOPMENTAL DIAPAUSE: some organisms manifest an arrest in one particular developmental stage (egg, larva or pupa) in response to harsh conditions. For instance flesh flies (Sarcophagidae) arrest their development at the pupal stage, if their mothers are exposed to diapause inducing conditions (Denlinger, 1972), whereas the silkworm (*Bombyx mori*) manifests a maternally controlled embryonic diapause (Yamashita, 1996).

    – REPRODUCTIVE DIAPAUSE: other insects overwinter as adults and when in diapause they stop their reproduction.

The first experiments aimed at elucidating the dynamics of diapause induction in *D. melanogaster* date back to 1989, when Saunders and colleagues exposed *D. melanogaster* flies to a range of temperatures and photoperiods usually not used in laboratory experiments. They concluded that *D. melanogaster*, like most *Drosophila* species, is characterised by a facultative, winter, reproductive diapause (Saunders *et al.*, 1989), which has been defined "weak" or

"shallow" given that it can be easily terminated by increases in temperature. A long standing debate has ensued, as to whether it would be more appropriate to define it as diapause (because it is characterised by an anticipatory component) or quiescence (because it is terminated as soon as warm temperature is restored) (Tatar *et al.*, 2001a). For simplicity in this work this phenotype will be referred to as diapause.

1.2.2 DIAPAUSE TRIGGERS

As mentioned before, diapause is not simply a reaction to harsh condition. Organisms react to the gradual environmental changes that happen throughout the seasons, and get ready for winter before unfavourable conditions catch them unprepared. Furthermore, a fly needs to know exactly how much time is left before the trade-off between energy expenditure and offspring survival makes it unreasonable for the fly itself to start reproduction. So what are the *stimuli* which allow flies to perceive the oncoming winter?

1.2.2.1 TEMPERATURE

Different species differ in the developmental stage at which they are sensitive to diapause inducing *stimuli*, and in the duration of this sensitive stage (Tauber *et al.*, 1986). In the case of *D. melanogaster* it has been shown that diapause can be induced by exposing newly eclosed female flies to short photoperiods (less than 14 hours of light) and low temperature (12°C) (Saunders *et al.*, 1989). As for temperature, analysis of populations of *D. melanogaster* showed that it plays a much more important role than photoperiod in determining diapause levels (Emerson *et al.*, 2009b). No diapause is observed in *D. melanogaster* flies when kept at 14°C, regardless of photoperiod. Furthermore, a 2°C temperature drop (from 12 to 10°C) causes a remarkable increase in diapause incidence (from 20 to 80%). Nonetheless, in natural conditions temperature is not constant, but it cycles throughout the day. It has been shown that in many insect species even thermocycles

influence diapause levels: in the firebug *Pyrrhocoris apterus* thermoperiods induce a higher

diapause level than constant temperature (Kalushkov *et al.*, 2001), and not surprisingly in the

parasitic wasp *Nasonia vitripennis*, thermoperiods with a long cold-phase provoke a higher

diapause than those with a long warm-phase (Saunders, 1973). Furthermore it has been

shown that exposing *D. melanogaster* flies to thermocycles similar to those experienced in

temperate regions in mid-autumn (9°C-23°C) significantly increases the capability of flies to

survive 1 h exposure to sub zero temperatures (-7°C) (Kelty & Lee Jr, 2001). Overall, at least in

long days insects (in which reproduction occurs under long photoperiods), low temperature

exposure represents a condition *sine qua non* for diapause induction and overwintering.

1.2.2.2 PHOTOPERIOD

As shown in Figure 1.1, photoperiod changes predictably throughout the year. Hence

it does not come as a surprise that most insects use the length of the night to measure the

approach of winter (Saunders, 2002). If low temperatures are necessary for diapause

expression, photoperiod determines the extent of it. Figure 1.2 shows how a laboratory strain

of *D. melanogaster* responds to photoperiod. A Photoperiodic Response Curve (PPRC) is

obtained by exposing flies to a constant temperature (12°C), and to a range of photoperiods

(from total darkness: DD, to constant light: LL) (Saunders *et al.*, 1989). The decrease in

diapause incidence in DD conditions and in long photoperiods (>14 hours of light) is typical of

long day species, which reproduce in summer, whereas the high levels of diapause in LL

conditions are much less common amongst insects and give *D. melanogaster* PPRC a typical

wave shape (Saunders *et al.*, 1989).
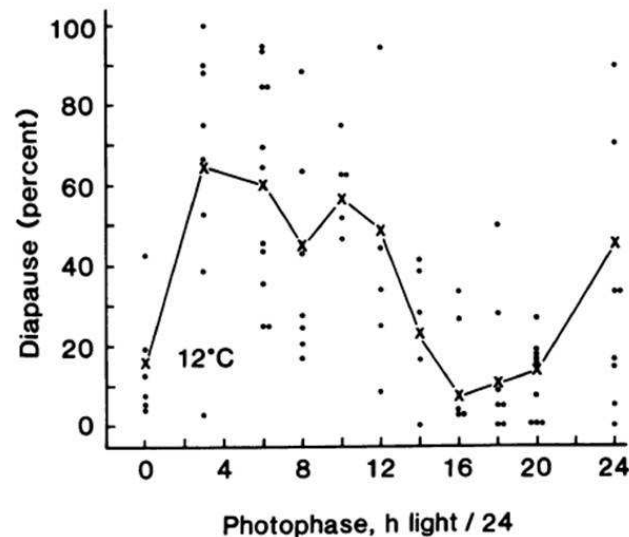
Figure 1.2
Photoperiodic Response Curve (PPRC) performed at 12°C of a laboratory strain of *D. melanogaster*. The decrease in diapause incidence in long photoperiods and in DD conditions is typical of long day species, whereas the increase in LL conditions in less common and unique to *D. melanogaster* flies. Figure from (Saunders *et al.*, 1989).

The PPRC allows us to determine the Critical Photoperiod (CPP): the borderline between what is perceived as short and long day, the day length at which half the flies are in diapause. In many *Drosophila* species the CPP depends on two parameters: <u>temperature</u> and <u>latitude</u> of collection of the flies used in the experiment. These two parameters affect the PPRC in a similar way: as experimental temperature gets lower, or latitude of collection gets further away from the Equator, the whole PPRC elevates along the y-axis, thus leading to longer CPP (Saunders *et al.*, 1989; Pittendrigh & Takamura, 1987; Lumme & Oikarinen, 1977) and suggesting that in particularly cold conditions diapause is initiated when the day is still relatively long (Figure 1.3).
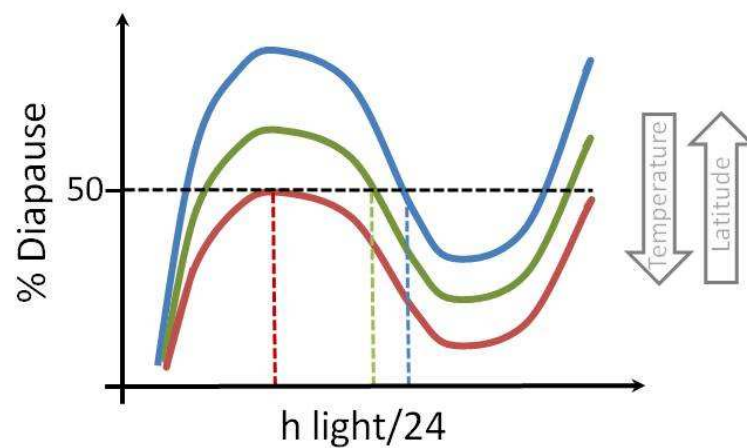
Figure 1.3
Effect of temperature and latitude of collection on the PPRC. Decreases in temperature and increases in latitude shift the PPRC upwards, along the y-axis, thus leading to an increased CPP.

Only recently, natural lines of *D. melanogaster* have been tested in respect to their capability to distinguish between different photoperiods. Interestingly, whereas in European flies the level of diapause predictably decreases when the photoperiod grows longer (Tauber *et al.*, 2007), this does not hold true for American lines, which do not respond differently when exposed to two experimental photoperiods (Emerson *et al.*, 2009b). The question as to what is causing this phenotypic difference will be addressed in Chapter 6 of this dissertation.

1.2.2.3 OTHER CUES

Dormancy is a widespread strategy also among tropical insects. Photoperiod obviously does not represent a strong seasonal clue in the Tropics, nor is temperature, whose fluctuations are more pronounced on a daily basis than throughout the year. Nevertheless tropical regions are subjected to alternating periods of rainfall and drought. This leads to cycles in vegetation flush, creation of puddles and presence of parasites and predators, thus cyclically affecting insect habitat, food availability and oviposition sites. Developmental and reproductive dormancy is a common (but not universal) strategy adopted by tropical insects, including some dipteran species, to survive this challenging environment (Denlinger, 1986). The triggers for this response include drop in relative humidity (caused by the lack of rainfall), and changes in moisture, protein and lipid content of host plants (Denlinger, 1986).

As for *D. melanogaster*, the species originated in sub Saharan Africa, around 2-3 My ago, when a split occurred from a *D. melanogaster* and *D. simulans* common ancestor (David & Capy, 1988). Nevertheless it has been suggested that diapause appeared upon *D. melanogaster* colonisation of temperate climates since this phenotype has not been detected in African flies, nor in the *D. melanogaster* sibling species *D. simulans* (Schmidt *et al.*, 2005a).

1.3 LIFE HISTORY TRAITS

Before venturing into the molecular mechanisms regulating diapause, it is important to look at the life history traits associated with it, to better understand why it is advantageous in temperate regions.

1.3.1 INDUCED BY DIAPAUSE STATE

D. melanogaster flies kept in diapause inducing conditions (11°C) for 3, 6 or 9 weeks exhibit mortality rates similar to those of newly eclosed flies, they are more resistant to heat and methyl viologen-induced oxidative stress and they have a lower fecundity, regardless of the photoperiod at which they were exposed, and on the duration of such exposure (Tatar *et al.*, 2001a). These characteristics, other than being due to the "low rate of living" merely induced by the low temperature, are promoted by the diapause state, since they are lost upon treatment with a Juvenile hormone analogue (which has been shown to stop reproductive diapause in *D. melanogaster*, see below). Generally, during the diapause period flies do not seem to be aging and they acquire the capability to resist stress. This makes the phenotype very advantageous in temperate regions characterised by harsh winter conditions.

1.3.2 ENDOGENOUS TRAITS

Diapause is a very variable phenotype, and among natural lines collected in the same location, some will enter diapause if exposed to the right conditions, some will maintain normal rates of reproduction, and the vast majority of them will display intermediate levels of diapause induction (Williams & Sokolowski, 1993). Schmidt and coworkers divided natural inbred isofemale lines in diapausing (D) and non-diapausing (ND, 100% and 0% diapause respectively, at 12°C). Remarkably, many life history traits differ between these lines, even when the flies were exposed to normal laboratory conditions (25°C, LD 12:12) throughout their entire lifespan (Schmidt *et al.*, 2005a). ND lines survived significantly longer, showed a higher resistance to starvation and cold stress, had a higher proportion of viable eggs, but were characterised by lower fecundity and a longer mean egg to adult developmental time (Schmidt *et al.*, 2005a).

These results highlight underlying constitutive differences between diapausing and non-diapausing lines, which are independent of the diapause state itself. Given these fitness related differences it is easy to understand the advantages of non-diapausing genotypes, in southern, non climatically challenged environments, since short developmental time and high early life fecundity certainly contribute in increasing the fitness.

1.4 CLINES

1.4.1 WHAT DO THEY TELL US?

The term "cline" comes from the Greek word klīnein which means "to possess or exhibit gradient", "to lean". It was first used by the English evolutionary biologist Julian Huxley in 1936 to describe a gradual change of a specific characteristic of a species over a geographical area, as a result of environmental heterogeneity. Examples of phenotypic clines in *D. melanogaster* include heat susceptibility and cold resistance (Hoffmann *et al.*, 2002), egg size, ovariole number (Azevedo *et al.*, 1996) and body size (James *et al.*, 1997; Gockel *et al.*, 2001).

Not surprisingly many allele frequencies have also been found to vary significantly with latitude. The most extensively studied *D. melanogaster* clinal gene is the *Alcohol dehydrogenase* gene (*Adh*) which presents two allozymes (*Adh-S* and *Adh-F*). The two alleles differ in the their kinetic properties, thermostability and expression levels, and flies characterised by different *Adh* genotypes show different fitness related phenotypes, in terms of survivorship, developmental time and ethanol tolerance (Heinstra, 1993). The frequency of the allozyme *Adh-F* increases significantly with latitude in North America (Vigue & Johnson, 1973), Australia and Asia (Oakeshott *et al.*, 1982), tropical Africa (David & Mercot, 1986) and Europe (David *et al.*, 1989). The two allozymes seem to be maintained in the population by balancing selection, with the allele *Adh-F* increasing the flies' alcohol tolerance in more temperate regions (David & Mercot, 1986). Nevertheless some ambiguity remains as to which selective pressure is maintaining the two alleles. Many other metabolic genes have been shown to be characterised by clinal variation (Sezgin *et al.*, 2004), together with heat shock proteins (Frydenberg *et al.*, 2003), clock genes (Tauber *et al.*, 2007; Costa *et al.*, 1992), and chromosomal inversions (Knibb, 1982).

The presence of a latitudinal cline in a phenotype or allele gene frequency is often used to infer adaptation to the environmental conditions which change with latitude. Nevertheless even demographic effects such as migration, random colonisation events, founder effect, genetic drift and bottle-necks might give rise to latitudinal gradients. The presence of similar clines in different continents (Azevedo *et al.*, 1996; Sawyer *et al.*, 2006) or in sibling species (Hallas *et al.*, 2002), the observation that the cline under study changes over time in response to climate change (Bradshaw & Holzapfel, 2001; Anderson *et al.*, 2005; Umina *et al.*, 2005), and the non-clinality of neutral loci (Sezgin *et al.*, 2004; Berry & Kreitman, 1993) are all indications that the cline has an adaptive meaning, as opposed to being merely driven by population dynamics such as bottle-necks, migration or founder effects.

1.4.2 DIAPAUSE RELATED CLINES

Photoperiod and temperature change throughout the year, but their absolute values and the extent of their annual fluctuation depend on latitude (Figure 1.1). It does not come as a surprise that diapause incidence (and traits associated with it), also shows clines. Figure 1.4 summarises the effect of latitude on diapause in *D. melanogaster* natural lines collected along the US and Australia East coast (A and B) and Europe (C). When *D. melanogaster* natural lines collected in USA are kept at 12°C for 28 days, they show a diapause incidence which predictably increases with latitude of collection (Figure 1.4 A) (Schmidt *et al.*, 2005a). The same holds true for Australian flies (Figure 1.4 B), except in this case a surprising increase in diapause levels is observed in flies collected close to the Equator (latitude below 25 degrees, white data points in Figure 1.4 B) (Lee *et al.*, 2011). As for European flies, a significant latitudinal cline in diapause levels is only observed when flies are kept in diapause inducing conditions for 12 days (filled data points in Figure 1.4 C), whereas the traditional-28-day protocol does not show any significant correlation between diapause levels and latitude of collection (white data points in Figure 1.4 C) (Pegoraro *et al.*, unpublished). The reasons for

these discrepancies are partly due to experimental approaches, and partly to environmental

differences between these continents. These issues are further discussed in Chapter 6 of this

thesis.
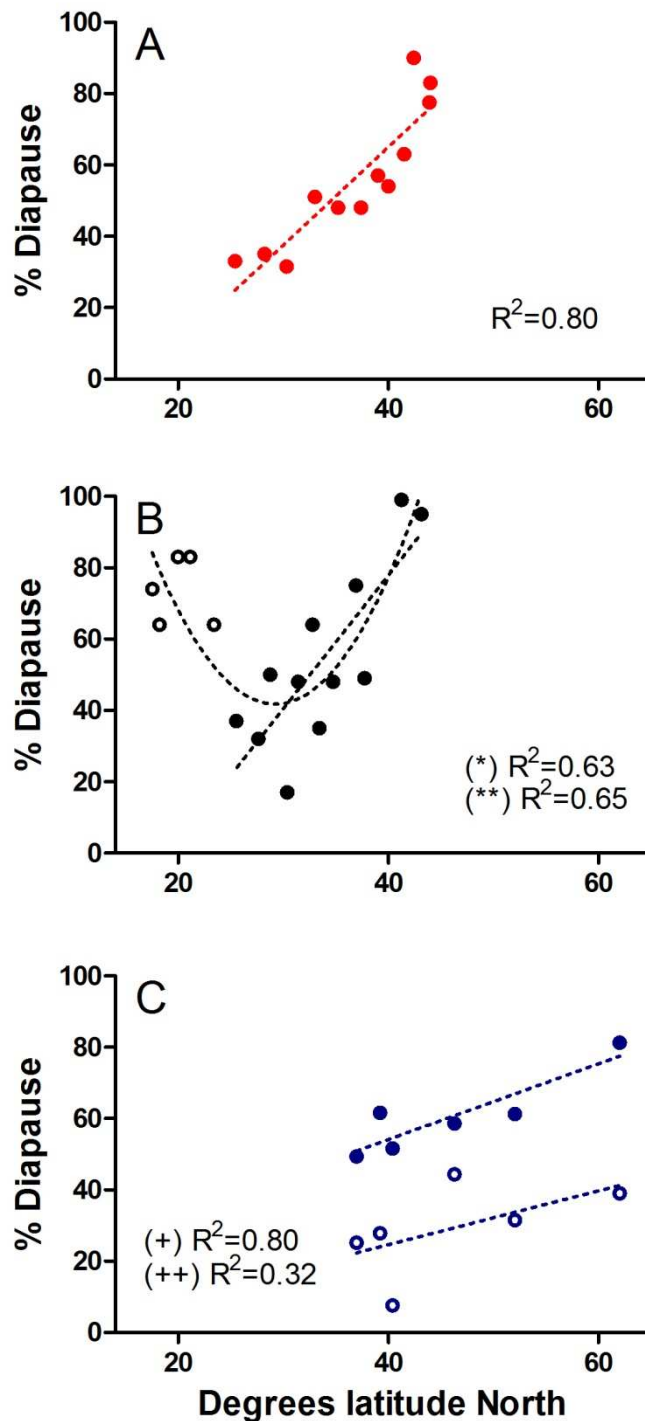
Figure 1.4

Diapause level of natural *D. melanogaster* lines over latitude of collection. A) Flies collected along the US East coast (data re-plotted from Schmidt *et al.*, 2005a). B) Flies collected along the Australian East coast. (*) Quadratic regression line fitted through all the dataset. (**) Linear regression fitted through datapoints from latitudes>25 degrees (solid dots; data re-plotted from Lee *et al.*, 2011). C) Flies collected in Europe and analysed after 12 (+) or 28 (++) days in diapause inducing conditions (filled and white data points respectively; Pegoraro *et al.*, unpublished)

Not only does diapause incidence significantly vary with latitude, at least in American and Australian natural populations, but not surprisingly also life history traits associated with this phenotypes show the same geographical distribution. As mentioned before, desiccation and starvation resistance (Hoffmann *et al.*, 2001), longevity and timing of reproduction (Mitrovski & Hoffmann, 2001) all increase with latitude, and so does ovariole number, (Schmidt *et al.*, 2005a; Azevedo *et al.*, 1996), but not influencing overall female fecundity.

1.5 NEUROENDOCRINE CONTROL

After addressing the phenotypical and evolutionary aspect of diapause we can now turn our attention to the neuroendocrine control of this complicated phenotype.

Three major events happen upon diapause-inducing environmental input (Figure 1.5):

1. Insulin-like-peptide (ILP) secretion by neurosecretory cells in the dorsal midbrain.

2. Secretion of Juvenile hormone (JH) by the *corpora allata* (CA) in the ring gland.

3. Ecdysteroid (Ecd) released by the ovaries stimulates yolk protein uptake.
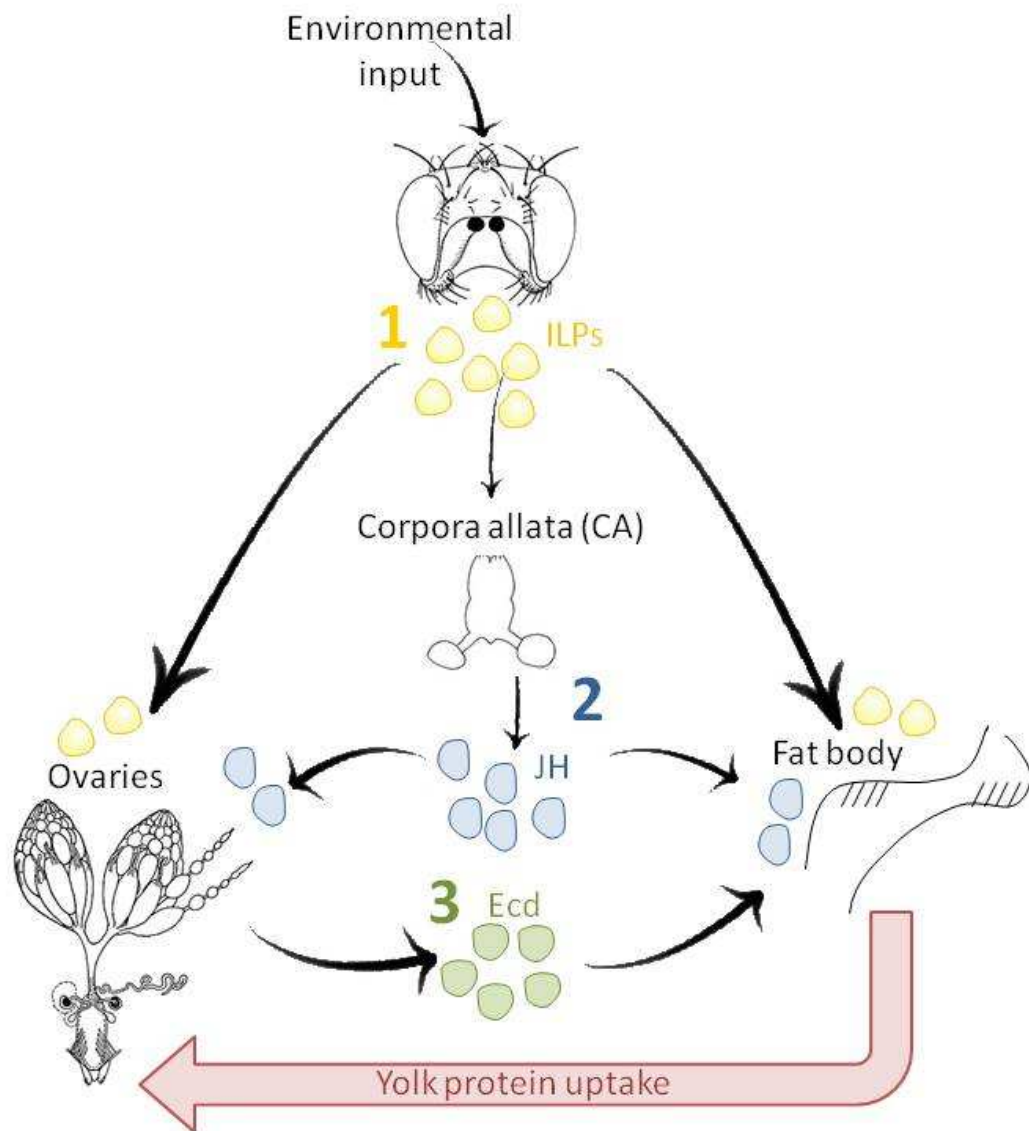
Figure 1.5
Three different steps of the neuroendocrine control of diapause, triggered by the appropriate environmental conditions. 1) Secretion of ILPs by dorsal midbrain neurons. 2) JH is released by the CA. 3) Ecd is secreted by the ovaries and reaches the fat bodies. All these events eventually lead to yolk protein uptake and ovarian development. See text for a more detailed explanation of each single step.

1.5.1 WHAT HAPPENS IN THE BRAIN? (1)

Not much is known about the molecules and mechanisms that allow *D. melanogaster* flies to measure shortening photoperiods and lowering temperatures. Nevertheless the neuroendocrine changes happening after the trigger has been pulled have been characterised, and involve, at first, the activation of the Insulin-like signaling pathway.

The Insulin signaling pathway is well conserved among mammals, nematodes and insects. In *Drosophila*, upon appropriate environmental triggering, neurosecretary cells in the *pars intercerebralis* of the brain, produce Insulin-like peptides (ILPs) (Cao & Brown, 2001; Broughton *et al.*, 2005). The ILPs are released in the haemolymph and eventually reach the ovary, the *corpora allata* and the fat body. *D. melanogaster* possesses 7 *Ilp* genes (*Ilp1-Ilp7*) which show a significant degree of homology with the human *Insulin* gene (Claeys *et al.*, 2002). At least 3 of them (*Ilp2*, *3* and *5*) are expressed by the *pars intercerebralis*, whereas one (*Ilp5*) is also expressed in the ovaries (Broughton *et al.*, 2005).

Not surprisingly also the *Insulin-like Receptor* gene is very well conserved, and the *D. melanogaster* gene shows 30-44 % similarity to the human homologue (Yenush *et al.*, 1996). Activation of the receptor by the ILPs triggers a cascade of events, mainly phosphorylations, which eventually lead to the inactivation of the transcription factor FOXO (Giannakou & Partridge, 2007) (Figure 1.6).
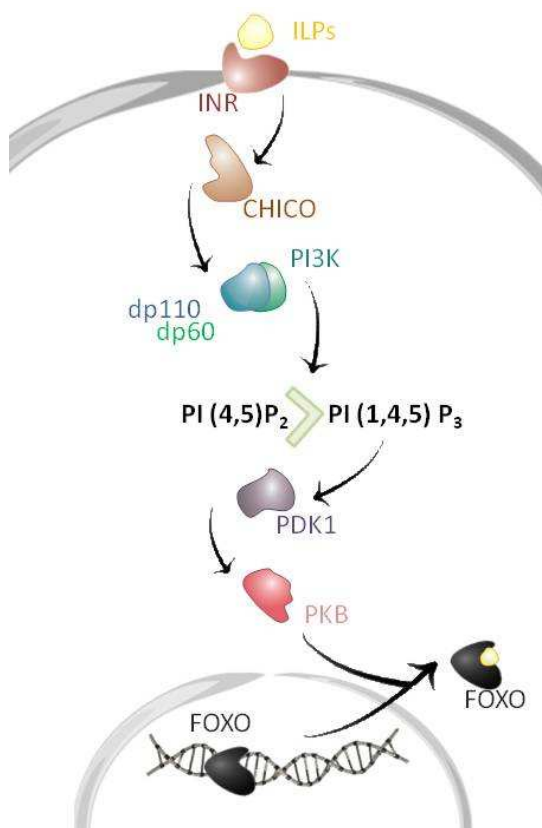
Figure 1.6
ILPs bind to the Insulin-like Receptor (INR), which transduces the signal to its molecular partner CHICO. CHICO activates the PI 3-kinase (which consists of two subunits, a catalytic and a regulatory one, dp110 and dp60 respectively), which converts phosphatidylinositol (4,5)-biphosphate to phosphatidylinositol (1,4,5)-triphosphate. This in turns activates a series of kinases which eventually phosphorylate the transcription factor FOXO, leading to its inactivation and translocation to the cytoplasm.

Insulin signaling is involved in many diverse processes, and it probably appeared early in evolution since its roles are conserved among different species: in mammals, nematodes and insects, downregulation of the insulin signaling pathway leads to slower growth rates, longer lifespan, compromised reproduction and higher resistance to stress (Tatar *et al.*, 2003; Kenyon, 2005). Given its relevance in influencing lifespan, this pathway has been studied extensively, and many mutations have been found in *Drosophila* which compromise the insulin-like-signaling pathway thus affecting the above mentioned life history traits. Loss of *chico* (Figure 1.6) leads to smaller, sterile, longer-lived flies, which are more resistant to acute oxidative stress (Clancy *et al.*, 2001). Other mutations to have similar effects include those affecting the *Insulin-like-Receptor* gene (Tatar *et al.*, 2001b), those causing the loss of the ILP producing neurons (Broughton *et al.*, 2005), and the overexpression of the transcription factor FOXO (Hwangbo *et al.*, 2004). Other than affecting life history traits associated with diapause incidence, mutations of *InR* in flies lead to the eclosion of females with extremely immature ovaries, which remain previtellogenic in adults, strongly resembling diapausing female ovaries (Tatar *et al.*, 2001b). Furthermore in *C. elegans, daf-2* mutations (the *InR* homologue), arrest development at the dauer larval stage, the nematode equivalent of diapause, regardless of environmental conditions (Kimura *et al.*, 1997), and this leads to a substantial lifespan lengthening. Finally, the neuronal overexpression of the *Dp110* gene, which encodes the catalytic subunit of the PI3K kinase (see Figure 1.6) has been shown to significantly decrease diapause levels in *D. melanogaster* (Williams *et al.*, 2006). All these results are consistent with a *scenario* where the insulin-like pathway has a pleiotropic effect on many fitness related phenotypes, and the activation of this pathway is linked to the more stress resistant/less reproductive active diapause state.

1.5.2 SECONDARY HORMONE: JH (2)

How does the insulin-like-signaling pathway control diapause levels? What are the secondary hormones involved in triggering (or preventing) ovary maturation? ILPs released in the hemolymph reach the ovaries and the fat body, but they also travel down the frontal ganglion to reach the ring gland. The ring gland is a complex structure which surrounds the *D. melanogaster* esophagus, and it consists of three endocrine organs joined together: the *Corpus allatum* (CA), the Thoracic gland (TG) and the *Corpora cardiaca* (CC). The ring gland is directly innervated by neurons in the *Pars intercerebralis* (Siegmund & Korge, 2001), and when reached by the ILP triggered signaling, the CA produces Juvenile Hormone (JH). Silencing the expression of the *InR* in the CA represses the expression of a key enzyme in JH synthesis, thus providing a direct link between insulin signaling and JH production (Belgacem & Martin, 2006). JH has been known for many years to play a central role in regulating reproduction, promoting vitellogenesis and thus inhibiting diapause (Tatar *et al.*, 2003; Tatar & Yin, 2001). In particular JH stimulates yolk protein (YP) synthesis by ovaries and fat bodies and their uptake from the hemolymph by the developing oocyte. JH deficient mutants are characterised by severely reduced vitellogenesis, which is restored upon application of the JH analog methoprene (Postlethwait & Weiser, 1973). Similarly, diapausing flies are characterised by low levels of JH, and the application of JH to the abdomen of these flies restores vitellogenesis (Saunders *et al.*, 1990).

1.5.3 ECDYCOSTEROIDS AT WORK (3)

The other secondary hormone involved in this signaling cascade is ecdysone (Ecd). It is a steroid hormone required at various stages of *D. melanogaster* development (Garen *et al.*, 1977). In adults Ecd is produced by follicle cells in the ovary, upon stimulation by JH. It is then transported to the fat body, where it triggers the production of YP (Garen *et al.*, 1977; Kozlova

& Thummel, 2000) which are then taken up by the ovary, thus starting vitellogenesis. Interestingly, mutations affecting the concentration of Ecd, the expression of its receptor *EcR* or its subunit *dare*, have similar phenotypes, leading to female sterility (Carney & Bender, 2000). Not surprisingly the application of Ecd to diapausing females triggers vitellogenesis (Richard *et al.*, 1998; Richard *et al.*, 2001).

So what is the link between the insulin like signaling, the JH cascade, and ecdysone? *InR* mutants are JH deficient and characterised by low levels of Ecd (Tatar & Yin, 2001; Tu *et al.*, 2002b). Richard and coworkers demonstrated that Ecd alone is responsible for initiating vitellogenesis. As mentioned before methoprene application to JH mutants restores the production of YP, but this is mediated by the stimulation of ovarian Ecd synthesis promoted by the JH analog (Richard *et al.*, 1998; Richard *et al.*, 2001).

1.6 HUNTING FOR DIAPAUSE GENES

The next paragraphs will describe the Mendelian properties of the diapause trait (Paragraph 1.6.1), and the characteristics of three diapause genes which have been the focus of our interest in this project (Paragraphs 1.6.2-4). What these three diapause genes appear to have in common is that they have been targeted by natural selection, resulting in a fine modulation of the phenotype with latitude. All the three genes described here are characterised by natural polymorphisms which affect diapause levels and present interesting geographical distributions in their frequencies.

1.6.1 Genetic control of diapause

The first genetic analysis of the diapause phenotype was performed in 1993 by Williams and Sokolowski (Williams & Sokolowski, 1993). They crossed natural lines characterised by dramatically different diapause levels, and scored the phenotype of the progeny. By doing so, the authors excluded any contribution of X chromosome, maternal effect and permanent cytoplasmatic factors (e.g. mitochondrial genes) in determining diapause differences between the lines, and they concluded that diapause follows a recessive autosomal model of inheritance (Williams & Sokolowski, 1993). This conclusion partly disagrees with findings obtained more than 10 years later. Schmidt and coworkers determined that the diapause phenotype does not follow a simple autosomal dominant trait hereditary pattern (Schmidt *et al.*, 2005b). The offspring of crosses between lines with the same phenotype (D X D and ND X ND; D (diapausing) and ND (non-diapausing)), behave exactly like the parental lines. Furthermore the F1 hybrids all show diapause, suggesting that diapause is dominant over non-diapause. Nevertheless backcrossing F1 hybrids to ND lines does not produce the ND:D=1:1 ratio expected for a simple autosomal dominant trait (Schmidt *et al.*, 2005b). Additionally the authors determined that the phenotype of a specific line was solely determined by the origin of the 3rd chromosome (Schmidt *et al.*, 2005b).

1.6.2 *timeless*

The gene *timeless* (*tim*) has been identified almost 20 years ago (Myers *et al.*, 1995; Sehgal *et al.*, 1995; Sehgal *et al.*, 1994) as a key component of the molecular mechanism regulating the circadian clock in *D. melanogaster* (from the Latin circa, meaning "about", and diem, day).

1.6.2.1 CIRCADIAN VERSUS SEASONAL CLOCK

What is the circadian clock, and what makes it different from the seasonal time measurement machinery? There are four main differences between the two kinds of biological time measurement machineries:

1)   PERIOD OF OSCILLATION:

- The circadian clock regulates all those phenotypes which cycle with a period of about 24 hours.
- The seasonal clock is characterised by a rhythmicity of 12 months.

2)   FREE RUN:

- The circadian clock is endogenous and keeps ticking even in the absence of any external cues.
- The output of the seasonal clock is triggered by appropriate environmental conditions.

3)   TEMPERATURE COMPENSATION:

- The period of the circadian clock is constant over a relatively wide range of physiologically relevant temperatures.
- The occurrence of seasonally related phenotypes strongly depends on temperature.

4)   MOLECULES:

- The first circadian gene was identified in 1971 (Konopka & Benzer, 1971) and since then many other genes have been discovered and characterised, as well as the localisation of circadian neurons and their network.
- Not much is known about the mechanisms behind the seasonal clock machinery.

1.6.2.2 ROLE OF *timeless* IN THE CIRCADIAN CLOCK

As mentioned above *tim* was the second clock gene to be identified (Myers *et al.*, 1995; Sehgal *et al.*, 1995), and together with *period* (*per*), it is part of the molecular dimer often referred to as the "negative element". Figure 1.7 shows the general, and very simplified mechanism underlying the circadian clock in *D. melanogaster*. It consists of a negative feedback loop, where positive molecules bind the DNA and activate the transcription of clock-genes. Among these clock genes are the negative elements (*tim* and *per*), whose products form a dimer capable of translocating back into the nucleus, binding to the positive elements, thus inhibiting their own transcription (Hardin, 2005).

A lag between the mRNA production of *tim* and *per* and the inhibition of the positive elements occurs due to posttranslational modifications which delay TIM and PER protein accumulation. In the cytoplasm, PER is efficiently phosphorylated by the kinase DOUBLE-TIME (DBT) and is targeted for degradation *via* the ubiquitin-proteasome pathway; thus the accumulation of monomeric PER is seriously reduced (Price *et al.*, 1998). Concomitantly TIM concentration increases and it binds the DBT/PER heterodimer, thus preventing PER degradation and allowing the nuclear entry of the complex TIM/DBT/PER, and the repression of their own transcription. This lag makes it so that the mRNA and protein levels of clock genes cycle throughout the day, with a period of roughly 24 hours and a lag of 6 hours. Also TIM is posttranslationally modified allowing the phase of the clock oscillation to be synchronised with the external conditions. A circadian clock would be of little or no use at all if it could not detect daily environmental fluctuations and synchronise to them. The cytoplasmic concentration of TIM is critical as it sets the phase of the rhythm depending on the light/dark condition. TIM concentration is, in fact, affected by proteasome-mediated degradation triggered by light, thanks to its interaction with the photoreceptor CRYPTOCHROME (CRY). This blue light activated photoreceptor can be considered as a mediator between external environment and

the endogenous clock, setting the phase of the output rhythm. It belongs to the photolyase/cryptochrome family, which comprises proteins able to perceive blue light (300-500 nm) and to carry out different functions. Photolyases use the blue light to reverse the deleterious effects induced by far-UV light (200-300 nm), whereas cryptochromes act as light sensors able to reset and synchronise the circadian clock to the solar day. Cryptochromes possess two domains (Lin & Shalitin, 2003):

- N-terminal domain which contains two chromophore binding sites: one binding a pterin, the other a flavin adenin-dinucleotide (FAD).

- C-terminal domain of 20-200 amino acids which is involved in protein-protein interaction and signal transduction.
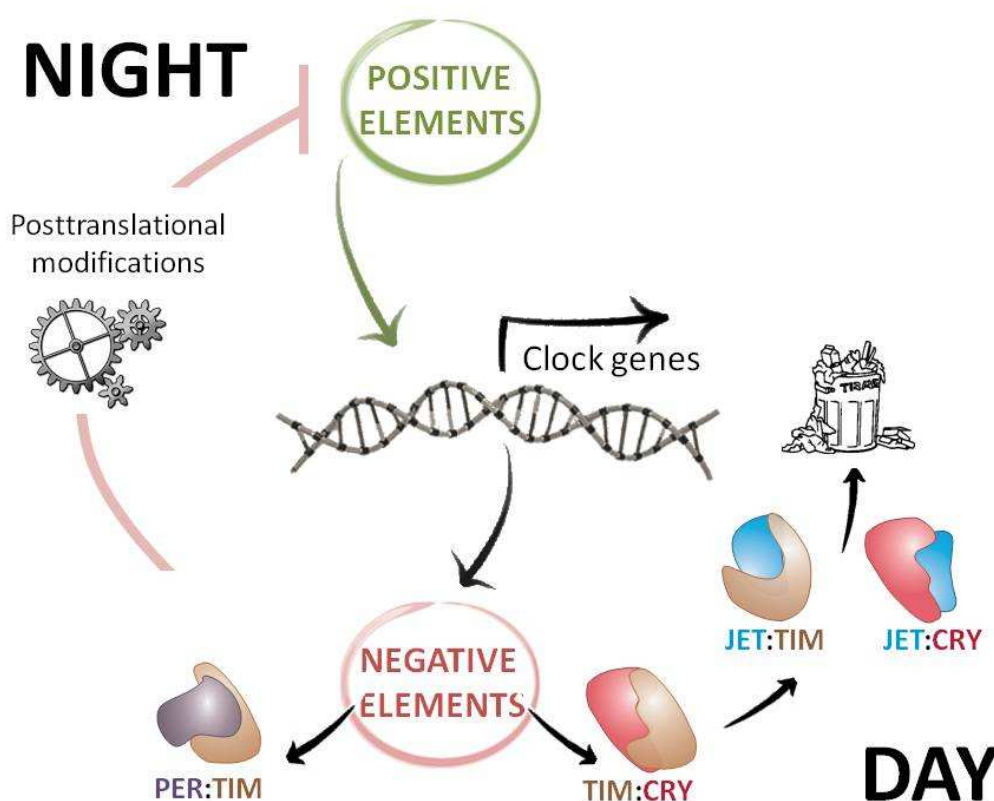
Figure 1.7
Schematic representation of the negative feedback loop, which represents the basis of the circadian clock in *D. melanogaster*. See text for a more detailed explanation.

CRY is activated by light which probably causes an inter- or intra- molecular redox reaction that might involve the flavin (Lin & Shalitin, 2003). Thanks to this activation CRY binds to TIM. As a result of TIM:CRY interaction, a posttranslational modification occurs in TIM, which in turn enables it to bind to JETLAG (JET), a F-box protein also involved in the resetting mechanism of the circadian clock (Peschel *et al.*, 2009). JET promotes TIM degradation *via* the ubiquitin-proteasome pathway, thus sequestering it from the interaction with PER and postponing the time of their entrance into the nucleus. CRY also binds to JET, but its affinity for JET is weaker than the one for TIM. Therefore, the CRY:JET dimer only occurs when the concentration of TIM in the cytoplasm has decreased. As a result of this interaction, CRY is also degraded *via* the ubiquitin-proteasome pathway (Figure 1.7) (Peschel *et al.*, 2009).

1.6.2.3 *timeless* ALLELES AND THEIR ROLE IN DIAPAUSE

D. melanogaster timeless is characterised by two initiating methionine codons, 23 codons apart (Rosato *et al.*, 1997b). The two *ATG* starting triplets theoretically translate a long and a short form of TIM, 1421 and 1298 aminoacids long respectively (Rosato *et al.*, 1997b). Interestingly, both laboratory and natural fly strains present an indel polymorphism in the region between the two starting codons, giving rise to two alleles, *ls-tim* and *s-tim*. In *ls-tim* (*long/short-tim*) both starting codons are used and as a consequence both the long and short form of TIM are expressed (Sandrelli *et al.*, 2007). In *s-tim* (*short-tim*) a single nucleotide deletion in position 294 introduces a premature stop codon 19 residues after the upstream starting codon. As a consequence only the shorter variant of TIM is expressed, together with a 19 aa N-terminal peptide (Figure 1.8).

The two alleles have been characterised both molecularly and phenotypically. From a molecular point of view L-TIM protein has been shown to interact weakly with its molecular partner CRY (Sandrelli *et al.*, 2007), thus is more stable and makes *ls-tim* flies less sensitive to short light pulses (Sandrelli *et al.*, 2007). More recently it has also been shown that the strong

S-TIM:CRY interaction prevents CRY from binding to and being degraded by JET, at least as long as the concentration of S-TIM in the cytoplasm is high. On the other hand, the lower affinity of L-TIM for CRY allows an earlier binding of CRY to JET, and an earlier degradation of CRY (Peschel *et al.*, 2009).
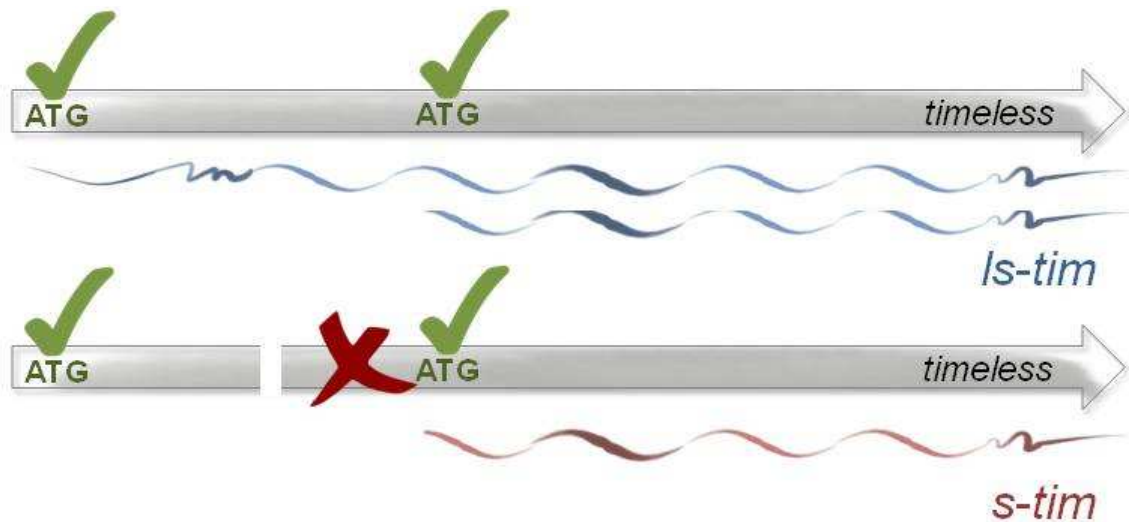
Figure 1.8
Schematic representation of the two *timeless* alleles, *ls-tim* (top, blue) and *s-tim* (bottom, red). In green the two alternative *ATG* start codons, the red cross represents the premature stop codon introduced by a single nucleotide deletion in position 294.

Phenotypically speaking, other than making flies less sensitive to light, the *ls-tim* allele confers a higher level of diapause (Tauber *et al.*, 2007). Figure 1.9 (panels A-C) shows the level of diapause of *ls-tim* (squares) and *s-tim* (triangles) lines collected in three different European locations, and tested at 6 different photoperiods. In all cases, *ls-tim* flies show a higher level of diapause than *s-tim* flies (Tauber *et al.*, 2007).

As for the frequency of the alleles in natural populations, counter-intuitively the high diapausing *ls-tim* allele is more common in flies collected at southern latitudes, and it decreases moving northwards. Nevertheless two outlier populations, collected in Israel and Crete, disrupt the negative correlation between latitude and *ls-tim* frequency (Figure 1.9 D) (Tauber *et al.*, 2007). A significant and robust cline is observed when overland distances from putative site of origin of the *ls-tim* allele is considered as the driving factors, instead of

latitude. In this case, the X axis represents the overland distance from Novoli, South of Italy, namely the location where the *ls-tim* frequency is the highest (Figure 1.9 E) (Tauber *et al.*, 2007). Given that the *ls-tim* allele does not seem to be present in African *D. melanogaster* populations, it probably derived from the ancestral form of *tim* (*s-tim*) through the random insertion of a single nucleotide (Tauber *et al.*, 2007). This insertion is likely to have happened in a single fly in the South of Italy, and given its (apparent) advantageous properties, it quickly increased in frequency. The reasons for the establishment and maintenance of the *ls-tim* distance cline, are discussed in Chapter 3.
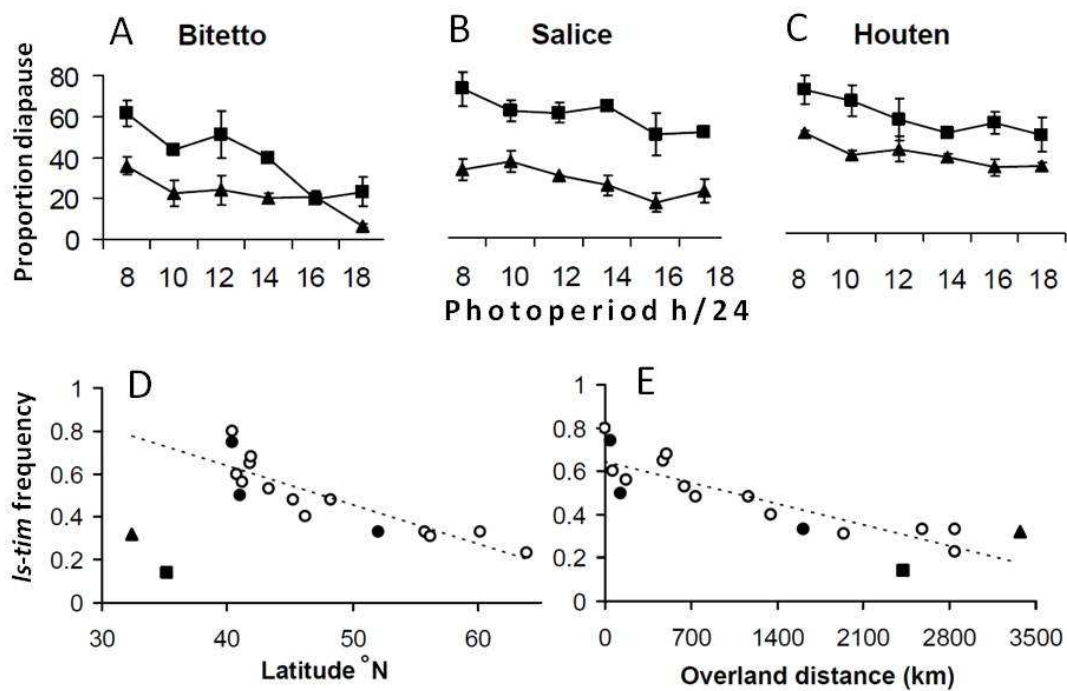
Figure 1.9
Top panel: diapause level of *ls-tim* (squares) and *s-tim* (triangles) flies, collected in three different locations (Bitetto, Salice and Huten; A-C) and tested at 6 photoperiods. Bottom panel: *ls-tim* frequency plotted against degrees latitude North (D) and overland distance from Novoli (E). Empty circles: flies collected in 1997. Full circles: flies collected in 2004. Full triangle and square: Israel and Crete respectively. Figures from (Tauber *et al.*, 2007).

1.6.2.4 COOPERATION OR INDEPENDENCE?

A debate has been going on for many decades, on the possible involvement of the circadian clock mechanism in the photoperiodic time measurement. Such controversy started in 1959, when E. Bünning and T. Lees, supporters of opposite theories, met at the Cold Spring Harbour Symposium. E. Bünning was a German biologist who first advanced the theory now referred to as 'Bunning's hypothesis'. It states that the same molecular mechanism controlling daily-circadian phenotypes is also in charge of sensing decreases in photoperiod, triggering diapause as winter approaches (Saunders *et al.*, 2004; Saunders, 2005). Such theory is both intuitive and parsimonious, as natural selection would have had to work on a single set of molecules in order to perform two relatively similar tasks: sensing transition from light-on to light-off (and *vice versa*) on a daily basis and transition from a long to a shorter photoperiod, on a seasonal time scale. Over the years, many models have been formulated to explain how the circadian clock machinery could control photoperiodism (Nunes & Saunders, 1999).

In contrast, T. Lees was a British entomologist who conducted extensive studies on diapause, mainly focusing on the green vetch aphid *Megoura viciae* and the fruit tree red spider mite *Metatetranychus ulmi* as model organisms. He was the first supporter of the so called 'hourglass model', which proposed that circadian systems have no influence on photoperiodic time measurements. The hourglass is triggered by the transition from light-on to light-off, it 'counts' the hours of dark and if these reach the threshold value, the diapause phenotype is initiated. The hourglass is then reset at the beginning of the next dark phase (Nunes & Saunders, 1999).

In order to discriminate between the 'Bünning hypothesis' and the 'hourglass model', several experiments have been designed over the years, which mainly involve scoring the diapause phenotype after the exposure of the organisms to a range of cleverly chosen photoperiods. However the results of these experiments are not unambiguous: the outcomes

may vary when the experiments are carried out at different temperatures. Furthermore the results of different experiments might be in disagreement (Nunes & Saunders, 1999).

Another way to elucidate the role of the circadian clock in photoperiodism, is to look at the photoperiodic response of clock mutants. Saunders and colleagues determined the photoperiodic response curve of different *per* mutants and showed that they are still able to distinguish between long and short days (Saunders *et al.*, 1989; Saunders *et al.*, 1990). As regards *timeless*, flies carrying the derived allele *ls-tim* show a higher incidence of diapause in all the populations tested (see paragraph 1.6.2.3). Nevertheless, *s-tim* and *ls-tim* homozygous flies show the same photoresponsive response (they have the same capability to distinguish different photoperiods, even if they show different absolute levels of diapause) suggesting that *timeless* is involved in the diapause response, but through a pathway which is independent of the photoperiodic timer. As pointed out by Emerson and colleagues (Emerson *et al.*, 2009a), photoperiodism and circadian rhythmicity can be considered as two different 'modules': a group of functionally related genes that interact with each other to control integrated processes (Figure 1.10). The results obtained with different *tim* allelic variants suggest that *tim* has pleiotropic functions: it is involved in two modules, circadian photoresponsiveness and diapause incidence. The lack of proof supporting 'modular pleiotropy' (mutations affecting a module influence the performance of a second module functionally linked to the first one, Figure 1.10 A) suggests that circadian and photoperiodic timers are two distinct processes (Figure 1.10 B) and that *timeless* is involved in both of them through gene pleiotropy. Nevertheless, due to the lack of any definite proof supporting one theory or the other, the debate goes on.
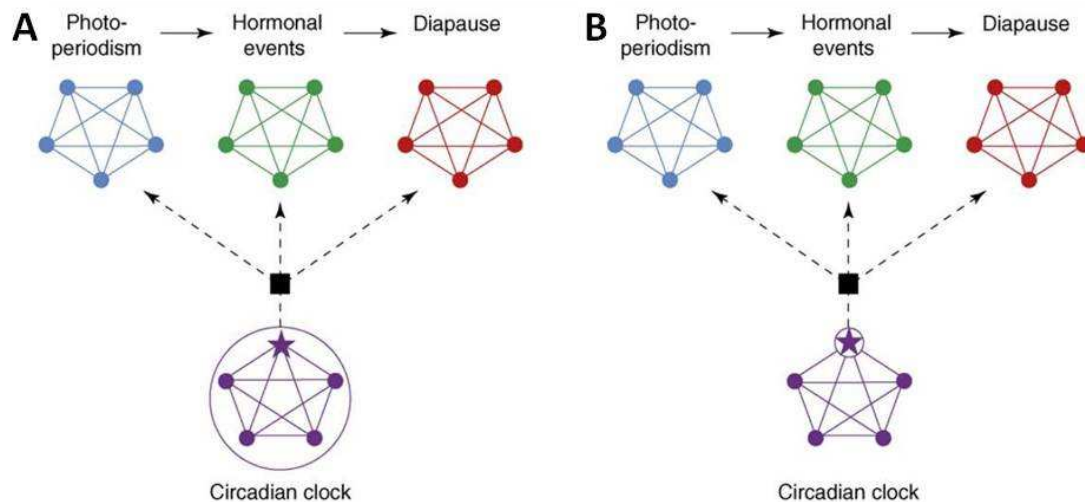
Figure 1.10
Processes leading to or influencing diapause. Single genes are represented by dots, connected to each others and organised in pentagrams, which indicate groups of functionally related genes. The star indicates a mutation in a clock gene. A) Modular pleiotropy: the overall functionality of a module (circadian clock) influences the performance of other modules (e.g. diapause). B) Gene pleiotropy: a single gene influences two processes (circadian clock and diapause) without these two having any direct effect one on the other. Figure from (Emerson *et al.*, 2009a).

1.6.3 *Couch potato*

1.6.3.1 ROLE OF *couch potato*

couch potato (*cpo*) is a gene which plays an essential role in *D. melanogaster* central nervous system (CNS) development (Bellen *et al.*, 1992b; Bellen *et al.*, 1992a). The curious name was chosen because of the general sluggishness of *cpo* viable mutants: they display slow development, increased seizure susceptibility, abnormal flight response, phototaxis, geotaxis and ether recovery (Bellen *et al.*, 1992b; Glasscock & Tanouye, 2005). The gene putatively encodes a protein which traslocates into the nucleus and is capable of binding to RNA molecules (Bellen *et al.*, 1992a). It is widely expressed: it has been detected in the peripheral and central nervous system (PNS and CNS) of embryos, larvae and adults, as well as in other tissues such as the midgut, glia, salivary glands and ring gland, the major endocrine organ in *D. melanogaster* (Bellen *et al.*, 1992a; Harvie *et al.*, 1998).

1.6.3.2 *Couch potato* ALLELES AND DIAPAUSE

The first connection between *cpo* and diapause was suggested by Schmidt and co-workers (Schmidt *et al.*, 2008) since it came up in a QTL screening performed in order to identify genes involved in regulating the diapause phenotype in *D. melanogaster*. It has been shown that *cpo* possesses Ecd binding sites, so its effect on diapause is likely to be exerted through modulations of the Insulin signaling pathway (Schmidt *et al.*, 2008). Furthermore high diapause incidence correlates with high levels of *cpo* expression, and a region at the end of *cpo* major exon was identified as being responsible for the modulation of the phenotype (Schmidt *et al.*, 2008). In particular, this region contains two non-synonymous SNPs at amino acid position 356 (Alanine to Valine) and 462 (Isoleucine to Lysine).

Given the potential relevance of these alleles in influencing seasonal adaptation to different environments, the frequency of the *cpo* variants has been studied in natural *D. melanogaster* flies. Along the US East Coast, frequency of SNP A356V shows a latitudinal cline with the cpo^{356Val} allele increasing in frequency with latitude (Schmidt *et al.*, 2008). The two SNPs are in strong linkage disequilibrium given their physical proximity, therefore it does not come as a surprise that at the extremes of the latitudinal range the frequency of cpo^{462Lys} was found to be 0.15 in the South and 0.61 in the North, suggesting that also SNP I462K is characterised by a latitudinal cline (Schmidt *et al.*, 2008). Furthermore the SNP I462K alone accounts for all the variance in diapause observed in the tested line, suggesting that it is that position, rather than SNP A356V, to be modulating the phenotype (Schmidt *et al.*, 2008). Interestingly, at the time this project was started, the database FlyBase (http://flybase.org/) reported 6 *cpo* transcripts, only one of them containing SNP I462K, but that transcript lacked the RNA binding domain (Chapter 4, Figure 4.2).

The frequency of SNP I462K in natural flies collected along the eastern Australian coast was also analysed, in an attempt to confirm the adaptive relevance of the cline. Lee and

colleagues found that the frequency of *cpo*^{462Lys} increases with latitude also in Australia, very similar to what happens in the USA (Chapter 4, Figure 4.2) (Lee *et al.*, 2011). Nevertheless this cline seems to be merely driven by the linkage disequilibrium between the SNP itself and the Payne chromosomal inversion (Lee *et al.*, 2011), which shows a latitudinal cline in Australia (Knibb, 1982). Moreover, unlike American populations, in Australian ones SNP I462K does not seem to be involved in regulating the diapause phenotype. The experiments aimed at shedding light on the relative contribution of the two SNPs and in their European geographical distribution are described in Chapter 4.

1.6.4 *Insulin-like Receptor*

The importance of the Insulin signaling pathway and the essential role it plays in diapause/reproduction has been described earlier. Paaby and coworkers focused their attention on two genes involved in the early steps of the cascade, namely *chico* and the *Insulin-like Receptor* (*InR*). Even though natural SNPs were identified in the *chico* sequence, none of them seems to present an interesting latitudinal cline (Paaby *et al.*, 2010). More interesting is the *scenario* which emerged for *InR*, in which a polymorphic region was identified, characterised by 6 indel polymorphisms. Interestingly, the frequencies of two of these alleles present very similar latitudinal clines (in terms of direction and steepness) in North America and Australia (Paaby *et al.*, 2010). To further suggest a role of these alleles in modulating the diapause phenotype, they were shown to confer phenotypes consistent with their geographical distribution: lines homozygous for the allele present at high frequencies at high latitudes, recover faster from oxidative stress and chill coma, and lay fewer eggs than flies homozygous for the other allele (Paaby *et al.*, 2010). This result indirectly suggests that the indel might be modulating the expression of the Insulin signaling pathway, downregulating it at northern latitudes thus conferring an evolutionary advantage. Chapter 5 describes the

experiments performed to analyse the frequency of these *InR* alleles in our set of European *D. melanogaster* lines.

1.7 PHENOTYPE ANALYSED

Diapause in *D. melanogaster* is scored by looking at the developmental stage of the ovary. Flies are dissected and, according to Saunders *et al.*, 1998, a fly is considered in diapause if its most advance follicle is previtellogenic. This protocol offers the possibility to perform the experiment on any fly line, but presents a few disadvantages. First of all the scoring is never fully objective, which makes data comparison difficult and, at times, inaccurate. The process of dissecting the ovaries and finding the most developed follicle can be very slow and therefore time consuming. But the most important disadvantage is the loss of information which happens upon converting a continuous value (ovary development) in a discrete one (diapause or reproductive state). Two samples which show the same percentage of diapause induction might hide profound differences at the level of actual ovary development. One way to overcome this problem is to score the precise developmental stage of the follicles rather than simply drawing a line between pre- and post-vitellogenic. This is the approach adopted for instance by Lee and colleagues (Lee *et al.*, 2011) which solves the loss-of-information issue but makes the whole procedure much more lengthy. We think a new and reliable tool for diapause studies should be developed, to make the ovary scoring process faster and more objective. Such a tool could take advantage of reporter genes, put under the control of relevant promoters which would switch on their transcription only when vitellogenesis occurs. Such an approach would also allows the estimation of the level of vitellogeneis, rather than merely classifying flies as in diapause or non-diapause.

1.8 WHY BOTHER?

The work described in this thesis was carried out using *D. melanogaster* as a model organism. This might look like a disputable choice, given the weak seasonal response expressed by this species (Saunders *et al.*, 1989). Other than many advantages that *D. melanogaster* shares with other insect model systems (like the short generation time and the reasonably cheap costs associated with maintaining the fly stocks), its long history as a genetics model organism provides us with unmatched genetic and molecular tools.. First of all vast collections of stocks are available: these include deficiencies, mutants generated by P-element insertion, chemical treatment or radiation exposure, and transgenic lines for tissue-specific gene down-regulation or transgene expression. Furthermore the *D. melanogaster* genome has been fully sequenced. User-friendly and well annotated online platforms and databases are available and regularly used by the fly genetics community.

In addition, one might wonder why it is relevant to study the seasonal phenotype in a fruit fly, and more specifically the genetic adaptation of the phenotype to different environments. The first reason is one out of scientific curiosity. This project offers the possibility to observe evolution in action, while it shapes a complex phenotype in time and in space. Secondly, the presence of clines driven by environmental conditions, and the possibility to follow them over time, can provide us with an insight on the extent of climatic change and global warming (Bradshaw & Holzapfel, 2001; Anderson *et al.*, 2005; Umina *et al.*, 2005), and its impact on the eco-system.

Last but not least, even human behaviour and physiology are affected by the changing photoperiod. As we all experience regularly, our appetite, mood, weight, sleep length and fertility change throughout the year. In some cases, especially in extreme environmental conditions (Rosen *et al.*, 1990), the extent of these cycles can become pathological. The

correlation between seasons and depression, suicide and suicide attempts has been known for years (Kevan, 1980) but the syndrome of Seasonal Affective Disorder (SAD) was first described in 1984 (Rosenthal *et al.*, 1984). It is characterised by symptoms such as depression, irritability, increased sleep length and weight gain, which recur annually in wintertime. SAD is triggered by the lack of light exposure (Molin *et al.*, 1996) and many studies showed that it can be alleviated by light treatment (Rosenthal *et al.*, 1984; Stewart *et al.*, 1991). Interestingly, variations in human clock genes have been found to correlate with human fertility, seasonality and SAD (Kovanen *et al.*, 2010; Partonen *et al.*, 2007). Studying the molecular mechanism of the seasonal clock in the fruit fly, and the way it adapts to the environment, can provide us with better understanding of how the human seasonal clock works, and can contribute in developing new techniques to deal with season-related health issues.

2. MATERIALS & METHODS

2.1 FLY STRAIN HUSBANDRY

2.1.1 NATURAL FLY COLLECTIONS

The Spanish lines used in this thesis were collected in early October 2008. The collection technique varied according to substrate and of flies available. In the arid southern Spanish regions flies were individually captured from rotten fruits (mainly prickly pears, apples and peaches) by means of a mouth aspirator. In the central/northern regions flies were much more abundant, due to the late grape harvest season and flies were mainly collected from the winemaking waste, piled up in the premises of the wineries. Due to their abundance, flies could be easily collected by means of sweep nets. In both cases, upon their collection flies were promptly transferred to plastic vials containing maize food (72 g/L maize meal, 80 g/L glucose, 50 g/L brewer's yeast, 8.5 g/L agar , 2 g/L of Nipagine [methyl p-hydroxybenzoate], dissolved in 10 ml 100% ethanol). Once in the laboratory males were fixed in 100% Ethanol and kept at -80°C. Single gravid natural females were instead moved to individual vials, and their progeny were allowed to interbreed. This process generates isofemale lines which can be kept in the laboratory and be used for behavioural experiments, while maintaining their natural genetic variation.

After the males had been frozen and the isofemale line had been set up, flies were screened to separate *D. melanogaster* from its sibling species, *D. simulans* by examining the morphology of the male genitalia. Since male flies were fixed in ethanol and the visual discrimination was difficult, a species-specific PCR was established, with primers designed in a region of *timeless* which differs between *D. melanogaster* and *D. simulans* (5'-*TTCACCTCTGGTACCACGTTAAG*-3' and 5'-*AATGGCTGGGGATTGAAAAT*-3', nucleotide coordinates 3504651..29 and 2304295..314 respectively). Figure 2.1 shows the ClustalW alignment between the *D. simulans* (sim) and the *D. melanogaster* (mel) DNA regions considered for the species-specific PCR. Primer sequences are indicated in red. All the males and a single fly from each isofemale line were then screened *via* species-specific PCR.

```
sim        AATACATATATTTGGTACCATTTCACAACGTATTTAAATGTGTCATACGTAACAAATATT 60
mel        AGTTTAAATGTGTAATATTATTTAAATTCACCTCTG-----GTACCACGTTAAGAATATT 55
           *  *   * ** *  *   **   **** *    *   * *       **    **** *  ******

sim        TGACTTTTAACAACAATTAAATTAATATTTTTATATGCTTGTCTTTACAGTTACCGCGCA 120
mel        TAAGTTTTAATAACAATTAAATATACATTTTTGTTTACTTATAGTTACAGATACCGCGCA 115
           * * ****** **********  * ****** * * *** *  ****** ********

sim        AATTGCTAAGAAGTACCTCAATGTTCACAGTCGACAATGAGCAGAGTTAGGCAGCTCAAC 180
mel        AATGGCTAAGAAGTACCTCAATGTTCGCAGTCGACAATGAGCAGAGTTAGGCAGCTCCAC 175
           *** ********************** ***************************** **

sim        AATCACATCTGGAATAATCAGAACTTT-ATCAAGTGAAATCGGTTATGGACTGGTTACTA 239
mel        AATCACATCTGGAATAATCAGAACTTTGATAAAGTGAAATCGGTTATGGACTGGTTACTA 235
           *********************** ** ** *************************

sim        GCAACTCCGCAGTTGTACAGCGCGTTCTCCTCCTTGGGTTGCTTGGAGGGCGATACCTAT 299
mel        GCAACTCCGCAGTTGTACAGCGCGTTCTCCTCCTTGGGTTGCTTGGAGGGCGATACCTAT 295
           *************************************************************

sim        GTGGTCAACCCGAATGCATTGGGTAAGTAGTGAACATAGGA-TACTACATCATTTCGATT 358
mel        GTGGTCAACCCGAATGCATTGGGTGAGTAATCAATATAGGACTACTACATTAATTAGGTA 355
           ************************ **** * ** ****** ******** * ** * *

sim        TTTAACTTCTTTTCAATCCTCAGCCATTCTGGAGGAGATCAACTACAAGCTCACCTATGA 418
mel        TTTAACTTATTTTCAATCCCCAGCCATTCTGGAGGAGATCAACTACAAGCTCACCTATGA 415
           ******** ********** *****************************************
```

Figure 2.1
ClustalW alignment between the *D. simulans* (sim) and the *D. melanogaster* (mel) DNA regions considered for the species-specific PCR. The position of the primers is indicated in red.

The DNA from the frozen males was used in the gene frequency studies. In some cases in order to increase the sample size, the DNA from one male from each isofemale line, extracted within two generations from collection, was included in the analysis.

2.1.2 FLY HUSBANDRY

Fly lines were maintained in plastic vials containing maize food (for recipe see paragraph 2.1.1). The vials were kept in temperature controlled rooms, at either 25°C or 18°C. *D. melanogaster* developmental time strongly depends on temperature. Flies to be used for imminent experiments were kept at 25°C, which allows a short developmental time, whereas flies were kept at 18°C for long-term maintenance. Flies were subjected to a 12 hour light-dark regime (LD 12:12).

2.1.3 FLY LINES

Details regarding the site of collection of the natural flies used in this project are provided in each chapter. The following list comprises all the natural flies used with details regarding source and date of collection.

- LINES FROM SPAIN: Used in chapters 3, 4, 5 and 6. Collected in October 2008.

- LINES FROM UK, DENMARK, SWEDEN AND FINLAND: Used in chapters 3, 4 and 5. Kindly provided by E. Tauber. Collected in Spring 2009.

- LINES FROM TREVISO (ITALY): Used in chapters 3, 4, and 5. Collected in Spring 2009 by S. Vanin.

- DNAs FROM OTHER EUROPEAN LOCATIONS: Used in chapter 4 and 5. Kindly provided by E. Tauber. Collected in October 2004.

- LINES FROM USA: Used in chapter 3, 4 and 6. Kindly provided by P. Schmidt. Collected in 2009.

- LINES DROM AFRICA (*D. simulans* and *D. melanogaster*): Used in chapter 5 and 6. Kindly provided by C. Shlötterer. Collected in 2001 and 2006 (*D. simulans*) and in 2008-2009 (*D. melanogaster*).

2.2 DNA EXTRACTION

2.2.1 WHOLE BODY DNA

Genomic DNA of *D. melanogaster* whole flies was extracted according to Gloor *et al*. (Gloor *et al.*, 1993). Single flies were anaesthetised and placed in individual 0.2 ml PCR tubes. Flies were then smashed with a yellow tip for 5-10 s, in 50 µl of Squishing Buffer (10 mM Tris-HCl pH 8.2, 1 mM EDTA, 25 mM NaCl, 200 µg/ml Proteinase K, diluted fresh from a frozen stock). The samples were then incubated for 1 h at room temperature and finally heated to 95°C for 2 min to allow the inactivation of the Proteinase K. 1 µl of this preparation was normally used in a 10 µl PCR reaction volume.

2.2.2 FLY WING DNA

When flies needed to be genotyped in order to be crossed later (e. g. to create the lines used in Chapter 4), genomic DNA was extracted from the fly's wings. Wings were gently removed from anesthetised flies by means of fine tweezers. In order to maximise the yield and concentration on DNA, wings were incubated overnight at room temperature in 20 µl (instead of 50) of Squishing Buffer. Furthermore 3X the normal concentration of Proteinase K was added to the reaction. This was then heat inactivated as described above.

2.3 RNA EXTRACTION

Flies were collected in a 1.5 ml tube at the desired time points and snap frozen in liquid nitrogen. The samples were repeatedly vortexed and frozen, and finally the heads were removed from the bodies by means of a fine metal sieve kept over dry ice. They were then

homogenised in 500 µl of TRIzol® (Invitrogen) using a plastic pestle. The homogenates ware allowed to stand for 5 min at room temperature before the addition of 200 µl chlorophorm. The mixtures were shaken vigorously for 15 s and left to stand for 10 min before being centrifuged in a microfuge for 15 min at 4°C. The upper phase was then carefully moved to a new RNase-free tube. This was filled up to 1.5 ml with isopropanol and left to stand at room temperature for 5-10 min. The precipitated RNA was collected by centrifugation for 15 min at 4°C. The RNA pellet was washed with 80% ethanol and then resuspended in 20 µl of diethylpyrocarbonate (DEPC) treated water. The quality and quantity of the isolated RNA were checked by electrophoresis on a 0.7% agarose gel.

2.4 GEL ELECTROPHORESIS

The extracted RNA and the amplified PCR fragments were analysed in agarose gels containing Ethidium Bromide (0.5 µg/ml). Agarose (0.5%-2% depending on the size of the fragment to be visualised), was melted in 0.5X TBE (for 1 L 5X: 54 g Tris base, 27.5 g Boric Acid, 20 ml 0.5 M EDTA pH 8.0). Normally 2 µl of 6X Loading Dye (Bromophenol Blue 0.25%, Xylene Cyanol 0.25%, Sucrose 40%) were added to 10 µl of PCR reaction and applied to the Agarose gel. Gels were run at 100 V in tanks containing 0.5X TBE, until the samples had reached the desired distance from the wells. The DNA was visualized under UV light and the images elaborated by Gene Genius from Bio Imaging System. The molecular dimension of the DNA fragments was determined by comparison with fragments from a known molecular weight marker loaded alongside the samples.

2.5 DNASE TREATMENT

The RNA extracted as described above was mixed with 0.2 volume of 10X TURBO™ DNase Buffer and with 2 µl of TURBO™ DNase (Invitrogen). The samples were the incubated at 37°C for 30 min. 0.2 µl of resuspended DNase Inactivation Reagent were added to the solution and this was incubated at room temperature for 5 min in a table shaker. Finally, the samples were centrifuged for 1.5 min at 10,000xg and the supernatant was carefully transferred into a fresh tube.

2.6 CDNA SYNTHESIS

1 µg of the DNase treated-RNA solution was mixed with 0.5 µg of oligo d(T) primers. The samples were incubated at 72°C for 5 min and then chilled on ice for 5 min. To this, the following reagents were added: 4 µl of 5X Reaction Buffer, 2.4 µl of 25 mM $MgCl_2$, 1 µl of dNTPs mix (10 mM each), 1 µl of Rnasin Ribonuclease Inhibitor (Promega®), 5.6 µl of DEPC treated water.

These were mixed before adding 1 µl of Improm-II Reverse Transcriptase (Promega®). The reaction was incubated on a thermocycler at 25°C for 5 min, followed by 42°C for 1 h; the reaction was then terminated by heat inactivation at 70°C for 15 min. The products were used directly for PCR amplification.

2.7 PCR REACTION

In order to amplify genomic DNA (or cDNA), primers were designed using the Software Primer3 version 0.4.0 (Rozen & Skaletsky, 2000), which optimises the primer and amplicon lengths, the primers melting temperature, GC content and self and 3' complementarity.

PCRs were performed by setting up the following reaction mix: 0.9 µl of 11.1.X PCR Magic Buffer , 0.2 µl of Forward Primer (10 µM), 0.2 µl of Reverse Primer (10 µM), 0.1 µl of Taq Polymerase (5 Units/µl), 1 µl of DNA (obtained as described above), 7.6 µl of Sterile double distilled water.

The 11.1X PCR Magic Buffer was preprared as follows: 45 mM Tris-HCl pH 8.8, 11 mM Ammonium sulphate, 4.5 mM Magnesium chloride, 6.7 mM 2-Mercaptoethanol, 4.4 µM EDTA pH 8.8, dNTPs 1mM each, 113 µl/ml BSA. The general PCR program used for the amplification was the following:

INITIAL DENATURATION	92°C		2 Minutes
START CYCLES			
DENATURATION	92°C	30 sec	
ANNEALLING	-	30 sec	Repeat (30-40) times
EXTENSION	72°C	-	
END CYCLES			
FINAL EXTENSION	72°C		10 Minutes
STORE	10°C		Indefinitely

Each new couple of primers was initially run using an annealing temperature gradient in order to identify the best conditions, namely the temperature which would provide maximum product yield, and minimum unspecific amplification. The extension time varied according to the length of the final product (1 min per kb). The number of cycles was optimised in each PCR, and normally varied between 30 and 40.

After the amplification the PCR products (or part of them) were run on an appropriate percentage agarose gel in order to check their size, purity and quantity.

2.8 RESTRICTION ENZYME REACTION

Flies were genotyped for *cpo* (see Chapter 4) according to Schmidt and colleagues (Schmidt *et al.*, 2008). After the PCR reaction, 5 µl of product were run in an agarose gel, in order to make sure that the amplification reaction worked successfully. The remaining PCR products were incubated at 60°C for 90 min with 0.1 µl of *BsiE*I enzyme (New England *BioLabs®Incs.*), 3 µl of New England *BioLabs®Incs* Buffer 4, 0.3 µl of BSA 100 µg/µl and 16.6 µl of H_2O. The result of the digestion was then checked on an appropriate percentage agarose gel.

2.9 DNA PURIFICATION AFTER PCR

Any dNTPs unconsumed in the PCR amplification, and primers remaining in the product mixture will interfere with the sequencing reaction. Therefore, prior to sequencing, PCR products were purified with the ExoSAP-IT® product, which allows fast PCR clean-up without any loss of amplified DNA. It contains two hydrolytic enzymes, Exonuclease I and Shrimp Alkaline Phosphatase, in a specially formulated buffer. Exonuclease I removes residual single-stranded primers and any extraneous single-stranded DNA produced in the PCR. Shrimp Alkaline Phosphatase removes the remaining dNTPs from the PCR mixture. 2 µl of ExoSAP-IT® were added to 5 µl of PCR product, and incubated at 37°C for30 min. The enzyme was then heat-inactivated by incubating the samples at 95°C per 5 min. The last step prevents the

enzyme from degrading the sequencing primer which will be later added to the reaction. Before setting up the sequencing reaction, the purified products were quantified in a Thermo Scientific Nanodrop 8000 Spectrophotometer.

2.10 DNA SEQUENCING

Sequencing reactions were carried out in our laboratory using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystem). The sequencing mix was prepared as follows: 1 µl of Big Dye®, 3.5 µl of Sequencing Buffer, 4 µl of primer 1pmol/µl, 20-50 ng of purified PCR product (depending on its length), water up to 20 µl.

The reaction was carried out with the following thermal cycle:

INITIAL DENATURATION	96°C		1 Minute
START CYCLES			
DENATURATION	96°C	10 sec	
ANNEALLING	50°C	5 sec	Repeat 30 times
EXTENTION	60°C	4 min	
END CYCLES			
STORE		10°C	Indefinitely

The samples were then purified using Edge Bio Performa® Gel Filtration Cartrages, in order to get rid of the unincorporated dNTPs, and finally sent to the Protein and Nucleic Acid Chemistry (PNACL, University of Leicester) to be run on an ABI 377 sequencer.

2.11 DNA FRAGMENT ANALYSIS

DNA fragment analysis was used to genotype natural flies for the indel polymorphism in the *InR* gene, and was performed as in Paaby *et al.*, (2010). The region of interest was amplified, the forward primer being 6-FAM-fluorescent-tagged. After having checked the success of the amplification by running part of the sample in a 1.5% agarose gel, the rest of the PCR reaction was subjected to A-tailing. This procedure assures that all the amplified molecules are characterised by the extra *A* (adenine) added by the Taq polymerase at the end of transcription. This is achieved by incubating 9 µl of PCR product with 0.05 µl of Taq, 0.4 µl of PCR buffer and 3.55 µl of water. The mixture was incubated at 72°C for 45 min.

Just before being run in a Applied Biosystems 3730 sequencer at the PNACL (Protein Nucleic Acid Chemistry Laboratory) facilities of the University of Leicester, the samples were diluted 1:10 and the following mix was prepared: 9.25 µl of formamide, 0.25 µl of GeneScan™ 500 LIZ ® Size Standard and 0.5 µl of diluted-A-tailed PCR product. Finally, the results were analysed with the Peak Scanner Software v1.0.

2.12 OVARY DISSECTION

2.12.1 POPULATION SET-UP

In order to assess the diapause incidence of a natural population from a specific location, equal numbers of non-virgin females from different isofemale lines were combined together. Flies were reared in glass bottles with "Sugar food" (4.6% w/v Sucrose, 1% w/v Agar, 0.8% w/v dried yeast, 0.04% w/v Nipagen), kept at 25°C and moved to a new bottle every 2-3 days, in order to quickly expand the population. Glass bottles were chosen over plastic vials

because they can accommodate more flies, thus limiting genetic drift. When the population was big enough, flies were collected for the experiment. Bottles with emerging flies were emptied in the morning. After 6 h the bottles were emptied again, and all the flies which emerged in the 6 h window were collected in plastic vials and used for the experiment. Since at 25°C *D. melanogaster* flies do not mate until 8 hours post eclosion (Ashburner *et al.*, 1989), this approach allowed us to collect virgin flies. Around 60 flies were collected in the same vial and exposed to diapause-inducing-conditions.

2.12.2 DIAPAUSE INDUCING CONDITIONS

Diapause was induced by exposing flies to 12°C, at two different photoperiods, LD 8:16 (winter) and LD 16:8 (summer). Plastic vials were placed in fluorescent-tube-light boxes, controlled by timers. The light boxes were kept in incubators, which allowed the temperature to be controlled. In each incubator a Temperature log was placed inside one of the light boxes, in order to check the temperature fluctuations due to light-on and light-off.

Figure 2.2 shows a typical temperature profile measured in a LD 16:8 light box, with each day having a temperature oscillation in the range of 0.4-0.5°C. 4-6 replicates were performed for each condition. In order to verify if potential differences in diapause incidence between the two photoperiods were due to the exposure to two slightly different temperatures (see Figure 2.2), at least two DD (Constant darkness) controls were performed in each condition. Vials were prepared as described above, but wrapped in aluminum foil before being placed in the light boxes. Vials were kept in diapause-inducing-conditions for 12 or 28 days.
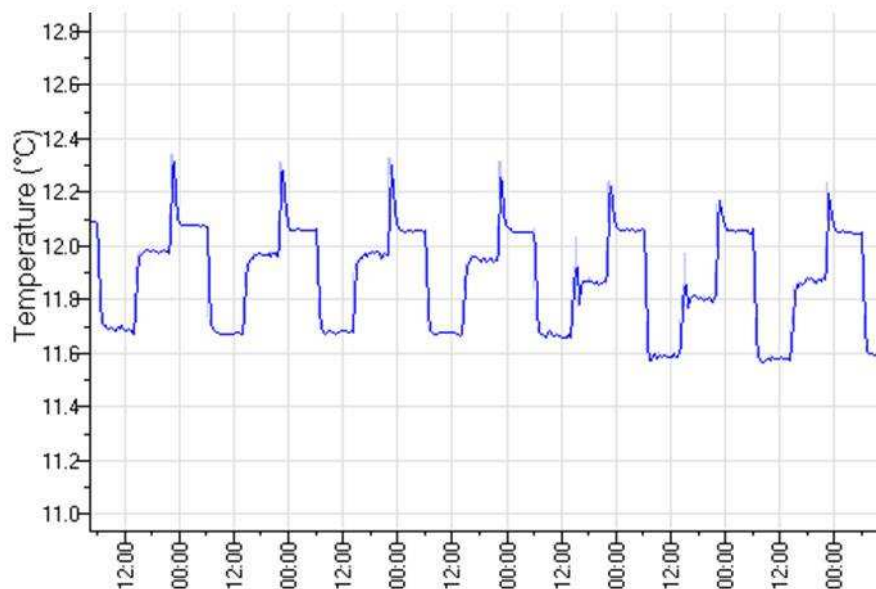
Figure 2.2

Temperature profile recorded over 7 days inside one of the light boxes kept at 12°C (LD 16:8). The first temperature peak occurs at light on in the light box where the sensor was kept. The second temperature peak is due to the lights switching on in the light box adjacent to the one being monitored, and having a winter light-dark cycle (LD 8:16). The average temperature difference between the two conditions (LD 16:8 and LD 8:16) is 0.2°C.

2.12.3 DISSECTIONS

After 12 or 28 days in diapause-inducing-conditions, the vials were retrieved and the flies anesthetised by cold exposure (4°C). Around 30 females per replicate were then dissected in Phosphate-Buffered Saline, PBS, prepared according to (Wulbeck & Helfrich-Forster, 2007). Diapause was scored according to the developmental stage of the most developed oocyte, as described by Saunders and colleagues (Saunders *et al.*, 1989). When all the oocytes, from both ovaries, were previtellogenic (stages 1-7, Figure 2.3), the individual was considered to be in diapause. If even one oocyte was more developed (stage >7), the fly was scored as "non diapausing". Figure 2.3 shows the 14 egg developmental stages.
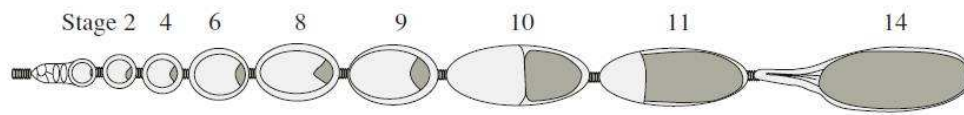
Figure 2.3
Egg chamber development is divided in 14 stages, based on morphological differences. The oocyte is shown in gray. Stages before 8 are classified as previtellogenic. Figure from Ogienko and coworkers (Ogienko *et al.*, 2007).

The results of diapause experiments in this thesis are always shown as a percentage of diapausing flies but arcsine square root transformation was applied to the dataset before any statistical tests.

3. TIMELESS

3.1 INTRODUCTION

As mentioned in Chapter 1, *tim* is a key component of the circadian clock. A natural allele, *ls-tim* has been discovered and compared, molecularly and phenotypically, to the original allele *s-tim* (Rosato *et al.*, 1997; Tauber *et al.*, 2007; Sandrelli *et al.*, 2007). The table below highlights the main differences between the two alleles.

ALLELE	PROTEIN LENGTH (AA)	CRY INTERACTION	TIM DEGRADATION	PHOTOSENSITIVITY	DIAPAUSE LEVEL
ls-tim	1421/1398	Weak	Slow	Weak	High
s-tim	1398	Strong	Fast	Strong	Low

Table 3.1
Differences between the two *timeless* alleles.

Tauber *et al.* (2007) analysed the frequency of the new allele *ls-tim* in European *D. melanogaster* populations from southern Italy to Sweden and found that, counter-intuitively, *ls-tim* decreases in frequency with latitude. Given that *ls-tim* increase diapause incidence, this allele was expected to be found in higher frequencies at northern latitudes, where winters are more severe and start earlier in the year. Nevertheless, the negative correlation between *ls-tim* frequency and latitude was disrupted by two outliers from areas further South than Italy, a population collected in Israel and one collected in Crete. Interestingly, when the *ls-tim*

frequency data were plotted against overland distance from the putative site of origin (namely the location where the *ls-tim* frequency is the highest, Novoli, southeastern Italy), a significant and robust distance cline was observed (Figure 1.9). This finding led the authors to formulate the hypothesis that *ls-tim* originated in the South of Italy and then spread throughout Europe thus generating the observed distance cline. Nevertheless three different processes could have caused the peculiar geographical distribution of *ls-tim*, namely:

DIRECTIONAL SELECTION: one allele is advantageous over the other(s), and slowly increases in frequency over time until eventually it reaches fixation. In this case *ls-tim* would be advantageous over *s-tim* at all latitudes and the present situation would represent a mere evolutionary transient before the ancestral allele *s-tim* is completely taken over by the new, more adaptive variant (blue arrows in Figure 3.1).

BALANCING SELECTION: two (or more) alleles are maintained in the population at intermediate frequencies. Balancing selection can be maintained by heterozygote advantage, frequency-dependent selection, variation in time and/or space. In this *scenario ls-tim* would confer an evolutionary advantage in the South of Italy, where it reaches high frequencies. In contrast, flies carrying the *s-tim* allele would be more adapted in the North of Europe (red arrows).
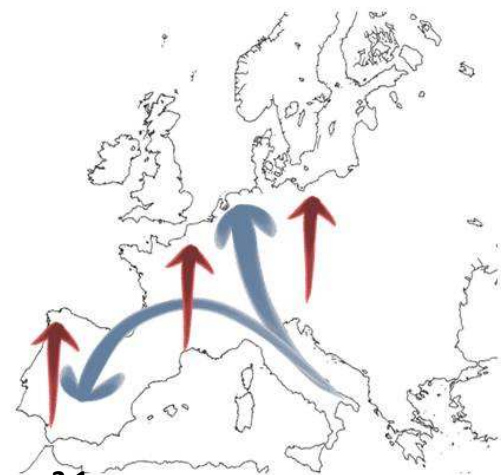
Figure 3.1
Schematical representation of the two selective *scenarios* which might have led to the geographical distribution of *ls-tim* in Europe. Blue arrows: directional selection; Red arrows: balancing selection. See text for further explanation.

GENETIC DRIFT: the observed distribution of *ls-tim* in Europe might be the result of genetic drift. After its first appearance, *ls-tim* might have increased in frequency merely because of random sampling. This *scenario* rules out the action of natural selection and attributes the reason of *ls-tim* cline to chance.

From Figure 3.1 we can see that for the Iberian peninsula, a balancing selection *scenario* would be supported if a pattern of *ls-tim* frequency similar to the Italian peninsula was observed (i.e. high levels in the South and lower levels in the North). On the other hand, if the 'directional selection by diffusion' model based on distance from the origin is supported, an opposite pattern should be observed in Iberia. Thus I sought to disentangle latitude and distance from putative site of origin by analysing the frequency of the new allele in Spain. I further investigated the evolutionary history of the new allele by performing neutrality tests on the region of interest and analysing the frequency of neutral alleles at loci scattered around the *D. melanogaster* X and second chromosomes.

3.2 MATERIALS AND METHODS

3.2.1 FLY LINES

In this work natural *D. melanogaster* lines were used. Methods of collections and husbandry are reported in Chapter 2. To perform the analysis of the geographical distribution of *ls-tim* in Spain, where possible, males collected directly from the wild and kept in EtOH 100% at -20˚C were used. In some cases, in order to increase the sample size, one male per isofemale line was also included in the analysis.

3.2.2 *ls/s-tim* GENOTYPING

ls/s-tim genotypes were determined using the same allele-specific-PCR approach as in Tauber *et al*. (2007). Two PCRs were performed for each DNA sample, using a forward primer specific either for *ls-* or *s-tim* (*5'-TGGAATAATCAGAACTTTGA-3'* and 5'-*TGGAATAATCAGAACTTTAT-3'* respectively, nucleotide coordinates: 3504493..74), together

with a common reverse primer (*5'-AGATTCCACAAGATCGTGTT-3'*, nucleotide coordinates: 3503801..20). The following primers, amplifying a syntenic region, were added to the reaction to control the PCR efficiency: *5'-CATTCATTCCAAGCAGTATC-3'* and *5'-TATTCATGAACTTGTGAATC-3'* (nucleotide coordinates: 3501740..21 and 3501254..73 respectively). The following thermal profile was used to carry out the amplification: 35 cycles with 92°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec.

3.2.3 NEUTRALITY TESTS AND PHYLOGENY

Around 70 individuals, collected in three Spanish locations (SP22, SP35, SP53) were used for the analysis. To make flies hemizygous for the *timeless* locus, virgin females from each isofemale line were crossed to the deficiency strain *DF(2L)tim^{02}/CyO*, obtained from the Bloomington Stock Center. Three regions in the second chromosome were amplified and sequenced with the primers listed in Table 3.2. The sequences were analysed with the software Geospiza FinchTV Version 1.4.0 (Geospiza, Inc.; Seattle, WA, USA; http://www.geospiza.com), and the Neutrality Tests were carried out with the program DNAsp version 5.10.01 (Librado & Rozas, 2009).

The phylogenetic tree was realised using the software MEGA version 4.0.2 (Tamura *et al.*, 2007).

	Primer Name	PCR Primers (5'-3')	Sequencing Primer (Name and Sequence 5'-3')	Annealing Temperature
5'	5-F	TCCCAACTGGAGAGGTATGC (3525005..4986)	*5-Seq* AGCTTTTGGCTGAGCTTGAT (3524974..55)	60.5°C
	5-R	GCAGCTCAACAGCGTTTGTA (3524178..97)		
LS	Tim-F1	TGATGTGGAATGGGAGTTG (3504919..01)	*Tim-Seq* GCAGCACATTGCATGGACTG (3504869..50)	60°C
	Tim-R1	GGTGAGCTTGTAGTTGATCTCCTC (3504268..91)		
	Tim-F2	CGGTTATGGACTGGTTACTAGC (3504462..41)	*Tim-R2* AGATTCCACAAGATCGTGTT (3503801..20)	55.3°C
	Tim-R2	AGATTCCACAAGATCGTGTT (3503801..20)		
3'	3-F1	GAAGCGGACATCTTAGCTGAT (3478626..06)	*3-F1* GAAGCGGACATCTTAGCTGAT (3478626..06)	56°C
	3-R1	CCTCTTGACTCCCAGACAGC (3477937..56)		
	3-F2	GTTAGCCAACAGCACTAAACG (3478067..47)	*3-R2* CTGCTTGGATCCACCCTATG (3477312..31)	60°C
	3-R2	CTGCTTGGATCCACCCTATG (3477312..31)		

Table 3.2

List of the primers used to amplify the region including the polymorphism in *timeless* (LS) and the upstream and downstream region (5' and 3' respectively). The table also reports the annealing temperature used in the PCR reaction and the primer used in the sequencing reaction. In brackets the nucleotide coordinates of the primers used.

3.3 Results

3.3.1 Frequency of *ls-tim* in Spain

The flies collected in Spain were genotyped for *ls/s-tim*. Figure 3.2 shows an example of the outcome of a genotyping reaction. Table 3.3 reports the information regarding the location of collection, the number of individuals analysed and the allele and genotype frequencies. As shown in Table 3.3, after applying the Bonferroni correction for multiple comparisons, all the populations were in Hardy Weinberg Equilibrium.
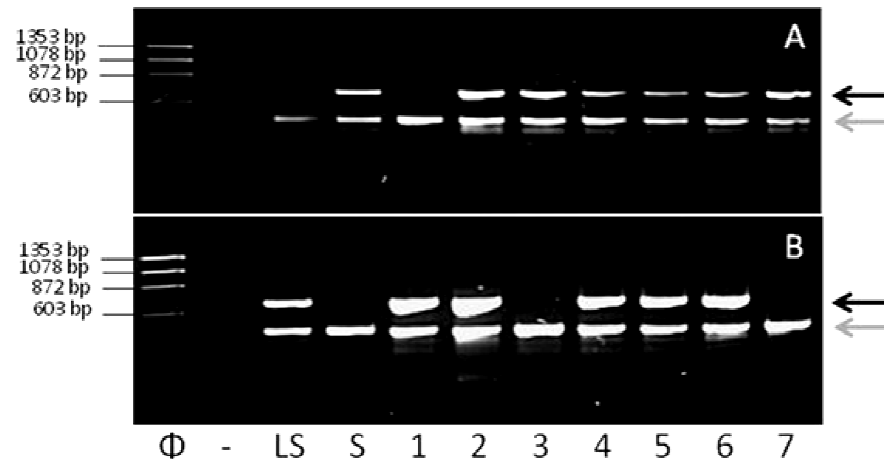
Figure 3.2

1.5% Agarose gel showing an example of *ls/s-tim* genotyping reaction. A) Master mix specific for *s-tim*. B) Master mix specific for *ls-tim*. Black arrows: specific bands (693 bp) for *s-tim* (A) or *ls-tim* (B). Gray arrows: control bands (488 bp). The genotypes of the samples are 1: *ls/ls*; 2: *ls/s*; 3: *s/s*; 4: *ls/s*; 5: *ls/s*; 6: *ls/s*; 7: *s/s*. Φ: Phi X174 *Hae III*: molecular weight marker. (-): negative control; LS: positive control for *ls-tim*; S: positive control for *s-tim*.

The *ls-tim* frequencies were then plotted both against latitude of collection and overland distance from Novoli. The plots are shown in Figure 3.3 A and B. In Spain *ls-tim* significantly increases in frequency with latitude (R^2=0.40; p=0.01) and decreases as the distance from South of Italy grows larger (R^2=0.40; p=0.01). Figures 3.3 C and D show how the genotype frequencies change with latitude and distance from South of Italy. The frequency of the homozygotes for *s-tim* decreases significantly with latitude (R^2=0.36; p=0.01), whereas the heterozygote frequency significantly increases (R^2=0.25; p<0.05). Not surprisingly, an opposite pattern is observed when the data are plotted against overland distance from Novoli (R^2=0.36; p=0.01 and R^2=0.26; p=0.04 respectively). The frequency of the homozygotes for *ls-tim* does not vary significantly with latitude (p=0.40) nor distance (p=0.44).

LINE	LOCATION	N	LAT	LONG	ALT	DISTANCE	f(ls)	f(s)	f(ls/ls)	f(ls/s)	f(s/s)	# ls/ls	#ls/s	#s/s	# ls/ls	#ls/s	#s/s	χ^2
SP-14;15	Dalias/Algarrobo	70	36.82	-2.87	423	2367	0.17	0.83	0.06	0.23	0.71	2	8	25	1.03	9.94	24.03	1.34
SP-22	Nijar	100	36.97	-2.21	345	2262	0.21	0.79	0.08	0.26	0.66	4	13	33	2.21	16.59	31.21	2.34
SP-23	Herreiras	30	37.10	-2.03	205	2242	0.23	0.77	0.00	0.47	0.53	0	7	8	0.82	5.37	8.82	1.39
SP-24	Lorca	24	37.67	-1.70	335	2172	0.17	0.83	0.08	0.17	0.75	1	2	9	0.33	3.33	8.33	1.92
SP-25;26	Abaran	52	38.21	-1.40	189	2106	0.25	0.75	0.15	0.19	0.65	4	5	17	1.63	9.75	14.63	6.17
SP-28	Jumilla	82	38.48	-1.32	520	2078	0.24	0.76	0.05	0.39	0.56	2	16	23	2.44	15.12	23.44	0.14
SP-33	Ontinyent	28	38.82	-0.61	378	2017	0.32	0.68	0.14	0.36	0.50	2	5	7	1.45	6.11	6.45	0.46
SP-35	Fontanares	116	38.78	-0.79	630	2025	0.32	0.68	0.05	0.53	0.41	3	31	24	5.90	25.20	26.90	3.07
SP-38	Requena	168	39.49	-1.10	716	1971	0.38	0.62	0.10	0.57	0.33	8	48	28	12.19	39.62	32.19	3.76
SP-41;42	Libros	62	40.16	-1.23	770	1919	0.32	0.68	0.13	0.39	0.48	4	12	15	3.23	13.55	14.23	0.40
SP-43	Alcaniz	58	41.05	-0.13	317	1784	0.29	0.71	0.10	0.38	0.52	3	11	15	2.49	12.02	14.49	0.21
SP-44	Vandeltormo	88	40.99	0.08	478	1778	0.27	0.73	0.07	0.41	0.52	3	18	23	3.27	17.45	23.27	0.04
SP-45	Gandesa	150	41.05	0.44	351	1755	0.29	0.71	0.09	0.40	0.51	7	30	38	6.45	31.09	37.45	0.09
SP-51	La Rapita	144	41.31	1.65	168	1687	0.31	0.69	0.07	0.49	0.44	5	35	32	7.03	30.94	34.03	1.24
SP-52	S. Sadurni d'Anoia	150	41.42	1.76	199	1671	0.33	0.67	0.12	0.43	0.45	9	32	34	8.33	33.33	33.33	0.12
SP-53	Manresa	162	41.73	1.82	278	1640	0.27	0.73	0.05	0.44	0.51	4	36	41	5.98	32.05	42.98	1.23

Table 3.3

Details of the Spanish flies used in the *ls/s-tim* study. N: number of alleles analysed. Lat: latitude in degrees North. Long: longitude in degrees (Negative values: West; Positive values: East). Alt: altitude in meters above sea level. Distance: overland distance from Novoli (Italy), calculated "as the crow flies" (km). The last column shows the results of the Hardy Weinberg test.
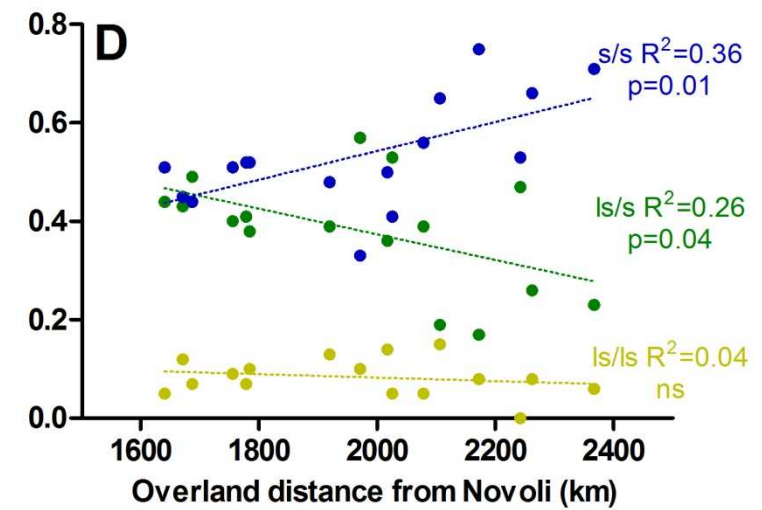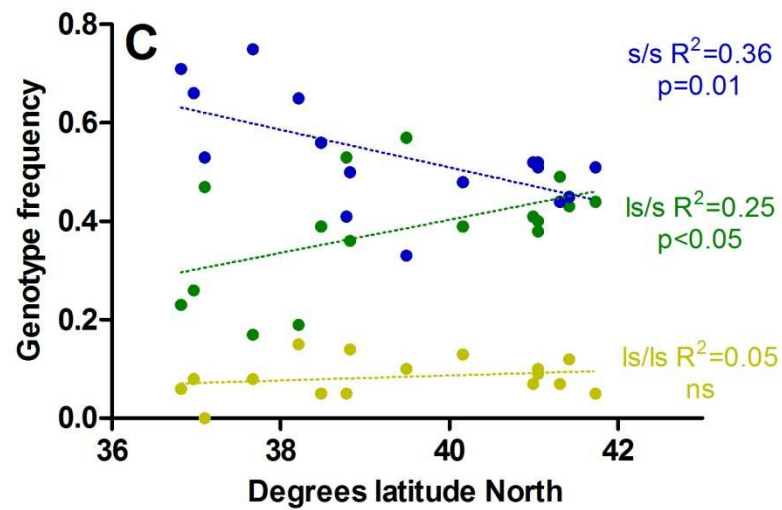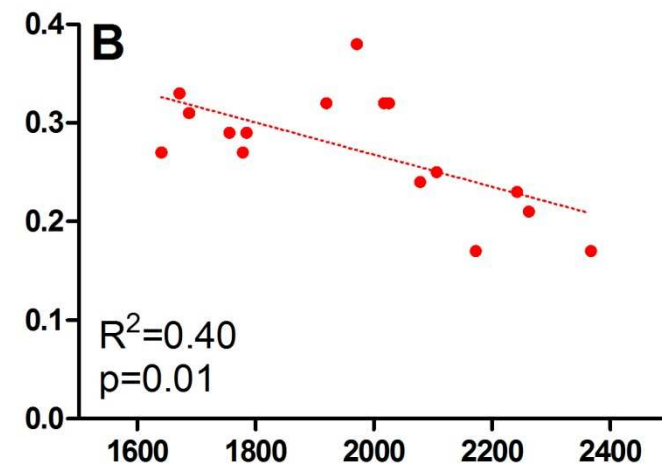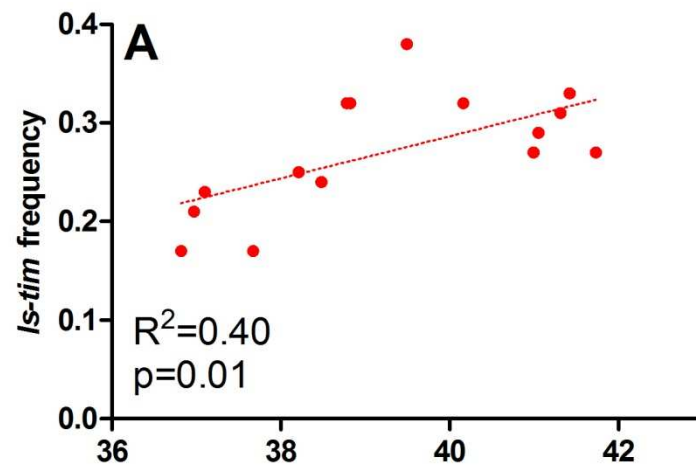
Figure 3.3

Frequency of *ls-tim* in Spanish populations, plotted against latitude (A) and overland distance from Novoli (B). Genotype frequencies over latitude and distance are shown in plots C and D respectively.

3.3.2 SPAIN VS REST OF EUROPE

The populations reported in Table 3.4 were collected by Dr Eran Tauber and were therefore genotyped in order to enrich the European dataset previously published (Tauber *et al*., 2007). Also in this case, after applying the Bonferroni correction for multiple comparisons, all the populations are in Hardy Weinberg Equilibrium.

The frequency of *ls-tim* in the new locations was plotted against latitude and overland distance from Novoli (Figure 3.4 A and B respectively) and compared to the data obtained for the Spanish populations. The latitudinal cline in *ls-tim* allele frequency follows an opposite pattern in the Iberian peninsula and in the rest of Europe. In Spain *ls-tim* is relatively low in frequency, ranging from 17% in the South and reaching 30-40% in the North. In the rest of Europe it goes from as high as 80% in the South of Italy, decreasing to as low as 20-30% in Scandinavia.

As for the distance-from-putative-site-of-origin cline, a similar pattern is observed in Spain and in the rest of Europe: the frequency of the derived allele decreases moving away from the South of Italy (R^2=0.68; p<0.0001). Interestingly the Spanish data set trend line falls beneath the European one (R^2=0.40; p=0.01), highlighting a ~10% decrease of *ls-tim* frequency in the Iberian peninsula for the same distance from Novoli.

LINES DETAILS							FREQUENCIES					OBSERVED			EXPECTED			
LINE	LOCATION	N	LAT	LONG	ALT	DISTANCE	f(ls)	f(s)	f(ls/ls)	f(ls/s)	f(s/s)	# ls/ls	#ls/s	#s/s	# ls/ls	#ls/s	#s/s	χ^2
TRV	Treviso (IT)	48	45.71	12.26	7	834	0.50	0.50	0.25	0.50	0.25	6	12	6	6	12	6	0.00
MAR	Market Harborough (UK)	248	52.48	-0.92	82	1990	0.41	0.59	0.19	0.44	0.37	24	54	46	20.98	60.05	42.98	1.26
KIL	Kilworth (UK)	312	52.53	0.98	94	1995	0.38	0.62	0.14	0.48	0.38	21	75	60	21.94	73.13	60.94	0.10
HФJ	Hфjbjerg (DEN)	100	56.11	10.21	n.a.	2027	0.41	0.59	0.08	0.66	0.26	4	33	13	8.41	24.19	17.41	6.63
GOT	Goteborg (SWE)	56	57.70	11.97	10	2284	0.36	0.64	0.11	0.50	0.39	3	14	11	3.75	12.86	11.57	0.22
KOR	Korpilahti (FIN)	134	62.02	25.55	104	3193	0.40	0.60	0.18	0.43	0.39	12	29	26	10.48	32.04	24.48	0.60

Table 3.4

Details of the new, non-Iberian European flies used in the *ls-tim* study. N: number of alleles analysed. Lat: Latitude in degrees North. Long: Longitude in degrees (Negative values: West; Positive values: East). Alt: altitude in meters above sea level (n.a. not available). Distance: overland distance from Novoli (Italy), calculated "as the crow flies" (km). The last column shows the results of the Hardy Weinberg test.
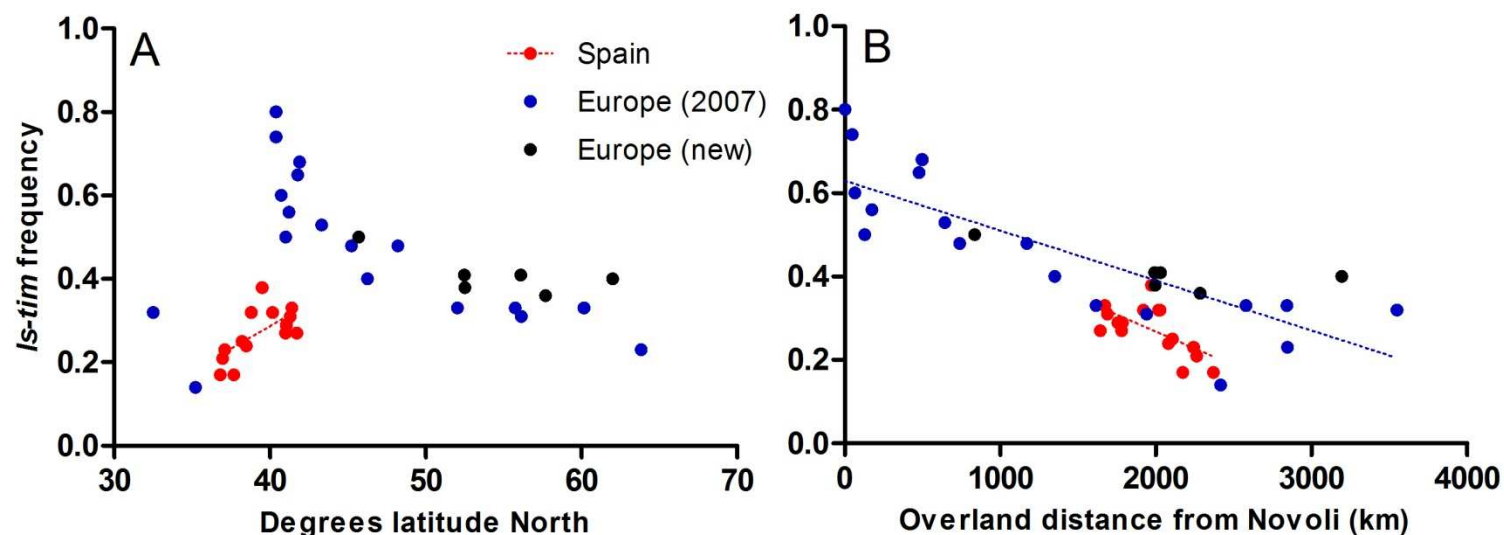
Figure 3.4

Frequency of *ls-tim* in Europe over latitude (A) and distance from Novoli (B). Red dots: Spanish data. Blue dots: European data from Tauber and colleagues (2007). Black dots: newly genotyped European populations.

3.3.3 EUROPE VS USA

A number of isofemale lines, collected in five different locations along the eastern coast of the USA, were available in the laboratory. Recently captured frozen males were not available for these lines, so in this case only one fly per isofemale line has been genotyped for *ls/s-tim*. Two of the five populations shown here have been genotyped by N. Patel, an undergraduate project student in our laboratory. Details regarding the site of collection, allele and genotype frequency and number of flies analysed are shown in Table 3.5. Three populations out of the five analysed are not in Hardy Weinberg equilibrium and are characterized by an excess of heterozygotes. In Figure 3.5 the frequencies have been plotted over latitude of collection (A and C). In Figure 3.6 B the frequency of the allele in the USA has been compared to its frequency in Europe.

The way *ls-tim* changes with latitude in the USA shows some analogies with the cline found in Spain (see section 3.1). The *ls-tim* frequency is relatively low, but significantly increasing with latitude (from 15% to 30%; R^2=0.96 p=0.01). Similarly, in the USA the frequency of homozygotes for *s-tim* significantly decreases with latitude (R^2=0.85 p=0.02).

	LINES DETAILS					FREQUENCIES					OBSERVED			EXPECTED			
LINE	LOCATION	N	LAT	LONG	ALT	f(ls)	f(s)	f(ls/ls)	f(ls/s)	f(s/s)	# ls/ls	#ls/s	#s/s	# ls/ls	#ls/s	#s/s	χ^2
[++]KIS	Kissimmee (FL)	164	**28.30**	-81.41	17	0.14	0.86	0.10	0.09	0.82	8	7	67	1.61	19.77	60.61	34.22***
[++]MOR	Morven (GA)	50	**32.72**	-83.70	73	0.18	0.82	0.08	0.20	0.72	2	5	18	0.81	7.38	16.81	2.60
RAL	Raleigh (NC)	100	**35.80**	-78.74	91	0.17	0.83	0.12	0.10	0.78	6	5	39	1.45	14.11	34.45	20.84***
PRI	Princeton (NJ)	132	**40.35**	-74.76	55	0.25	0.75	0.14	0.23	0.64	9	15	42	4.13	24.75	37.13	10.24**
[++]BOW(1)	Bowdoinham (ME)	100	**45.17**	-69.90	15	0.31	0.69	0.10	0.42	0.48	5	21	24	4.81	21.39	23.81	0.02

Table 3.5

Details of the American flies used in the *ls/s-tim* study. N: number of alleles analysed. Lat: Latitude in degrees North. Long: Longitude in degrees (Negative values: West; Positive values: East). Alt: altitude in meters above sea level. Populations marked with ([++]) were genotyped by V. Zonato, the others were genotyped by N. Patel (unpublished data). The last column shows the results of the Hardy Weinberg test. **: $p<0.01$; ***: $p<0.001$.
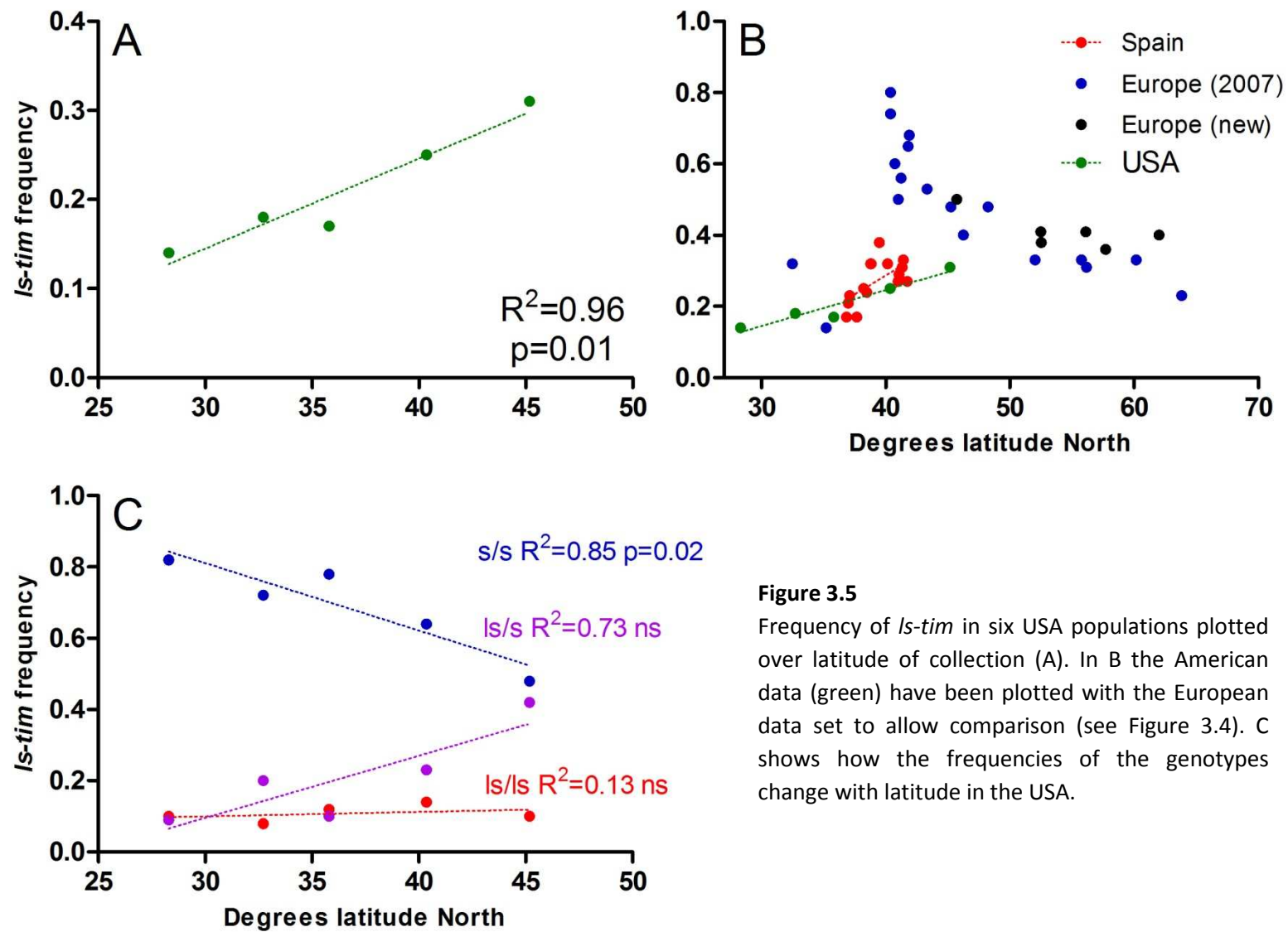
Figure 3.5
Frequency of *ls-tim* in six USA populations plotted over latitude of collection (A). In B the American data (green) have been plotted with the European data set to allow comparison (see Figure 3.4). C shows how the frequencies of the genotypes change with latitude in the USA.

3.3.4 NEUTRALITY TESTS

In order to study further the selection hypothesis, neutrality tests were performed on the region under study and on two intergenic regions: 5', 19.7kb upstream and 3', 25.9kb downstream of the *ls/s-tim* polymorphic site. Around 70 individuals collected in three locations in Spain were made hemizygous for the *timeless* locus and the three regions of interest were amplified and sequenced (see Table 3.6).

DNA REGIONS			NUMBER OF INDIVIDUAL ANALYSED			
Name	Location relative to *ls/s*	Length	SP22	SP35	SP53	TOTAL
5'	19.7 kb UPSTREAM	828	18	30	19	67
LS		1119	22	27	20	69
3'	25.9 kb DOWNSTREAM	1315	23	31	16	70
LOCATIONS OF COLLECTION						
Code	Location	Latitude	Frequency *ls-tim*			
SP22	Nijar	36.97	0.21			
SP35	Fontanares	38.78	0.32			
SP53	Manresa	41.73	0.37			

Table 3.6
Upper panel: details regarding the three genomic regions amplified, sequenced and analysed. Their length is reported in base pairs. The last four columns report the number of individuals sequenced, from each of the three geographical locations under study. Bottom panel: details of the three geographical regions analysed. Latitude is reported in degrees North.

The null hypothesis of neutrality assumes the absence of population structure, constant population size, no recombination, and that all the alleles present in the data set are selectively neutral. When the null hypothesis is rejected, one of these assumptions is violated.

The tests were carried out with the Software DNAsp 5.10.01 (Librado & Rozas, 2009). Gaps were not considered in the analysis and all the statistics were computed using the total number of mutations (η). However, using the total number of segregating sites (S) would not

have changed the outcome of the analysis since in the data set η and S have the same value (only two variants are present per variable site).

3.3.4.1 GENETIC DIFFERENTIATION

Firstly, the genetic differentiation among the three populations was analysed, to decide whether all the sequences could be treated as one unique population or if it was more appropriate to analyse them separately. In particular, using the method described by Hudson *et al.* (1992), simulations were carried out to test the null hypothesis of no genetic differentiation between populations at different localities. The results are shown in Table 3.7.

	χ^2 (Df)	χ^2 P-value	H_S	H_{ST} (PM test P-value)	K_S	K_{ST} (PM test P-value)	K_S^*	K_{ST}^* (PM test P-value)	Z (PM test P-value)	Z* (PM test P-value)
5'	57.41 (54)	0.35 (ns)	0.90	0.002 (0.35 ns)	25.11	0.005 (0.27 ns)	2.80	-0.001 (0.45 ns)	1137.47 (0.34 ns)	6.76 (0.44 ns)
LS	21.27 (26)	0.73 (ns)	0.84	-0.02 (0.97 ns)	3.22	-0.003 (0.47 ns)	1.21	-0.01 (0.74 ns)	1177.56 (0.54 ns)	6.84 (0.78 ns)
3'	68.87 (66)	0.38 (ns)	0.89	-0.0002 (0.40 ns)	35.59	-0.0006 (0.42 ns)	2.83	0.004 (0.26 ns)	1205.28 (0.37 ns)	6.82 (0.28 ns)

Table 3.7
The table shows the results of the genetic differentiation statistics, for each of the three genomic regions considered. For an explanation of the different values, see Appendix 8.1.

As shown in Table 3.7 all the values calculated by DNAsp turned out to be non-statistically significant, for all the DNA regions analysed. This result suggests that the three geographical locations can be pooled together and considered as one metapopulation. Nevertheless, further analyses performed on these regions were carried out both, on local populations and at the metapopulation-level.

3.3.4.2 TAJIMA'S TEST

The purpose of this test, developed by Tajima (Tajima, 1989), is to verify whether a set of sequences is evolving neutrally or if there is evidence pointing towards external forces influencing the evolution of the sequences. Such external forces could be of various natures, from natural selection to population dynamics, such as migrations or population size changes.

When performing the Tajima's test, the null hypothesis is neutral evolution. In such a *scenario* the molecular differences observed in the dataset are to be considered selectively neutral, and present in the population only because of a combination of chance and genetic drift. The test is performed by measuring the DNA variability in the dataset with different approaches:

Nucleotide diversity (π): measures the frequency with which two randomly chosen alleles in the dataset differ at a specific site. It is equal to the sum of the nucleotide differences between all possible pairs of sequences, normalised by the number of sequences considered and their length.

Watterson's Θ (Θ_w): measures how many sites present variability. It is calculated by counting the number of polymorphic sites. Such a value is divided by the number of sites per sequence, and by a correction factor that takes into account the number of sequences analysed.

Under neutrality and assuming no recombination, these two values are expected to be equal. If they differ significantly then the null hypothesis of neutrality is rejected, and further investigation is required in order to understand which process is perturbing neutrality in the data set.

Tajima's D is defined by:

$$D = \left.(\pi - \Theta_w)\middle/ \sqrt{\hat{V}(\pi - \Theta_w)}\right.$$

A value significantly larger than zero is an indication of polymorphisms present in the dataset at a relatively high frequency (high π compared to Θ_w). That can be explained by either balancing selection acting on the sequences, or a recent decrease in population size. The opposite *scenario* suggests an excess of rare variants (singletons) and could be an indication of

a recent population expansion or positive selection. The analysis was carried out using the software DNAsp 5.10, and the results are reported in Table 3.8.

REGION	POPULATION	TOT SITES	SEGREGATING SITES	TAJIMA' S D	D_1	D_2	D_3	D_4
5'	META	680	38	0.75				
	SP22	681	33	0.25				
	SP35	703	31	1.02				
	SP53	705	28	0.88				
LS	META	977	27	-1.62	-1.43	-1.85*	1.25	-1.79*
	SP22	977	19	-1.55	-1.79	-2.14*	0.59	-1.70
	SP35	977	17	-1.31	-0.55	-1.01	1.08	-1.52
	SP53	977	12	-1.06	0.06	-1.16	1.26	-1.36
3'	META	1100	61	0.12				
	SP22	1100	57	0.20				
	SP35	1132	53	-0.69				
	SP53	1141	49	-0.06				

Table 3.8
The table shows the results of the Tajima's test, performed on the three genomic regions (5', LS and 3'), on the local populations (SP22, SP35, SP53) and metapopulation-level (META). For details about the locations of collection see Table 3.6. Total sites is given in base pairs and sites with gaps are excluded from the analysis. 'Segregating sites' refers to the number of positions which present variability in the dataset. D_1, D_2, D_3 and D_4 indicate the Tajima's D calculated only on a subset of sites. D_1: coding region; D_2: synonymous sites; D_3: non synonymous sites; D_4: silent sites (include synonymous sites and other sites which do not lead to an amino acid substitution, introns and non coding regions). * indicates significant values ($p<0.05$)

As reported in the table above, Tajima's D tends to be negative in the data set "LS", although it does not reach a significantly negative value. However, significantly negative values are reached when only synonymous or silent sites are considered in the analyses. These sites are less constrained therefore more likely to be characterised by singletons.

3.3.4.3 FU'S AND LI'S TEST STATISTICS

This test, developed by Fu and Li (Fu & Li, 1993), is a modification of the Tajima's test, based on the idea that the external branches of a tree will contain those mutations which have a deleterious effect, and are therefore present at low frequencies. On the other hand, mutations kept in the population by balancing selection for many generations will appear in

the internal branches of the genealogy. The test can be computed with or without the use of an outgroup. When it is calculated without, the statistics are:

D*: based on the difference between η_s (number of singletons, mutations that appear only in one sequence) and η (total number of mutations)

F*: based on the difference between η_s and k (average number of nucleotide differences between pairs of sequences)

An outgroup can be used to determine which mutations are in the external branches. In other words, by comparing the data set to the ancestral form of the sequences, the new variant of the polymorphism can be distinguished from the original allele. In this case the two statistics calculated are D and F, which correspond to D* and F* except that η_s is replaced by η_e (number of mutations in the external branches, which corresponds to the number of singleton variants which are not shared with the outgroup). As above, the statistical significance is determined by comparison to data obtained by computer simulations. The results of the test are shown in Table 3.9. Sequences of *D. simulans* were used as outgroup.

REGION	POPULATION	TOT SITES	SEGREGATING SITES	D*	F*	D	F
5'	TOTAL	680	38	0.51	0.72	0.09	0.34
	SP22	681	33	0.03	0.11	0.19	-0.06
	SP35	703	31	1.31	1.43	1.26	0.38
	SP53	705	28	1.06	1.17	1.01	1.05
LS	TOTAL	977	27	-1.71	-2.00	-0.77	-1.26
	SP22	977	19	-2.76*	-2.80*	-1.02	-1.36
	SP35	977	17	-1.86	-1.98	-0.76	-1.10
	SP53	977	12	-1.82	-1.86	-0.96	-1.14
3'	TOTAL	1100	61	1.14	0.89	1.16	0.90
	SP22	1100	57	0.02	0.08	0.03	0.12
	SP35	1132	53	-0.58	-0.72	-0.95	-1.05
	SP53	1141	49	0.51	0.40	0.94	0.82

Table 3.9
Results of the Fu and Li test, performed with and without an outgroup. For a description of the statistics D*, F*, D and F see text. * indicates significant values (p<0.05)

Consistent with the Tajima's test results, we observed a tendency for negative values in the "LS" region, although they do not always reach significance.

3.3.4.4 HKA TEST

The Hudson-Kreitman-Aguadè, HKA (Hudson *et al.*, 1987) test compares the observed number of variable sites between two unlinked loci, and asks whether differences can be explained by different mutation rates affecting the two regions (null hypothesis). If this is not fulfilled, then the null hypothesis is rejected because the sequences contain more (or less) mutations than explainable by variations in the mutation rates. These are taken into account by comparing the number of differences between the two loci analysed (intraspecies comparison) and the same loci in a sibling species (interspecies comparison). Expected number of segregating sites between the two loci and number of differences at the same locus in the two species are calculated and compared to the observed numbers. The significance of the deviation is then calculated with a contingency test. In our analysis, the region containing the polymorphic site was compared to the two flanking (not completely unlinked) regions, using the same approach used by Kreitman and Hudson to infer the evolutionary history of the *Adh* locus (Kreitman & Hudson, 1991). Sequences from *D. simulans* were used for the interspecies comparison and the results of the test are shown in Table 3.10. As shown in the table, when the two intergenic regions are compared, no significant difference is observed, regardless of the dataset used. On the other hand, when the "LS" region is compared to either of the two intergenic regions, significant (or almost significant) deviations from expected values emerge, with the "LS" region showing less intraspecies variability than expected, consistent with a signature of directional selection.

	DNA REGION	TOT SITES	S OBS	S EXP	DIFF OBS	DIFF EXP	χ² (P-value)
META	5'	598	23	15.74	26.56	33.82	3.674
	LS	977	27	34.26	83.68	76.42	(0.054)
	5'	598	23	25.84	26.55	23.71	0.34
	3'	1100	61	58.16	50.93	53.79	(0.56)
	LS	977	27	42.28	83.68	65.40	5.84
	3'	1100	61	45.72	50.94	66.22	(0.016*)
SP22	5'	599	16	10.45	26.48	32.03	2.66
	LS	977	19	24.55	83.18	77.63	(0.10)
	5'	599	16	20.47	26.48	22.01	0.80
	3'	1100	57	52.53	51.56	56.04	(0.37)
	LS	977	19	36.38	83.18	65.8	6.34
	3'	1100	57	39.62	51.56	68.94	(0.01*)
SP35	5'	598	18	11.06	26.56	33.5	4.22
	LS	977	17	23.94	84.04	77.1	(0.04*)
	5'	598	18	20.41	26.56	24.15	0.26
	3'	1132	53	50.59	55.71	58.12	(0.61)
	LS	977	17	32.66	84.04	68.37	6.36
	3'	1132	53	37.34	55.71	71.37	(0.01*)
SP53	5'	598	12	7.04	26.62	31.58	3.27
	LS	977	12	16.96	85.75	78.79	(0.07)
	5'	598	12	16.60	26.62	22.01	1.07
	3'	1141	49	44.40	56.56	61.17	(0.30)
	LS	977	12	29.05	83.75	66.7	7.74
	3'	1141	49	31.45	56.56	73.62	(0.005**)

Table 3.10

Results of the HKA test performed on the Spanish data set (META) and on the 3 sub populations (SP22, SP35, SP53). The 5' and 3' regions were both compared to the LS region, and to each others. Total number of sites is given in base pairs and sites with gaps are excluded from the analysis. S OBS: observed number of segregating sites (intraspecies comparison). S EXP: expected number of segregating sites (intraspecies comparison). DIFF OBS: observed number of differences (interspecies comparison). DIFF EXP: expected number of differences (interspecies comparison). The analysis was carried out with the software DNAsp 5.10.01.

3.3.5 SELECTION VS DEMOGRAPHY

As mentioned in paragraph 3.3.4, a significant deviation from neutrality does not necessarily point to the action of natural selection. Several processes are able to perturb neutrality, such as population structure, reductions or increases in population size and migration. Given the geographical distribution of *ls-tim* in Europe and the results of the

neutrality tests, we can consider two other possible situations which might have led *tim* to show the present allelic distribution:

> 1) *ls-tim* appeared in the South of Italy and it increased in frequency because of genetic drift. Human overland migration from the South of Italy enhanced the genetic flow to the North and contributed to shape the distance-from-site-of-origin cline that we observe today.
>
> 2) *ls-tim* appeared in the South of Italy and it increased in frequency because of the selective advantage it confers over *s-tim*. Human migration helped the spread of the new allele in Europe where it will probably slowly increase in frequency due to its evolutionary advantage.

One way to discriminate between these two hypothesis is to focus attention on neutral loci spread throughout the *D. melanogaster* genome. If the present *ls-tim* distribution is merely due to a demographic effect and *ls-tim* is selectively neutral (option 1), then many other neutral loci will show the same geographical allelic distribution. Nevertheless if the current cline is mainly due to the selection affecting *ls-tim* (option 2), then neutral loci will not show similar geographical pattern.

The frequency of alleles of a set of neutral microsatellite located on the X and 2nd chromosomes of *D. melanogaster* was analysed in relation to the overland distance between Novoli (putative site of origin of the *ls-tim* allele) and site of collection of the flies analysed. The data set was kindly provided by Prof. C. Schlötterer and the details about the lines and genotyping technique can be found in Caracristi and Schlötterer (Caracristi & Schlötterer, 2003) and in Nunes *et al.* (Nunes *et al.*, 2008). Table 3.11 reports the details regarding the eight locations which were actually considered in our study. These were chosen because of their immediate physical proximity to some of the locations used in the *ls/s-tim* frequency study (Paragraph 3.3.2).

LINE CODE	LOCATION	COUNTRY	N	LATITUDE	DISTANCE
MAR	Marrakech*	Morocco	28	31.63	1689
KR	Crete	Greece	18	35.25	2583
NA	Naples	Italy	64	40.83	320
WR	Weil am Rhein	Germany	60	47.58	1180
KAT	Katowice	Poland	60	50.27	1640
TEX	Texel	Holland	60	53.08	1801
KBH	Copenhagen	Denmark	60	56.16	1933
FI	Harjavalta	Finalnd	38	64.85	2885
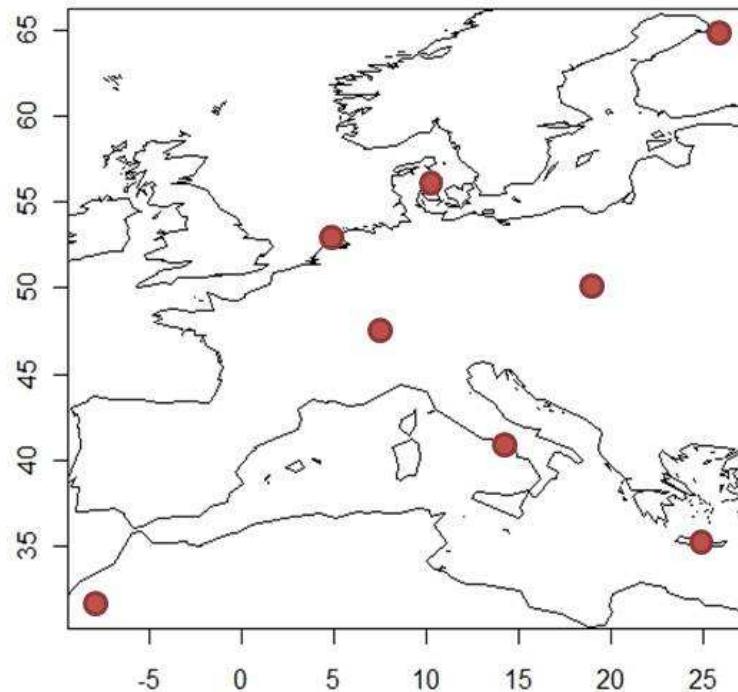
Table 3.11

Top: details of the populations used in the microsatellite study. N: number of alleles genotyped. Lat: Latitude in degrees North. Distance: overland distance from Novoli (Italy), calculated "as the crow flies" (km). (*) Distance of Marrakech from Novoli was calculated through the Iberian Peninsula to minimise overseas migration. Bottom: map showing the locations of collection of the flies used in the microsatellite study. The X and Y axis indicate degrees longitude and latitude respectively.

A total of 48 loci were analysed, 23 located on the X and 25 located on the 2nd chromosome. The frequency of each of the alleles present at each locus was correlated to both latitude of site of collection and distance from Novoli. A significant correlation was considered only if the allele frequency was larger than zero in at least four locations out of eight.

As shown in Table 3.12, the vast majority of the loci analysed do not show any significant correlation with latitude nor distance. Only nine of them are characterised by at least one allele whose frequency varies significantly with latitude. This might be due to their physical proximity to some loci involved in temperature/photoperiod adaptation, or it may simply be due to chance. Three loci have one allele which varies in frequency with distance from Novoli. One of them, ft-CA, is located on the second chromosome relatively close to *tim* (whose cytological position is 23F6). At first sight it seems reasonable to assume that the two loci are in linkage disequilibrium (LD), therefore share the same allelic geographical distribution. Nevertheless ft-CA is located around 900 kb away from the *s/ls-tim* polymorphic site, abundantly out of the range of linkage disequilibrium in *D. melanogaster*, estimated to be around 500 bp (Macdonald *et al.*, 2005; Miyashita *et al.*, 1993).

Reference	Locus	Chr	Cyt Position	Number of Alleles	Correlation with Latitude	Correlation with Distance
147	DS09020	X	15A	11 (2)	ns	ns
1310	AE002566-gt	X	3A	5 (1)	ns	ns
1385	X17869774gt	X	17A	16 (2)	ns	p<0.05 (1)
1395	X19942721gt	X	19C	12 (3)	ns	ns
1486	X13039889ca	X	11E	15 (3)	ns	ns
1488	X13203739gt	X	11F	18 (2)	p<0.05 (1)	ns
1496	X14425888gt	X	12F	7 (3)	ns	ns
1500	X15146508gt	X	13C	4 (1)	ns	ns
1502	X15279912atc	X	13E	9 (3)	ns	ns
1780	X3439769ca	X	3E	9 (2)	ns	ns
1784	X3829513gt	X	4A	12 (2)	ns	ns
1794	X3306698ca	X	3E	12 (1)	p<0.05 (1)	ns
1796	X3343263ca	X	3E	8 (1)	ns	ns
1800	X3516772ga	X	3F	6 (3)	ns	ns
1804	X3655941ga	X	3F	10 (3)	ns	p<0.05 (1)
1828	X4944599ca	X	4E	9 (2)	p<0.05 (2)	ns
1834	X5326452ct	X	5A	14 (4)	ns	ns
1836	X5179712gt	X	4F	13 (1)	ns	ns
1842	X5973753gt	X	5D	15 (2)	p<0.01 (1)	ns
1852	X7028104ga	X	7A	8 (3)	p<0.05 (1)	ns
1864	X8022709ca	X	7D	9 (2)	ns	ns
2061	X15854539ta	X	14A	3 (2)	ns	ns
2257	X15149564gt	X	13C	3 (2)	ns	ns
157	Drogpad	2nd	47A	10 (1)	ns	ns
406	Dm0620	2nd	51E	5 (2)	ns	ns
426	cact-TC	2nd	35F	5 (1)	ns	ns
428	cact-TG	2nd	35F	6 (2)	ns	ns
430	Adh-TC	2nd	35B	11 (3)	ns	ns
436	ft-CA	2nd	24E	6 (2)	ns	p<0.05 (1)
440	Pkc53E-GA	2nd	53D	9 (2)	ns	ns
444	Dm0600-TC	2nd	24C3-D1	7 (2)	ns	ns
446	Pkg-TC	2nd	23A	6 (3)	p<0.05 (1)	ns
448	Pkg-GT	2nd	23A	12 (2)	p<0.05 (1)	ns
450	Cad-GA	2nd	38D4-E1	6 (2)	ns	ns
620	Ote-GA	2nd	55A2-B1	7 (2)	ns	ns
656	Dm0332-TC	2nd	29F	3 (2)	ns	ns
681	Dm0600-TA	2nd	55F	6 (2)	ns	ns
699	tor-TA	2nd	43B3-C5	3 (2)	ns	ns
858	AC005270	2nd	24E1-F1	17 (3)	ns	ns
1972	2L20794024	2nd	38F6-39A1	20 (4)	ns	ns
2008	2R5196790	2nd	46F	7 (1)	ns	ns
2020	2L10056972	2nd	31A	11 (2)	ns	ns
2085	2R5377736ca	2nd	47A	9 (4)	ns	ns
2087	2R5394848ca	2nd	47A	6 (3)	ns	ns
2091	2R5442209ta	2nd	47A	8 (2)	p<0.05 (1)	ns
2093	2R548177ta	2nd	47A	9 (2)	ns	ns
2095	2R5491247ta	2nd	47A	4 (2)	ns	ns
2097	2R5510443ca	2nd	47A	5 (1)	ns	ns
2099	2R5514150ca	2nd	47A	3 (2)	p<0.05 (1)	ns

Table 3.12

Details regarding the 48 loci analysed in the microsatellite study. The table reports locus name, chromosome and cytological position, and number of alleles present in the dataset. In brackets are the number of alleles whose frequencies were included in the analyses (correlation performed only when the allele frequency was larger than zero in at least 4 locations out of 8). The last two columns report the results of the correlation analysis between allele frequency and latitude or distance. ns: correlation not significant. If Bonferroni correction for multiple testing is applied (106 correlations were performed), none of these results are significant.

3.3.6 AGE OF THE NEW ALLELE

Three different dating methods have been used to estimate the age of the new allele.
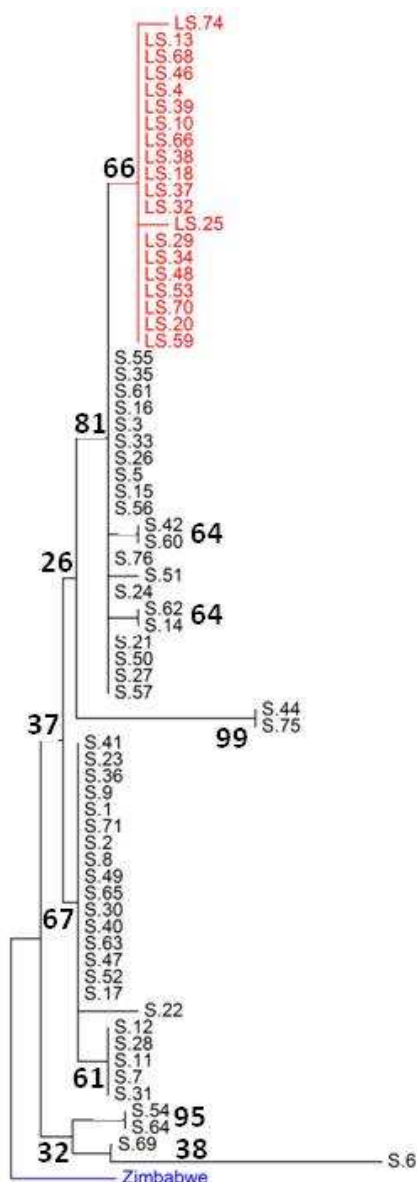
Previous phylogenetic estimations used the sibling species *D. simulans* (diverged from *D. melanogaster* 2-2.5 million years ago (Russo *et al.*, 1995)), to calibrate the age of *ls-tim* as ~8000-10,000 years (Tauber *et al.*, 2007). To achieve a higher resolution, we used a sub-Saharan African *D. melanogaster tim* sequence as an ancestral outgroup (*D. melanogaster* colonised Europe from sub-Saharan Africa about 15Kya (Li & Stephan, 2006)) and compared it with the Spanish sequences. As shown in Figure 3.6 all the *ls-tim* haplotypes (in red) cluster at the top of the Neighbour Joining tree regardless of location of collection, suggesting that *ls-tim* is the more recent, derived allele. The linearised tree branch length to the ancestral African sequence (in blue) suggests that *ls-tim* is ~3000 years old.

Figure 3.6

Neighbour Joining tree with Kimura 2- parameter model, where gaps were treated as pair wise deletions. Red, *ls-tim* sequences, black *s-tim*, blue, African (Zimbabwe). Numbers represent bootstrap values. Sequences marked 1-23, 24-44 and 46-75 were collected from Spanish populations SP-22, SP-35 and SP-53 respectively (Table 3.6).

The software GENETREE (Griffiths & Tavaré, 1999) uses a coalescent-based method to generate a likelihood estimate of the temporal distribution of mutations. Using this software we obtained an estimated age (±SD) of 2,130 (899) years assuming population size of 10^5 and 20 generations per year. The tree obtained with this software is shown in Figure 3.7.
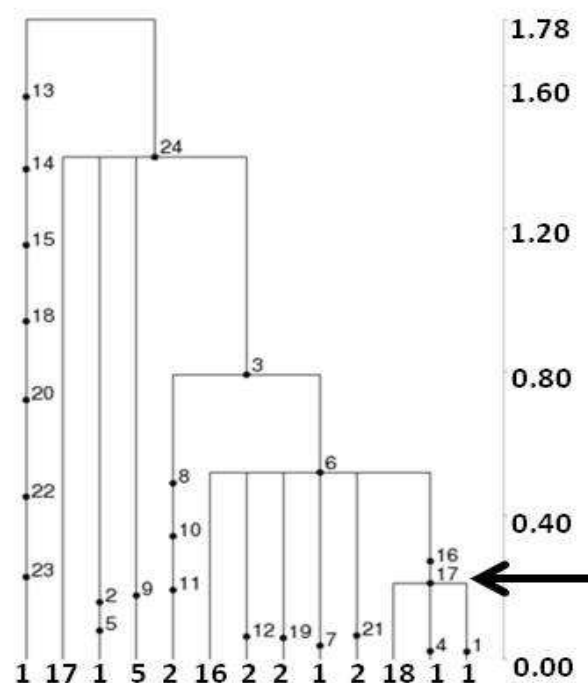
Figure 3.7

Tree generated with GENETREE. Numbers at the end of each branch indicate the frequency of each haplotype. Numbers on the branches indicate the segregating sites. The *ls/s-tim* mutation is represented by number 16 or 17 (black arrow, the two mutations are 3 bp apart therefore in strong linkage disequilibrium and impossible to distinguish) and includes three haplotypes, two of which are unique. The vertical timescale on the right indicates coalescent units.

The third method used is based on the amount of linkage disequilibrium between the mutation of interest and flanking markers. The reasoning behind this method is that the age of the allele is proportional to the amount of LD with flanking sites. If the mutation is very young, LD will decay more slowly than expected as we move from the mutated site, because not

enough recombination events have occurred to make the two loci (mutation and marker) independent. The method, reviewed by Slatkin and Rannala (Slatkin & Rannala, 2000), is based on the following formula:

$$t = \frac{1}{\ln{(1-c)}} \, ln\left(\frac{x-y}{1-y}\right)$$

where c is the number of crossover events per generation between the two loci, and x and y are the frequencies of one of the marker alleles on the new (*ls-tim*) and old (*s-tim*) allele background respectively. We calculated t using different markers at a range of distances from *ls/s-tim*. The results are shown in Table 3.13.

Overall, and despite substantial variation between the age estimates produced by the three different methods, they all point to a relatively young allele which was introduced well after the colonisation of *Drosophila* in Europe (10-15 thousand years ago).

SNP POSITION	DISTANCE FROM *ls/s-tim*	N	ALLELES	x	y	c	t	YEARS
3508427	4948 (5')	60	G/A	0.37	0.37	$8.6*10^{-5}$	64173.48	3208.67
3508188	3715 (5')	60	T/C	0.47	0.83	$6.4*10^{-5}$	6083.71	304.19
3508818	4345 (5')	60	G/C	0.05	0.27	$7.5*10^{-5}$	19631.35	981.57
3478495	25978 (3')	63	T/G	0.28	0.49	$4.5*10^{-4}$	2728.25	136.41
3478009	26469 (3')	63	A/G	0.33	0.44	$4.6*10^{-4}$	3901.44	195.07
3477430	27043 (3')	63	Indel	0.22	0.36	$4.7*10^{-4}$	3757.86	187.89

Table 3.13

Estimates of the age of *ls-tim* polymorphism obtained using 6 markers upstream and downstream the polymorphic site. The distance of the marker from the *ls/s-tim* site is reported in bp. In brackets is the direction relative to *ls/s-tim*. N refers to the number of haplotypes used for the calculation. The last marker is an indel polymorphism consisting of 7 bp. The number of cross-overs between the two loci (c) has been calculated by multiplying the recombination rate (r=3.47 cM per Mbp in this region, (Fiston-Lavier *et al.*, 2010)) by the distance between the two loci (the recombination rate was divided by two to take into account the lack of recombination in male flies). t is given in generations. The estimation in years was obtained by assuming 20 generations per year.

3.4 DISCUSSION

A latitudinal cline in phenotype or allele frequency is often used as an indication that the character/gene under study is under natural selection, and the selective pressure is to be sought for in those environmental conditions which vary with latitude, most likely photoperiod and temperature. Nevertheless a latitudinal cline alone should not be used as definitive proof for the action of natural selection. When the selective strength is mild, selection is a process that can go on for thousands of years before reaching equilibrium, therefore it cannot be monitored and studied while it is happening. But the process itself leaves traces behind, scattered in the genomes of the individuals subjected to selection. Other processes (such as population structures, migration, genetic drift, population size changes), also leave marks on the genome, and these signatures can be easily mistaken with the action of selection. In this chapter we show different approaches used in order to discriminate between selection and population dynamics affecting the *timeless* locus in natural populations of *D. melanogaster*.

The cline in *ls-tim* frequency in Europe reported by Tauber and colleagues (2007), already pointed at a more complex pattern than a simple cline in gene allele frequency. This was because the direction of the cline itself seemed to be the opposite of what was expected based on the phenotypes conferred by the allele under study (the new, high diapause inducing allele, decreasing in frequency with latitude). In order to illuminate the dynamics of the appearance and spread of *ls-tim*, I analysed the frequency of the new allele in the Iberian peninsula. Surprisingly, I found a reversed latitudinal cline in Spain, compared to the rest of Europe. This result supports the idea that the European cline did not arise due to different selective pressures in the North and South of Europe. More likely, the mutation first appeared in the South of Italy and spread through Europe as a result of diffusion (possibly helped by human migration), due to the selective advantage the new allele confers over *s-tim*, regardless

of latitude. This hypothesis was confirmed by the significant distance-from-putative-site-of-origin cline found in the European dataset (enriched with a few newly collected populations) and in the Spanish dataset. The frequency of *ls-tim* decreases as we move away from Novoli, the location assigned as site of origin since it is characterised by the highest frequency of the new allele.
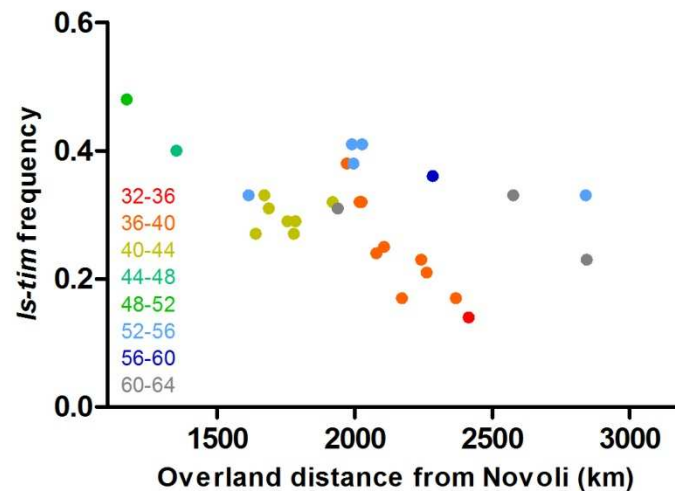
Figure 3.7
Subset of the data already shown in Figure 3.4 B. The colour code indicates the different latitudes (in degrees North) at which the flies were collected.

Remarkably, even if the distance cline shows the same trend in Spain as in the rest of Europe, there is a 10% frequency difference between the two trend lines, with Spain being characterised by a lower frequency of *ls-tim*. This might be explained by the fact that locations at approximately the same distance from Novoli are in fact situated at different latitudes. Figure 3.7 shows part of the data in Figure 3.4 B. In this case though, the colour code indicates the different latitude of collections. The plot suggests that if we keep one parameter constant (distance from Novoli), then latitude seems to be playing a role, favouring *ls-tim* in the North. This observation was confirmed by statistical analyses: the slopes of the two linear regression lines is identical ($F_{1,36}$=0.23; p=0.63) but their overall elevation is significantly different ($F_{1,37}$=21.54; p<0.0001). This suggests, for instance, that the new mutation might have reached Madrid and Copenhagen around the same time (same distance from Novoli), but in

Copenhagen it conferred an additional evolutionary advantage (due to enhanced seasonal selection) which accelerated its increase in frequency.
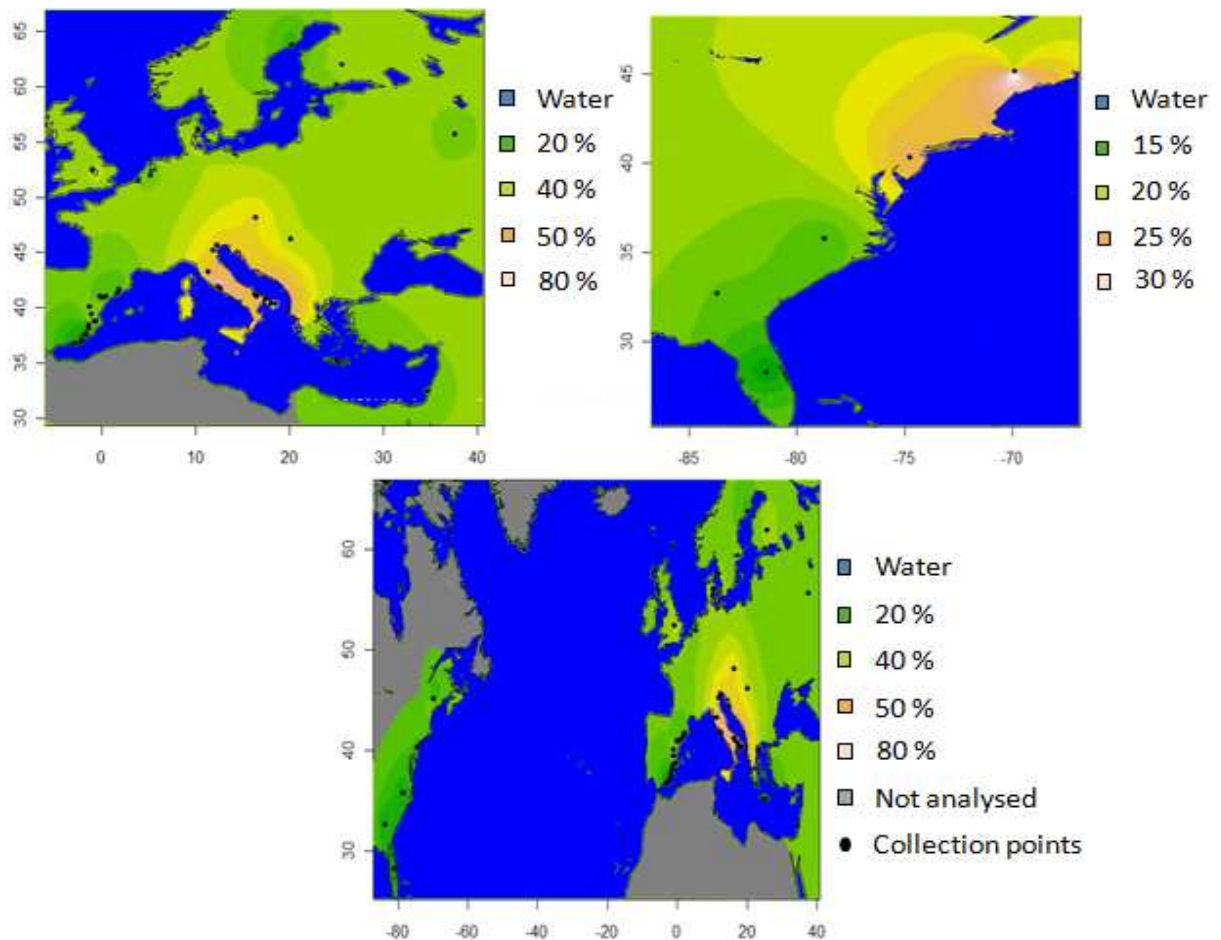
Figure 3.8
Images obtained with the R software (package "Maps") showing the distribution of *ls-tim* frequencies in Europe (top left), USA (top right; note that the scale differs in the two images) and a comparison between the two (bottom). Different colours represent different *ls-tim* frequencies. The black dots indicate the locations of collection and gray areas indicate regions where data is not available.

If this hypothesis is correct we can speculate that the present situation represents an evolutionary transient, and that this process is far from evolutionary equilibrium at the moment. One possibility is that slowly the cline in *ls-tim* frequency will disappear, the frequency of the allele reaching the same level (70-80%) throughout the whole of Europe or even going to fixation. Or perhaps the frequency might keep increasing in the extreme North,

until eventually a new very much flatter cline will emerge, opposite of the cline we observe now and more similar to that observed in the USA, but at much higher overall levels.

The frequency of *ls-tim* was also examined in North American populations. Not surprisingly, given *D. melanogaster* recent colonisation of the New World, the frequency of the new allele is much lower here than in Europe and yet it shows a significant latitudinal cline, with the new allele increasing in frequency with latitude. Figure 3.8 (realised with the Software R, package "Maps"), offers a visual representation of how the frequency of *ls-tim* varies in Europe and USA and a comparison between the two.

Flies were introduced into North America a few hundred years ago at most (David & Capy, 1988). When the colonisation process started, the frequency of *ls-tim* in Europe was presumably lower than it is now, reaching relatively high frequencies only in the South of the continent. For this reason, the lower frequency of the allele in the USA does not come as a surprise and it can be explained by the colonisation dynamics. David and Capy (1988) suggested that *D. melanogaster* flies were introduced into tropical America from central Africa through the slave trade, thus setting to zero the frequency of *ls-tim* in the South of the USA. Later on, probably around the 19th century, flies from Europe colonised North America, this time carrying with them a low frequency of the new allele (the original settlers came from northern Europe or Spain, where *ls-tim* frequencies are low). Thus the original low ~20% frequencies in New England may have risen in the past couple of hundred years to ~30%, whereas repeated introductions of *D. melanogaster* along the eastern coast have not generated high levels of *ls-tim*, presumably because in the South any seasonal selection will be relaxed. Clearly, the analysis of the frequency of the new allele in natural populations represents the first step towards the understanding of the forces driving the spread of *ls-tim* but further analysis is needed to discern the contribution of natural selection and demographic processes in North America.

The results of the Tajima's and Fu and Li's tests seem to point towards an excess of singletons, indicative of directional selection at the *timeless* locus. Such evidence is not present for the two flanking regions used as controls, suggesting that the phenomenon is peculiar to the *tim* locus, rather than involving the whole long arm of the 2nd chromosome. Negative Tajima's D values are obtained only when the LS region is analysed, but they reach statistical significance only when a subset of sites is considered: synonymous and silent sites. Singletons are more likely to be detected here because these sites are less constrained. As for the Fu and Li's test, again the values are substantially negative only when the LS region was considered, but they reach significance only when the most southern population (SP22) is included in the analysis. The HKA test highlighted a lower number of segregating sites than expected in the LS region, compared to both the two flanking regions. However, when the two intergenic regions are compared, the number of segregating sites does not deviate significantly from expectation. Again, this result is compatible with directional selection acting on the *timeless* locus.

Taken as a whole the above results are coherent with a *scenario* in which human migration is helping the spread of a new advantageous allele which is selected for in the population. This is not surprising given that *D. melanogaster* is a human commensal (Keller, 2007). As mentioned before, a negative Tajima's D indicates excess of low frequency polymorphisms, and could be indicative of recent population size expansion after a bottle neck or a selective sweep. Nevertheless, the fact that the vast majority of the microsatellites analysed do not show the same geographical distribution as *ls-tim*, further supports the view that selection is the main force in driving this process.

One could reasonably ask why the process is still ongoing and why it has not reached an equilibrium? The reason appears to lie in the relatively young age of *ls-tim*. My phylogenetic analyses all show that the new mutation is likely to have appeared around three

thousand year ago at most. Indeed depending on the markers, this could even be a few hundred years ago (Table 3.13). In any case, *ls-tim* is much more recent than originally suggested by Tauber and coworkers (Tauber *et al.*, 2007), who used *D. simulans* sequences as the outgroup. Using an African *D. melanogaster* sequence as the outgroup enhances resolution considerably. Given that *D. melanogaster* African lines are characterised by high genetic diversity, more African sequences should be included in the phylogenetic analysis as soon as they become available in the laboratory.

Our results regarding the spread of *ls-tim* were partly unexpected. Most clines are simply driven, shaped and maintained by latitude, or more precisely by factors associated with latitude, such as temperature and photoperiod. As mentioned before, latitudinal clines have been reported not only for many insects phenotypes (body size (Chown & Gaston, 2009), heat susceptibility and cold resistance (Hoffmann *et al.*, 2002), ovariole number (Azevedo *et al.*, 1996) just to name a few), but also for mouse pigmentation (Mullen & Hoekstra, 2008), rodents digestive flexibility (Naya *et al.*, 2008) and human skin pigmentation (Jablonski & Chaplin, 2010). The cline described in this chapter strongly resembles the one regarding human lactose persistence (LP). The allele for LP has been shown to have a recent origin and to have been positively selected (Ingram *et al.*, 2009). The persistence allele would only provide a selective advantage with a supply of fresh milk, therefore a gene-culture coevolutionary model was developed, where LP is only favoured in cultures practicing dairying, and dairying in more practiced in LP populations (Itan *et al.*, 2009). Homogeneous selection over space and time, in combination with human migration processes (the spread of farmers from the Balkans to the rest of Europe), are sufficient to explain the significant correlation between latitude and LP in Europe (Itan *et al.*, 2009).

Overall, I imagine that *ls-tim* established itself in southeastern Italy over a few centuries, and with human migration it was carried North and commenced its conquest of Europe.

Overall, the results presented in this chapter can be summarised as follows:

- *ls-tim* probably appeared in the South of Italy not more than three thousand years ago (and possibly considerably more recently), and spread from there thus generating a distance-from-site-of-origin cline

- the spread of the new allele was aided by human activity but could only happen because of the selective advantage conferred by *ls-tim* over *s-tim*, regardless of latitude. It is an ongoing process which has not yet reached equilibrium

- latitude does have an effect, further favouring *ls-tim* over *s-tim* when the distance from site of origin is the same

- the frequency of the new allele in the USA is much lower than in Europe and presents a latitudinal cline, with *ls-tim* increasing in frequency with latitude. The establishment of such a cline is likely to have been aided by the colonisation dynamics of North America.

4. COUCH POTATO

4.1 INTRODUCTION

couch potato (*cpo*) was first identified in a screen for genes involved in *D. melanogaster* central nervous system development (Bellen *et al.*, 1992b). The name was chosen because of the overall hypoactive behavior and sluggishness conferred by mutations in this gene. Viable mutant flies are characterised by a delayed development, compromised flight response, abnormal phototaxis, geotaxis and ether recovery, and increased seizure susceptibility (Bellen *et al.*, 1992b; Glasscock & Tanouye, 2005). *cpo* encodes a RNA recognition motif and a nuclear localization sequence (Bellen *et al.*, 1992a). CPO protein is expressed in the peripheral and central nervous system (PNS and CNS) of embryos, larvae and adults, as well as in other tissues such as the midgut, glia, salivary glands and ring gland, the major endocrine organ in *D. melanogaster* (Bellen *et al.*, 1992a; Harvie *et al.*, 1998).

Our attention to this gene was summoned by a paper from Schmidt and co-workers (Schmidt *et al.*, 2008) who performed a QTL screen to discover genes involved in controlling diapause in *D. melanogaster*, and identified *cpo* as a putative candidate gene. The level of the *cpo* expression was found to be linked to diapause (high diapause incidence seems to correlate with high levels of *cpo* expression). Furthermore, 35 third chromosomes from a natural population were put in a standardised genetic background; the diapause incidence of these

lines was then characterised, and a 3.5 kb *cpo* region was sequenced. By doing so, a region at the end of the *cpo* major exon was identified, which shows strong association with diapause. This region contains two non-synonymous SNPs at amino acid position 356 (Alanine to Valine) and 462 (Isoleucine to Lysine, Table 4.1). The latter was shown to display the strongest association with diapause and when it was taken in consideration no other SNP in this region accounted for any additional variance in the phenotype.

AMINO ACID POSITION	ALLELES		DIAPAUSE LEVEL	F (SOUTH)	F (NORTH)
	NUCLEOTIDE	AMINO ACID			
356	T	Valine	High	~0.20	~0.55
	C	Alanine	Low	~0.80	~0.45
462	A	Lysine	High	0.15	0.61
	T	Isoleucine	Low	0.85	0.39

Table 4.1
Non-synonymous SNPs present in the *cpo* region which showed a very strong correlation with the diapause phenotype. The last two columns show the frequency of the alleles in the southernmost and northernmost American population respectively (Schmidt *et al.*, 2008).

As mentioned in the Introduction, diapause levels have been shown to vary in flies collected along the East coast of the US (Figure 4.1), with the incidence of this trait predictably increasing with latitude (Schmidt *et al.*, 2005a). Interestingly, the frequency of SNP A356V also shows a cline (Figure 4.1), the *cpo*[356Val] allele increasing in frequency with latitude (Schmidt *et al.*, 2008).
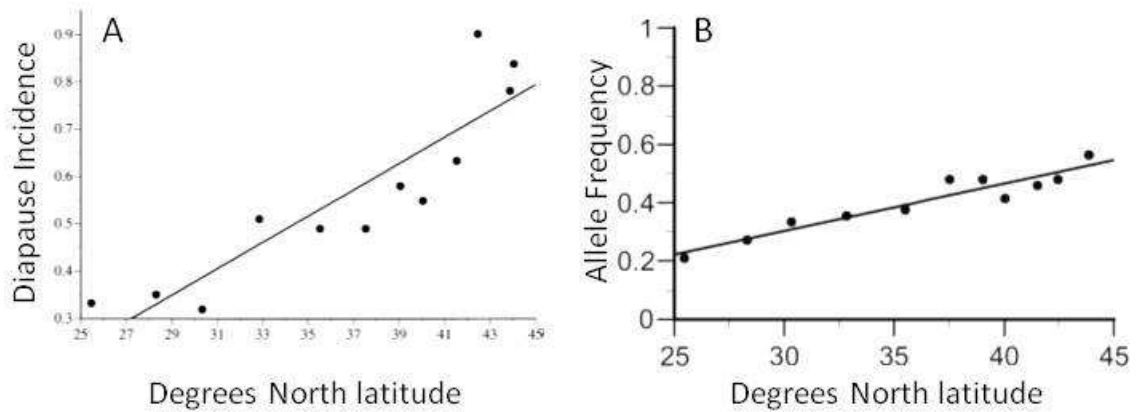
Figure 4.1
A: Incidence of diapause in *D. melanogaster* natural populations collected along the US East coast (Schmidt *et al.*, 2005a). B: Frequency of the allele *cpo*^{356Val} plotted against latitude (Schmidt *et al.*, 2008).

Given the physical proximity of the two SNPs and the strong linkage disequilibrium between them, it has been suggested that also SNP I462K is characterised by a latitudinal cline in frequency (with the *cpo*^{462Lys} allele increasing in frequency with latitude). The existence of SNP I462K cline was proven by sequencing the *cpo* region only in two populations at the extremes of the latitudinal range considered here: the frequency of *cpo*^{462Lys} was found to be 0.15 in the South and 0.61 in the North (23 and 24 haplotypes were sequenced respectively). This cline has been argued to be responsible for the diapause cline observed in North American flies, rather than the upstream A356V, given that alone it accounts for all the variance in diapause in the lines analysed.

cpo gene structure spans 84 kb and is characterised by 6 transcripts (Figure 4.2). The major coding region is exon 5 which, in most transcripts, encodes 449 amino acids. This exon is 37.5 kb from a 3' set of several small exons encoding around 286 residues, depending on transcript and annotation. These amino acids represent the RNA binding domain, which is absent in the short transcript *cpo*-RH. Interestingly, this transcript is the only one containing SNP I462K (blue arrow in Figure 4.2), that is located in the extra stretch of exon 5.
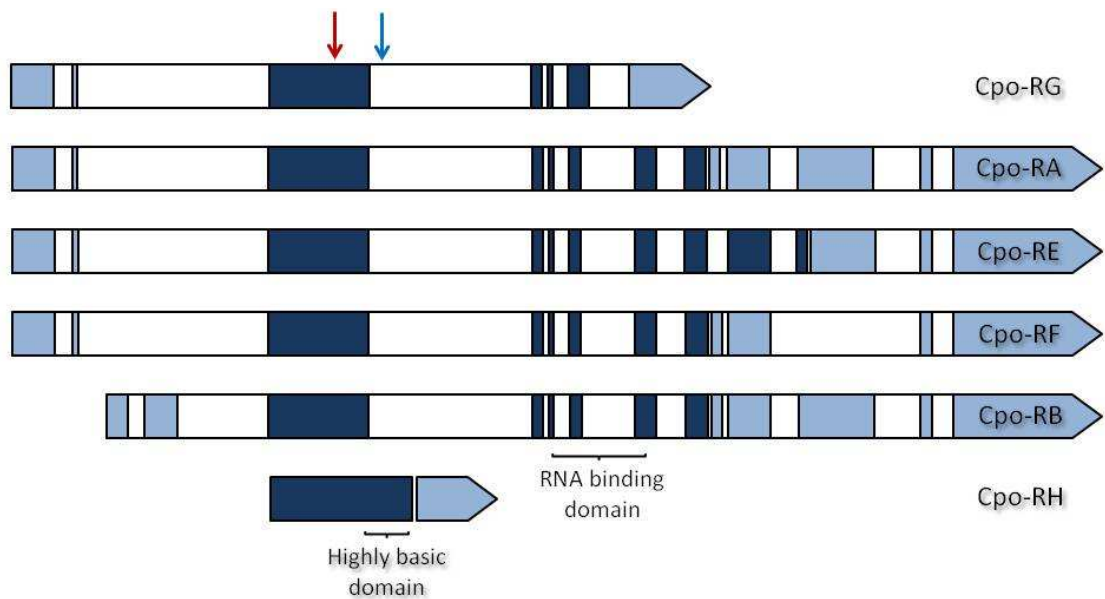
Figure 4.2
Schematic representation of the six *cpo* splicing variants. Dark blue boxes encode the protein, whereas light blue ones represent the 5' and 3' UTR regions. The red and blue arrows represent the position of SNP A356V and SNP I462K respectively. The figure is not to scale and was redrawn from the database Flybase as it appeared at the time this project was started (September 2008).

Lee and coworkers (2011) analysed the frequency of SNP I462K in natural flies collected along the eastern Australian coast, and found that, even if cpo^{462Lys} increases significantly with latitude (Figure 4.3), this is merely driven by the strong linkage disequilibrium between the SNP and the Payne inversion (Lee *et al.*, 2011). This is a cosmopolitan chromosomal inversion located in the right arm of chromosome 3 (*In(3R)P*) whose frequency has been shown to vary significantly with latitude in Australia (Knibb *et al.*, 1981). Furthermore, unlike American populations, SNP I462K does not seem to account for differences in diapause levels in different Australian populations.

In this chapter I characterise these two SNPs in European *D. melanogaster* populations and examine their effects on diapause incidence.
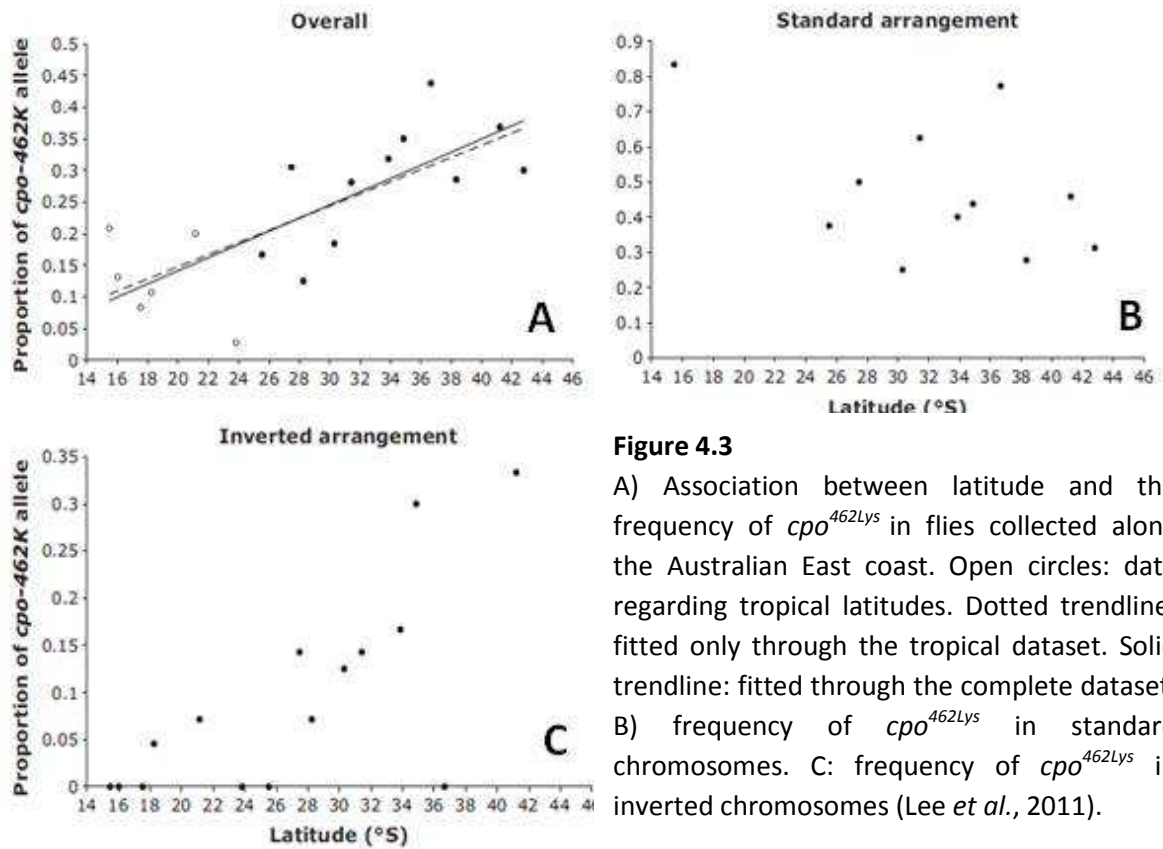
Figure 4.3
A) Association between latitude and the frequency of cpo^{462Lys} in flies collected along the Australian East coast. Open circles: data regarding tropical latitudes. Dotted trendline: fitted only through the tropical dataset. Solid trendline: fitted through the complete dataset. B) frequency of cpo^{462Lys} in standard chromosomes. C: frequency of cpo^{462Lys} in inverted chromosomes (Lee *et al.*, 2011).

4.2 MATERIALS AND METHODS

4.2.1 FLY LINES

As for the Spanish populations used in this chapter, methods of collections and husbandry are reported in Chapter 2. Where possible, males collected directly from the wild and kept in EtOH 100% at -20˚C were used. In some cases, in order to increase the sample size, one male per isofemale line was also included in the analysis. All the other flies used in this chapter were already available in the laboratory, details regarding location and date of collection are listed in Chapter 2.

4.2.2 SNP A356V GENOTYPING

SNP A356V was genotyped according to Schmidt and coworkers, 2008. The DNA region under study was amplified with the primers cpo-F 5'-*AACATCCGTTGCTGCTGTC*-3' and cpo-R 5'-*CCCCAAGCTGTCACTTTTGT*-3' (nucleotide coordinates 13793184..202 and 13793686..67 respectively). The following thermal profile was used to carry out the amplification: 40 cycles with 92°C for 35 sec, 55°C for 45 sec, 72°C for 30 sec. The PCR product was then subjected to treatment with the restriction enzyme *BsiEI*. The amplicons contain one *BsiEI* cutting site in the SNP 356T genotype (cpo^{356Val}), and two sites in the SNP 356C genotype (cpo^{356Ala}). The result of the digestion was then checked in a 1.5% agarose gel.

4.2.3 SNP I462K GENOTYPING

To genotype natural flies for this SNP, the DNA fragment under study was amplified, using the same primers and thermal profile as described in paragraph 4.2.2. The amplicons were then sequenced with the primer cpo-R and the sequences analysed with the software Geospiza FinchTV Version 1.4.0.

4.2.4 PAYNE INVERSION FREQUENCY

A SNP in complete linkage disequilibrium with the inversion was identified by Anderson and coworkers (Anderson *et al.*, 2005) at position 12253813 of the *D. melanogaster* genome sequence (Berkeley Drosophila Genome Project database). At this position standard and inverted chromosomes are characterised by an A or a C respectively. This SNP was used as a marker for the inverted chromosome, and was genotyped with a Bi-PASA approach (Bidirectional PCR Amplification of Specific Alleles (Liu *et al.*, 1997)).

Nevertheless, given that the frequency of inverted chromosomes falls to ~10% at latitudes > 35° North (Knibb, 1982), positive controls needed to optimise the technique were difficult to isolate, in our populations. The frequency of inverted chromosomes in a subset of European populations was checked by amplifying the region of interest with the primers Payne-F   5'-*TTTGCCGCAAATTATTGTGAG*-3'   and   Payne-R   5'-*ATCGCGTGCAGGTTGGC*-3' (nucleotide position 12253527..47 and 12254096..80 respectively) and the following thermal profile: 45 cycles with 92°C for 30 sec, 65°C for 30 sec, 72°C for 20 sec. The amplicons were then sequenced with the primer Payne-R.

4.2.5 NEUTRALITY TESTS

Around 35 individuals, collected in different locations in Europe, were used for this analysis. In order to make flies hemizygous for the locus under study, one female for each isofemale line was crossed to the deletion strain *Df(3R)Bsc566/Tmb* obtained from the Bloomington Stock Center. The *cpo* polymorphic region and one upstream intergenic region were amplified and sequenced with the primers listed in Table 4.2.

	PRIMER NAME	PCR PRIMERS (5'-3')	SEQUENCING PRIMER (NAME AND SEQUENCE 5'-3')	ANNEALING TEMPERATURE
5'	5'-F	*GGCCAGGGAAAATTGAGAAT* (13745290..309)	*5'-F GGCCAGGGAAAATTGAGAAT* (13745290..309)	57°C
	5'-R	*TGCGATGTTGTCTCTTGGAC* (13746299..80)	*5'-Seq-R TTCTCTTCCGACTCTGGCTCT* (13746278..58)	
cpo	cpo-NT-F	*AAAGCCATCACCATTTCGTC* (13792444..63)	*cpo-NT-F AAAGCCATCACCATTTCGTC* (13792444..63)	56°C
	cpo-NT-R	*CGCCCAACGCTATCTTTATC* (13793626..07)	*cpo-NT-Seq-R AGTCCATGCTCTGCGAAAGT* (13793521..02)	

Table 4.2
Primers used to amplify and sequence the polymorphic region in *cpo* and the upstream intergenic region (5'). In brackets the nucleotides coordinates are indicated.

The sequences were analysed with the software Geospiza FinchTV Version 1.4.0, and the Neutrality Tests were carried out with the program DNAsp version 5.10.01 (Librado & Rozas, 2009) and NeutralityTest version 1.1, kindly provided by Haipeng Li (Li & Fu, in preparation).

4.2.6 SNPs IN *D. simulans*

In order to amplify the region under study in *D. simulans* flies, the following primers were used: cpo.sim-F 5'-*AGCATCCGTTGCTGCTGT*-3' and cpo.sim-R 5'-*CCCAAGCTGTCAACTTTTGTC*-3' (nucleotide coordinates 7651979..97 and 7652419..398 respectively). The PCR reaction was performed using the following thermal profile: 45 cycles with 92°C for 30 sec, 55°C for 30 sec, 72°C for 20 sec.

The flies used for this analysis were kindly provided by Christian Schlötterer, and details about site and date of collection can be found in Chapter 2.

4.2.7 EXPRESSION STUDY

To detect and quantify the expression of *cpo* and in particular that of isoform RH, virgin flies from the Treviso population (see Chapter 2) were subjected to the desired temperature and kept in light boxes in a 12:12 LD regime. After 4 days they were collected, their RNA was extracted and cDNA was synthesised. Specific primers were designed in order to amplify either only isoform RH, or all the isoforms except RH. The following common forward primer was used: cpo-F 5'-*AACATCCGTTGCTGCTGTC*-3'. As a reverse primer cpo-R 5'-*CCCCAAGCTGTCACTTTTGT*-3' (nucleotide coordinates 13793184..202 and 13793686..67 respectively) was used to amplify cpo-RH, since it anneals to a region in Exon 5 not present in all the other transcripts (Figure 4.2). The primer cpo-R2 5'-*ACGAAAAGTGTGCGAACCTC*-3' (nucleotide coordinates 13832673..55) recognises a region in exon 6, thus allowing the

amplification of all the isoforms but *cpo*-RH (Figure 4.2). An extra couple of primers were added to the reaction, in order to control for PCR efficiency and allow quantification of the expression levels. These primers amplify a region of the *Gapdh-glyceraldehyde 3-phosphate dehydrogenase* gene and their sequences were obtained from Schmidt *et al.*, 2008: GapdhF 5'-*TGAAGAAGTCGGTGGAGACC*-3' and GapdhR 5'-*AGTGCTGGAGCCGAGTATGT*-3' (nucleotide coordinates 3679773..92 and 3680370..51 respectively). The thermal profile used for the amplification of all the isoforms but *cpo*-RH was the following: 36 cycles with 92°C for 30 sec, 56.3°C for 30 sec, 72°C for 30 sec. To amplify only isoform *cpo*-RH the following thermal profile was used instead: 31 cycles with 92°C for 30 sec, 59°C for 30 sec, 72°C for 30 sec. The result of the amplification reaction was then checked in a 1.5% Agarose gel.

4.2.8 FLY LINES FOR BEHAVIOURAL ANALYSIS

Four different lines with different combination of the *cpo* SNPs were created and characterised phenotypically. Males and virgin females were collected from the Treviso population (Chapter 2). Their wings were removed, the DNA extracted and characterised for the *tim* (*ls-tim* and *s-tim*) and *cpo* alleles. Flies with the required genotypes (the four SNPs combinations in *cpo*, in the same *tim* background) were then inter-crossed to obtain the desired populations. Several steps of genotyping-crossing were required to obtain the final, homogeneous populations which were then checked for their diapause phenotype.

4.3 RESULTS

4.3.1 SNP A356V IN EUROPEAN POPULATIONS

European *D. melanogaster* flies were genotyped for *cpo* SNP A356V. The region under interest was amplified and subjected to a restriction enzyme digest. The enzyme *BsiEI* recognises one cutting site in the genotype *cpo*^{356Val} and two cutting sites in the genotype *cpo*^{356Ala}. Figure 4.3 shows the genotyping strategy used, and an example of the outcome.
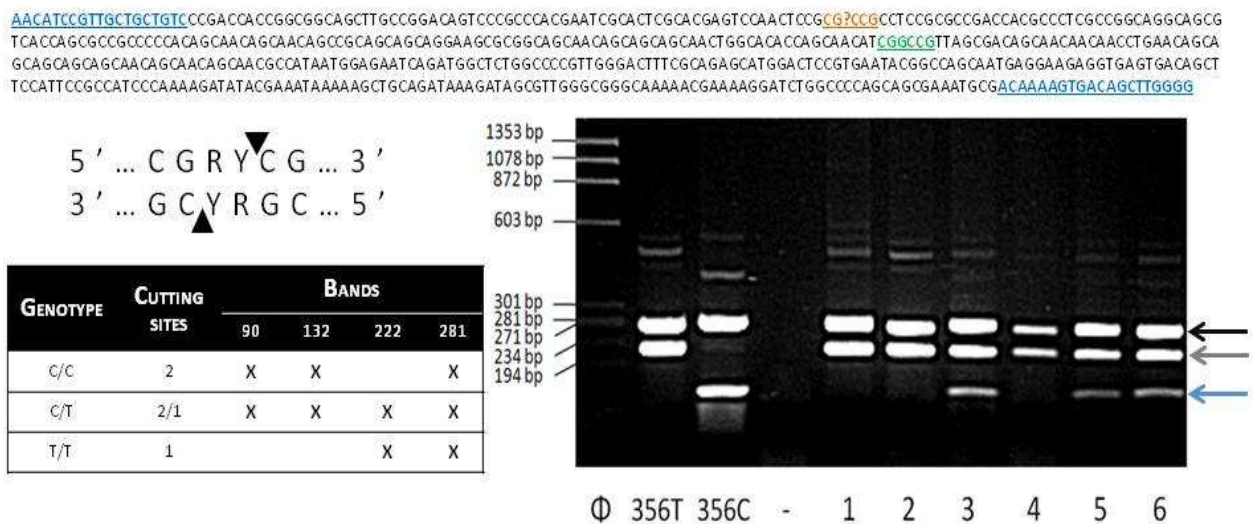
Figure 4.4
Strategy used for *cpo* SNP A356V genotyping. Top: Region of *cpo* amplified with the primers cpo-F and cpo-R (in blue). The region in green indicates the *BsiEI* cutting sites present in both genotypes. The region in orange includes the polymorphic site 356 (indicated by the question mark). This region is recognised by the enzyme *BsiEI* depending on the genotype. Bottom: On the left side, sequence recognised by the enzyme *BsiEI* (R: *A* or *G*; Y: *T* or *C*). The table illustrates the cutting sites and the bands expected for each genotype (sizes are given in bp). On the right side: 1.5% agarose gel, showing an example of outcome. 356T and 356C: internal controls, genotypes *T/T* and *C/C* respectively. The black, gray and blue arrows indicate the 3 possible bands of 281, 220 and 132 bp respectively (the 90 bp band is not visible). Genotypes of the samples: 1: *T/T*; 2: *T/T*; 3: *T/C*; 4: *T/T*; 5: *T/C*; 6: *T/C*. Φ: Phi X174 *Hae III*: molecular weight marker. (-): negative control.

Table 4.3 shows the details regarding the flies genotyped. Interestingly, after applying Bonferroni correction for multiple testing, 5 populations out of 19 are not in Hardy Weinberg Equilibrium, generally due to the frequency of heterozygotes being lower than expected. Heterozygote deficit could be an indication of population subdivision (Wahlund effect), inbreeding or positive assortative mating. Figure 4.5 shows how the allele and genotype frequencies change over latitude. In Figure 4.5 A, the data regarding the southern- and northernmost American populations are plotted in grey, in order to allow easy comparison with the European data. In Europe, the allele cpo^{356Val} shows a trend similar to the one observed in the USA, increasing significantly in frequency with latitude (R^2=0.33; p=0.01). The two trend lines show a similar slope, nevertheless the European data are more scattered and dispersed than the American data. One possibility is that SNP A356V is not a good marker for SNP I462K in Europe as it is in the USA. The linkage disequilibrium between the two sites might not be as strong in Europe, therefore leading to an underestimation of the putative cline in SNP I462K.

As regards the genotype frequencies, both the homozygotes *T/T* and *C/C* show significant latitudinal clines (R^2=0.32; p=0.01 and R^2=0.22; p=0.04 respectively), unlike the heterozygotes whose frequency does not change significantly with latitude.

		LINES DETAILS					FREQUENCIES					OBSERVED			EXPECTED			
LINE	LOCATION	COUNTRY	N	LAT	LONG	ALT	f(T)	f(C)	F(T/T)	f(T/C)	f(C/C)	# T/T	#T/C	#C/C	# T/T	#T/C	#C/C	χ^2
SP-14;15	Dalias/Algarrobo	Spain	56	**36.82**	-2.87	423	0.27	0.73	0.25	0.04	0.71	7	1	20	2.01	10.98	15.01	23.13***
SP-22	Nijar	Spain	50	**36.97**	-2.21	345	0.76	0.24	0.60	0.32	0.08	15	8	2	14.44	9.12	1.44	0.38
SP-38	Requena	Spain	194	**39.49**	-1.10	716	0.58	0.42	0.38	0.39	0.23	37	38	22	32.33	47.34	17.33	3.78
SAL	Salice	Italy	70	**40.38**	17.38	48	0.44	0.56	0.29	0.31	0.40	10	11	14	6.86	17.27	10.86	4.61
SP-44	Vandeltormo	Spain	70	**40.99**	0.08	478	0.50	0.50	0.23	0.54	0.23	8	19	8	8.75	17.50	8.75	0.26
BIT	Bitetto	Italy	82	**41.02**	16.75	149	0.72	0.28	0.59	0.27	0.15	24	11	6	21.23	16.55	3.23	4.61
SP-43	Alcaniz	Spain	38	**41.05**	-0.13	317	0.45	0.55	0.32	0.26	0.42	6	5	8	3.80	9.39	5.80	4.16
SP-52	S.Sadurni d'Anoia	Spain	68	**41.42**	1.76	199	0.66	0.34	0.53	0.26	0.21	18	9	7	14.89	15.22	3.86	5.68
CAV	Cavarzere	Italy	78	**45.13**	12.08	3	0.50	0.50	0.33	0.33	0.33	13	13	13	9.75	19.50	9.75	4.33
TRV	Treviso	Italy	66	**45.71**	12.26	7	0.62	0.38	0.58	0.09	0.33	19	3	11	12.73	15.53	4.73	21.48***
VIL	Villorba	Italy	30	**45.74**	12.23	39	0.40	0.60	0.33	0.13	0.53	5	2	8	2.40	7.20	5.40	7.82
VNN	Vienna	Austria	24	**47.79**	16.36	183	0.79	0.21	0.58	0.42	0.00	7	5	0	7.52	3.96	0.52	0.83
FUL	Fulda	Germany	14	**50.55**	9.68	267	0.64	0.36	0.57	0.14	0.29	4	1	2	2.89	3.21	0.89	3.32
HU	Houten	Holland	74	**52.03**	5.17	3	0.55	0.45	0.32	0.46	0.22	12	17	8	11.36	18.28	7.36	0.18
MAR	Market Harborough	England	88	**52.48**	-0.92	82	0.83	0.17	0.73	0.20	0.07	32	9	3	30.28	12.44	1.28	3.37
KIL	Kilworth	England	120	**52.53**	0.98	94	0.62	0.38	0.53	0.17	0.30	32	10	18	22.82	28.37	8.82	25.15***
HØJ	Højbjerg	Denmark	138	**56.11**	10.21	n.a.	0.85	0.15	0.84	0.01	0.14	58	1	10	49.60	17.80	1.60	61.47***
GOT	Goteborg	Sweeden	54	**57.70**	11.97	10	0.74	0.26	0.48	0.52	0.00	13	14	0	14.81	10.37	1.81	3.31
KOR	Korpilahti	Finland	110	**62.02**	25.55	104	0.81	0.19	0.73	0.16	0.11	40	9	6	36.00	16.99	2.00	12.17*

Table 4.3

Details of the flies used in the *cpo* SNP A356V study. N: number of alleles analysed. Lat: latitude in degrees North. Long: longitude in degrees (Negative values: West; Positive values: East). Alt: altitude in meters above sea level. The last column shows the results of the Hardy Weinberg test. *: p<0.01; ***: p<0.001.

Figure 4.5

Frequency of *cpo^356Val* in European populations, plotted against latitude (A). The gray dots represent the southern- and northern-most American populations (Schmidt *et al.*, 2008). Panels B, C and D show how the genotype frequencies (*T/T*, *T/C* and *C/C* respectively), change with latitude of collection.

4.3.2 SNP I462K IN EUROPEAN POPULATIONS

Since SNP A356V might not be a good marker for SNP I462K in Europe, natural *D. melanogaster* flies collected in Europe were also genotyped for the latter SNP. Despite many efforts made in order to develop a cheap and fast way to genotype SNP I462K, it was not possible to optimise a consistent and reliable technique. Therefore the DNA fragment under study was amplified and then sequenced. Details regarding the locations of collection of the flies used in this study are listed in Table 4.4. As shown in the table, 3 populations out of 17 are not in Hardy Weinberg Equilibrium, again because of a lack of heterozygous flies.

Figure 4.6 shows how the frequency of the allele cpo^{462Lys} and of the three different genotypes changes with latitude of collection. Contrary to expectations the alleles at SNP I462K do not vary significantly with latitude. The frequency of the allele cpo^{462Lys} is set around 60% at all the tested latitudes. Similarly, the genotype frequencies do not show any significant latitudinal cline.

	LINES DETAILS						FREQUENCIES					OBSERVED			EXPECTED			
LINE	LOCATION	COUNTRY	N	LAT	LONG	ALT	f(A)	f(T)	F(A/A)	F(A/T)	f(T/T)	# A/A	#A/T	#T/T	#A/A	#A/T	#T/T	χ^2
SP-14;15	Dalias/Algarrobo	Spain	56	36.82	-2.87	423	0.48	0.52	0.11	0.75	0.14	3	21	4	6.51	13.98	7.51	7.05
SP-22	Nijar	Spain	50	36.97	-2.21	345	0.76	0.24	0.60	0.32	0.08	15	8	2	14.44	9.12	1.44	0.38
SP-28	Jumilla	Spain	60	38.48	-1.32	520	0.52	0.48	0.40	0.23	0.37	12	7	11	8.01	14.98	7.01	8.52
SP-38	Requena	Spain	48	39.49	-1.10	716	0.38	0.63	0.21	0.33	0.46	5	8	11	3.38	11.25	9.38	2.00
SAL	Salice	Italy	26	40.38	17.38	48	0.54	0.46	0.46	0.15	0.38	6	2	5	3.77	6.46	2.77	6.20
SP-44	Vandeltormo	Spain	70	40.99	0.08	478	0.50	0.50	0.23	0.54	0.23	8	19	8	8.75	17.50	8.75	0.26
BIT	Bitetto	Italy	24	41.02	16.75	149	0.75	0.25	0.67	0.17	0.17	8	2	2	6.75	4.50	0.75	3.70
SP-43	Alcaniz	Spain	38	41.05	-0.13	317	0.61	0.39	0.37	0.47	0.16	7	9	3	6.96	9.08	2.96	0.00
SP-52	S.Sadurni d'Anoia	Spain	62	41.42	1.76	199	0.56	0.44	0.48	0.16	0.35	15	5	11	9.88	15.24	5.88	14.00*
CAV	Cavarzere	Italy	48	45.13	12.08	3	0.60	0.40	0.46	0.29	0.25	11	7	6	8.76	11.48	3.76	3.65
TRV	Treviso	Italy	66	45.71	12.26	7	0.58	0.42	0.45	0.24	0.30	15	8	10	10.94	16.12	5.94	8.37
HU	Houten	Holland	46	52.03	5.17	3	0.70	0.30	0.52	0.35	0.13	12	8	3	11.13	9.74	2.13	0.73
MAR	Market Harborough	England	88	52.48	-0.92	82	0.64	0.36	0.52	0.23	0.25	23	10	11	17.82	20.36	5.82	11.40*
KIL	Kilworth	England	120	52.53	0.98	94	0.59	0.41	0.38	0.42	0.20	23	25	12	21.00	28.99	10.00	1.14
HØJ	Højbjerg	Denmark	138	56.11	10.21	n.a.	0.73	0.27	0.46	0.54	0.00	32	37	0	36.96	27.08	4.96	9.26*
GOT	Goteborg	Sweeden	52	57.70	11.97	10	0.67	0.33	0.38	0.58	0.04	10	15	1	11.78	11.44	2.78	2.51
KOR	Korpilahti	Finland	110	62.02	25.55	104	0.52	0.48	0.20	0.64	0.16	11	35	9	14.77	27.46	12.77	4.14

Table 4.4

Details of the flies used in the *cpo* SNP I462K study. N: number of alleles analysed. Lat: latitude in degrees North. Long: longitude in degrees (Negative values: West; Positive values: East). Alt: altitude in meters above sea level. The last column shows the results of the Hardy Weinberg test. *: $p<0.05$.

Figure 4.6
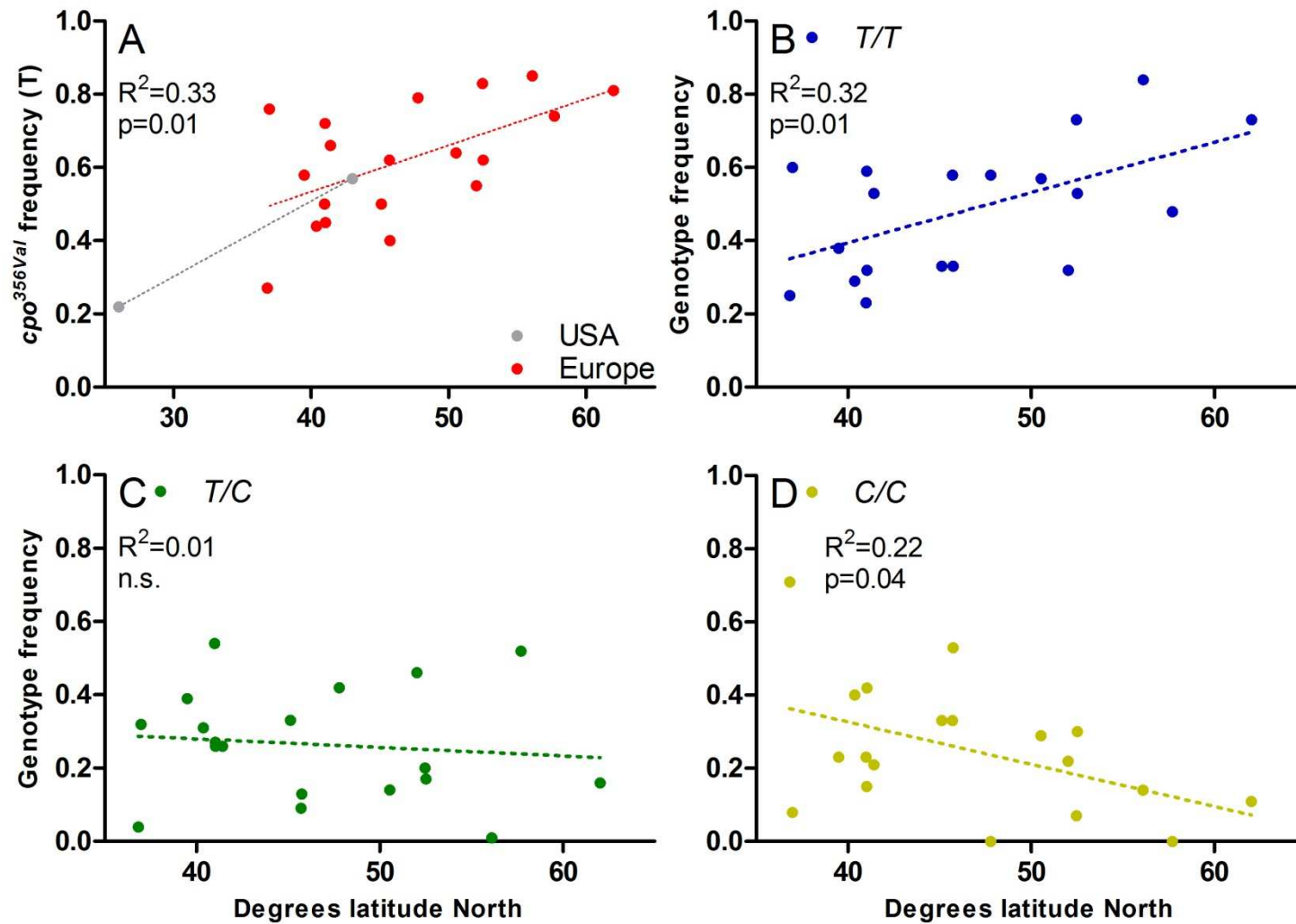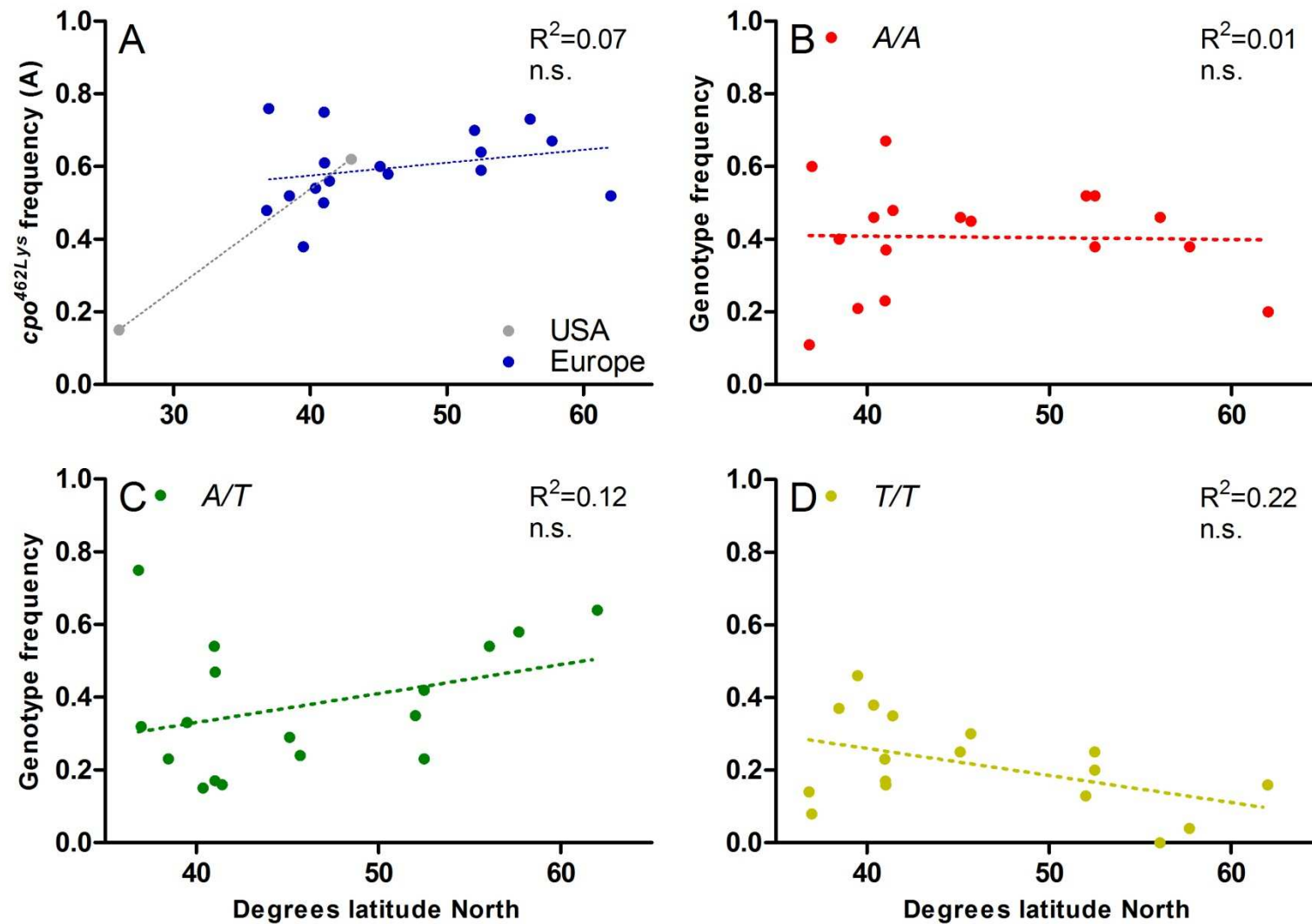
Frequency of cpo^{462Lys} in European populations, plotted against latitude (A). The gray dots represent the southern- and northern-most American populations (Schmidt *et al.*, 2008). Panels B, C and D show how the genotype frequencies (*A/A*, *A/T* and *T/T* respectively), change with latitude of collection.

4.3.3 Linkage disequilibrium

Linkage disequilibrium (LD) is defined as the non-random association of alleles which occurs between two markers if their variants are found in the same haplotype more or less often than expected based on their frequencies. A number of factors influence LD, such as the physical proximity between the two markers and selection. Proximity decreases the likelihood of recombination between two sites whereas selection might increase the chances of finding two variants together because they might interact at a functional level.

Schmidt and co-workers (2008) determined the LD between SNP A356V and SNP I462K to be D=0.216, p<0.0001. D measures the difference between the frequency of a specific haplotype, and its expected value, therefore D values significantly different from zero indicate that two variants are found in the same haplotype more often than expected based on their frequencies. Such a strong association between the two markers was expected given that they are only 317 bp apart.

The alleles at the two SNPs under study differ in their geographical distribution in Europe, therefore the strength of the linkage disequilibrium between SNP A356V and SNP I462K was analysed in our populations. The analysis was performed with the statistical program R (http://www.r-project.org) and the package "Genetics". Three estimators of D are computed, D, D' and r. Details about these three parameters can be found in Appendix 8.2. The results are shown in Table 4.5. The analysis was performed with two different approaches. First double heterozygotes (flies heterozygous at both SNPs) were not included in the analysis, since they do not allow the determination of haplotypes. Alternatively, double heterozygotes data were included in the analysis, and the observed frequency of each haplotype was estimated by the software, based on Maximum Likelihood (data obtained with this approach are indicated by "DH" in Table 4.5).

Population			Ld estimates			χ^2	
LOCATION	LAT	N	D	D'	r		
Dalias/Algarrobo	36.82	56	0.025	0.179	0.112	0.716	DH
		54	0.023	0.173	0.106	0.612	
Nijar	36.97	50	0.077	0.427	0.415	8.623**	DH
		44	0.051	0.317	0.304	4.082*	
Jumilla	38.48	60	-0.148	0.680	-0.596	21.264***	DH
		54	-0.140	0.653	-0.564	17.147***	
Requena	39.49	48	0.150	0.770	0.621	18.495***	DH
		38	0.141	0.730	0.586	13.045***	
Salice	40.38	26	0.134	0.581	0.538	7.532**	DH
		24	0.128	0.558	0.514	6.332*	
Vandeltormo	40.99	70	-0.220	0.882	-0.882	54.452***	DH
		36	-0.194	1	-0.778	21.779***	
Bitetto	41.02	12	0.091	0.497	0.473	5.365*	DH
		12	0.091	0.497	0.473	5.365*	
Alcaniz	41.05	38	-0.148	0.836	-0.607	14.007***	DH
		32	-0.133	0.809	-0.553	9.784**	
S. Sadurni d'Anoia	41.42	110	0.012	0.125	0.063	0.434	DH
		60	0.082	0.561	0.377	8.537**	
Cavarzere	45.13	48	0.080	0.634	0.402	7.745	DH
		44	0.066	0.593	0.352	5.455	
Treviso	45.71	66	0.141	0.650	0.585	22.613***	DH
		62	0.135	0.631	0.564	19.758***	
Houten	52.03	46	0.106	0.999	0.640	18.847***	DH
		40	0.072	0.999	0.541	11.693***	
Market Harborough	52.48	88	0.088	0.766	0.483	20.546***	DH
		80	0.070	0.719	0.417	13.920***	
Kilworth	52.53	120	0.120	0.550	0.503	30.349***	DH
		104	0.104	0.497	0.448	20.830***	
Hφjbjerg	56.11	138	0.030	0.273	0.187	4.800*	DH
		136	0.027	0.258	0.174	4.112*	
Goteborg	57.70	52	0.134	0.632	0.614	16.627***	DH
		38	0.084	0.454	0.439	7.318**	
Korpilahti	62.02	110	0.012	0.125	0.063	0.434	DH
		98	0.008	0.103	0.046	0.205	
ALL EUROPEAN POPULATIONS		1138	0.105	0.536	0.457	237.840***	DH
		950	0.090	0.480	0.398	149.540***	
ALL AMERICAN POPULATIONS		75	0.216	0.909	0.909	37.986***	

Table 4.5
LD values calculated for each individual population, and for the whole dataset. The analysis was carried out including the heterozygotes for both SNPs (DH, Double Heterozygotes), and excluding them from the dataset. The last row shows the results obtained by Schmidt and colleagues (2008) in American populations, using 75 extracted chromosomes (personal communication). N: number of haplotypes analysed. For LD estimates see Appendix 8.2. *: $\chi^2 > 3.84$; $p < 0.05$; **: $\chi^2 > 6.63$; $p < 0.01$; ***: $\chi^2 > 10.80$; $p < 0.001$.

The two approaches (considering or not the double heterozygotes) generated very similar results: not surprisingly most populations show a very strong linkage disequilibrium between the two sites, leading to an overall D' value of 0.536 (0.480 when double heterozygotes are excluded from the analysis). Remarkably the LD is much stronger in American populations, where D' reaches a value 0.909. Such a difference does not come as a surprise, given that in Europe SNPA356V does not seem to be a good marker to follow the distribution of SNP I462K. The stronger linkage disequilibrium in American populations might be the result of the bottle-neck experienced by *D. melanogaster* flies upon their colonisation of the new continent, a few hundred years ago.

4.3.4 PAYNE CHROMOSOMAL INVERSION

Lee and coworkers analysed the frequency of SNP I462K in Australian populations, and found that, similar to the American dataset, the cpo^{462Lys} allele increases in frequency with latitude (Lee *et al.*, 2011). Nevertheless the SNP is in strong linkage disequilibrium with the chromosomal inversion *In(3R)P*, which shows a strong latitudinal cline in frequency in Australian populations (Knibb *et al.*, 1981). The frequency of the inverted chromosome goes from 68% at a latitude of ~9° South, dropping to zero above 40° South. When only standard chromosomes were considered in the analysis, the association between cpo^{462Lys} and latitude ceased to be significant (Lee *et al.*, 2011). *In(3R)P* has been reported to vary significantly with latitude not only in Australian populations, but also in America and Asia (Knibb, 1982) where

again the frequency of inverted chromosomes is the highest close to the Equator, reaching

values of around 10% at latitudes above 35 degrees.

We therefore decided to analyse the frequency of inverted chromosomes in a subset

of our populations, to verify if the cline we observe at *cpo* SNP A356V is driven by the

inversion. In order to do that, flies were genotyped for a SNP in complete linkage

disequilibrium with *In(3R)P* (Anderson *et al.*, 2005). Table 4.6 shows the details of the flies

used in this analysis.

LINES DETAILS							FREQUENCIES	
LINE CODE	LOCATION	COUNTRY	N	LAT	LONG	ALT	f(st)	f(in)
SP-22	Nijar	Spain	40	**36.97**	-2.21	345	0.85	0.15
SP-44	Vandeltormo	Spain	58	**40.99**	0.08	478	0.91	0.09
BIT	Bitetto	Italy	18	**41.02**	16.75	149	1	0
SP-43	Alcaniz	Spain	60	**41.05**	-0.13	317	0.97	0.03
CAV	Cavarzere	Italy	48	**45.13**	12.08	3	0.98	0.02
HU	Houten	Holland	66	**52.03**	5.17	3	0.95	0.05
MAR	Market Harborough	England	62	**52.48**	-0.92	82	0.85	0.15
HФJ	Hфjbjerg	Denmark	68	**56.11**	10.21	n.a.	1	0

Table 4.6
Frequency of *In(3R)P* in a subset of the European populations. N: number of alleles analysed.
f(st) frequency of standard arrangements. f(in) frequency of inverted chromosomes.

As expected, the frequency of inverted arrangements is very low in our dataset,

(ranging from 0% to 15%) and it is in line with data collected at similar latitudes in USA, Asia

and Australia (Knibb, 1982). The frequency does not vary significantly with latitude (R^2=0.03),

therefore the cline in SNP A356V in Europe is unlikely to be driven by the frequencies of this

chromosomal inversion.
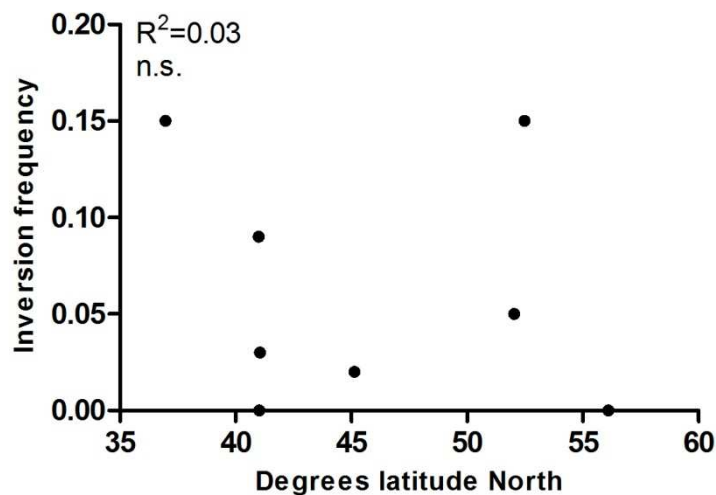
Figure 4.7
Distribution of the frequency of the inversion *In(3R)P* according to latitude of collection.

4.3.5 NEUTRALITY TESTS

In order to verify if the *cpo* locus is under selection we applied the Tajima's, Fu and Li's and HKA Test statistics to the region including the polymorphisms, and to an intergenic region upstream the polymorphic site. Table 4.7 shows the details of the DNA regions and flies analysed.

DNA REGIONS			N
Name	Location relative to SNP A356V	Length	
5'	97 kb UPSTREAM	876	39
cpo		1401	34
LOCATIONS OF COLLECTION			N
Code	Location		Latitude	N (5')	N (cpo)

Code	Location		Latitude	N (5')	N (cpo)
SP22	Nijar	Spain	36.97	11	10
BIT	Bitetto	Italy	41.02	6	5
TRV	Treviso	Italy	45.71	10	9
KOR	Korpilahti	FInland	62.02	12	10

Table 4.7
Upper panel: details regarding the two genomic regions amplified, sequenced and analysed. Their length is reported in base pairs. N: number of individuals sequenced. Bottom panel: details of the four geographical regions analysed. Latitude is reported in degrees North. N: number of flies analysed for the upstream and polymorphic region (referred to as 5' and *cpo* respectively).

Given the low number of sequences for each location, unlike Chapter 3, the neutrality tests were carried out by pooling all the sequences together and considering them as a single population.

4.3.5.1 Tajima's test

As already discussed in paragraph 3.4.2 in Chapter 3, this test is performed in order to verify if π (nucleotide diversity, measures the average number of pairwise differences in the dataset) and Ɵ (Watterson's estimate of average number of segregating sites) differ significantly (Tajima, 1989). Under neutrality these two values are expected to be the same, so any significant deviation points towards the action of natural selection, or any other phenomenon perturbing equilibrium (e.g. migration and population size changes).

An inspection of the alignment of the two genomic regions under study highlighted a remarkable number of insertion/deletion (indel) polymorphisms in the *cpo* region characterised by the two polymorphisms (Table 4.8). The Tajima's tests was therefore computed with the software NeutralityTest 1.1, kindly provided by Haipeng Li (Li & Fu, 2004) which, unlike DNAsp 5.10.01, offers the possibility of including indels in the analysis.

	TOTAL NUMBER INDELS EVENTS	AVERAGE INDEL LENGTH EVENT	AVERAGE INDEL LENGTH	NUMBER OF INDEL HAPLOTYPES
5'	4	2.250	2.333	7
cpo	21	14.524	6.922	21

Table 4.8
Details regarding the indel polymorphisms in the two genomic regions under study. The values were calculated with DNAsp v5.10.01, using a multialleic model (all alleles are considered, also the ones which overlap in sequence). The average indel length event is calculated averaging the length of every indel event in the alignment. The average indel is calculated considering every indel event in each single sequence. Values are given in nucleotides.

The results of the Tajima's test, calculated with NeutralityTest v1.1, are shown in Table 4.9. The shaded rows show the results obtained including the indel sites, whereas the non-shaded ones show the result of the analysis without including alignment gaps.

REGION	INDELS INCLUDED	TOT SITES	SEGREGATING SITES	TAJIMA'S D	D_1	D_2	D_3	D_4
5'	YES	876	38	0.579				
	NO	868	30	0.077				
cpo	YES	1401	292	-2.087*				
	NO	1146	29	0.090	0.57	1.05	-0.35	0.20

Table 4.9

Results of the Tajima's test performed with NeutralityTest v1.1, on the two genomic regions under study. Shaded cells show the results obtained including indel sites in the analysis, non shaded cells show the results obtained when those sites are not included. D_1, D_2, D_3 and D_4 indicate the Tajima's D calculated only on a subset of sites. D_1: coding region; D_2: synonymous sites; D_3: non synonymous sites; D_4: silent sites (include synonymous sites and other sites which do not lead to an amino acid substitutions, introns and non coding regions) * indicates significant values ($p<0.05$).

Tajima's D reaches a significantly negative value, indicative of directional selection, only in the 'cpo' region, and only when indels are included in the analysis. When DNAsp was used to calculate Tajima's D only in a subset of sites of the genomic 'cpo' (coding region, synonymous and non synonymous sites and silent sites), the result was not significant.

4.3.5.2 FU'S AND LI'S TEST STATISTICS

As in Chapter 3, also the Fu and Li's test statistic (Fu & Li, 1993) was calculated on this set of sequences. In order to include indel polymorphisms in the analysis, the NeutralityTest v1.1 software was used, and the results are shown in Table 4.10.

REGION	INDELS INCLUDED	TOT SITES	SEGREGATING SITES	D*	F*
5'	YES	876	38	0.618	0.670
	NO	868	30	0.301	0.191
cpo	YES	1401	292	-4.147**	-0.002**
	NO	1146	29	0.07	0.085

Table 4.10

Results of the Fu and Li's test. Shaded cells show the results obtained including indels, whereas non shaded rows show the results calculated not including them. ** indicates significant values ($p<0.01$).

As for the Tajima's test, the results are significant only in the genomic region including part of the *cpo* coding region. Furthermore only when indel polymorphisms are considered do the values reach significance.

4.3.5.3 HKA TEST

The Hudson-Kreitman-Aguadè test (Hudson *et al.*, 1987) can be computed with the NeutralityTest v1.1 software only if the two loci under study have exactly the same length. For this reason, as in chapter 3, this analysis was performed with DNAsp and does not include indel polymorphism. Sequences from *D. simulans* retrieved from the database FlyBase were used for the interspecific comparison.

DNA REGION	TOT SITES	S OBS	S EXP	DIFF OBS	DIFF EXP	χ^2 (P-value)
5'	876	29	28.11	19.23	20.12	0.04
CPO	1142	27	27.89	21.88	20.99	(0.84)

Table 4.11
Results of the HKA test performed on the polymorphic region (CPO) and on a upstream region (5'). Total number of sites is given in base pairs and sites with gaps are excluded from the analysis. S OBS: observed number of segregating sites (intraspecies comparison). S EXP: expected number of segregating sites (intraspecies comparison). DIFF OBS: observed number of differences (interspecies comparison). DIFF EXP: expected number of differences (interspecies comparison). The analysis was carried out with the software DNAsp 5.10.

As shown in Table 4.11 the result of the HKA test are not significant, the sequences do not contain more (or less) mutations than explainable by variations in the mutation rates affecting the two loci.

4.3.6 SNPs IN *D. simulans*

Schmidt and coworkers stated that, based on comparison with other *Drosophila* species, cpo^{356Val} and cpo^{462Lys} represent the derived allele (Schmidt *et al.*, 2008). Given the availability in the laboratory of *D. simulans* lines captured in different locations in Africa, the

cpo region including the two SNPs was amplified and sequenced in *D. simulans* flies in order to confirm which alleles are derived.

The sequences obtained suggest that the position referred to as 356 is not polymorphic in *D. simulans*. All the flies analysed are characterised by the nucleotide *C* in this position, thus encoding the amino acid Alanine and confirming that *cpo*^{356Val} is in fact the derived allele. As regards SNP I462K, an interesting situation emerged. As shown in Figure 4.8, the nucleotide position responsible for SNP I462K is not polymorphic in *D. simulans* (blue shaded column), but it is characterised by the *C* nucleotide, whereas in the sibling species *D. melanogaster* either a *T* or an *A* are present. Nevertheless the nucleotide position immediately upstream (yellow shaded column) was found to be polymorphic in *D. simulans*. When the amino acid sequence is considered, SNP462 is characterised by either an Alanine (*A*) or a Threonine (*T*) in *D. simulans*

Figure 4.8

Top: Nucleotide alignment of the *cpo* region including SNP 462 in *D. simulans* flies. The blue shaded position represents the polymorphic site responsible for SNP I462K in *D. melanogaster*. The yellow shaded box indicates the polymorphic site in *D. simulans*. Bottom: Protein alignment of the same region as above. In green the polymorphic site corresponding to SNP 462 in *D. melanogaster*. Table: Sequence of the possible codons generating SNP 462 in *D. melanogaster* and *D. simulans*.

To explain the current situation and account for the differences between African *D. simulans* sequences and the *D. melanogaster* ones, at least 3 independent changes in this codon are required. A high level of variability in this region, at the amino acid level and stop

codon position, has been reported also by Kankare and colleagues (Kankare *et al.*, 2011) who compared the 3' region of *cpo* exon 5 in five *D. virilis* group species. This high variability could be an indication of relaxed selection at this particular site which might be subjected to low structural/functional constraints.

4.3.7 EXPRESSION STUDY

At the time this project was initiated, *cpo* was reported to have 6 expression isoforms (Figure 4.2), only one of them (*cpo*-RH), containing SNP I462K. We therefore decided to check under which conditions *cpo*-RH was expressed at its highest level. Flies were kept at three different temperatures (12°C, 18°C and 25°C), they were then collected at different time points (3 and 15 hours after lights-on, in a regime of 12 hours of light followed by 12 hours of darkeness), and mRNA was extracted from the heads and the bodies, and retro-transcribed to cDNA. Two PCR reactions were then performed with specific primers to amplify either only isoform RH or all the isoforms but *cpo*-RH. Three independent replicates were performed for each condition and Figure 4.9 shows an example of the outcome of the semi-quantitative PCR reaction. Surprisingly *cpo*-RH could not be amplified in any of the aforementioned conditions.

There are many reasons which could explain the lack of amplification of *cpo*-RH. This isoform, which is much shorter than the others and lacks the RNA binding domain, might be expressed only in conditions which were not assessed in our experiment (e.g. after exposure to different temperatures and/or photoperiods). Alternatively it might only be expressed in a small subset of neurons, thus making its identification problematic. Finally, *cpo*-RH might only be expressed during earlier development, and its expression might be down regulated or even switched-off in adults.
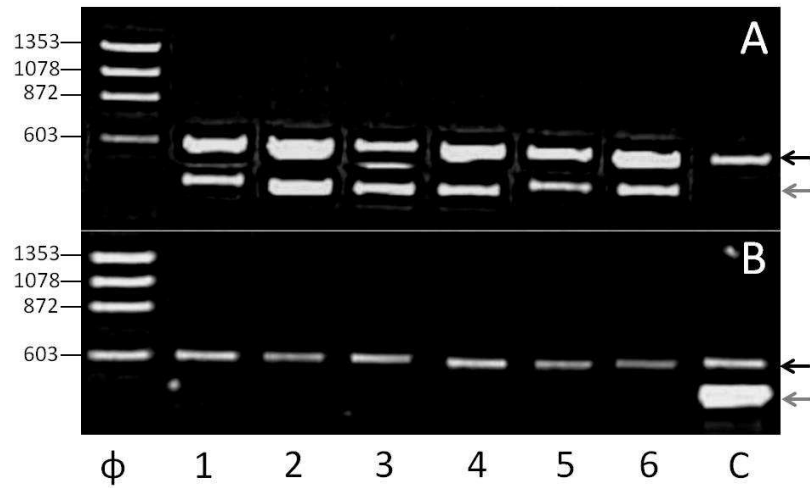
Figure 4.9
Example of outcome of a PCR reaction performed to check the expression of the isoform *cpo*-RH. A: reaction including primers to amplify all isoforms but *cpo*-RH (band indicated by the gray arrow) and the *GapDH* gene, used as internal control (black arrow). B: reaction to amplify only the isoform *cpo*-RH (gray arrow). The black arrow indicates the *GapDH gene*, amplified as a control of the efficiency of the reaction. Samples: 1: 12°C Head; 2: 12°C Body; 3: 18°C Head; 4: 18°C Body; 5: 25°C Head; 6: 25°C Body. C: genomic DNA used as a control. These samples were collected at ZT 15. Φ: Phi X174 *Hae III*: molecular weight marker.

At the time this thesis was written, the Flybase annotation of *cpo* was updated and it no longer included isoform *cpo*-RH. Remarkably the (now seven) isoforms vary in the 5' and 3' UTRs, the coding region remains unchanged, and none of them includes SNP I462K (Figure 4.10).
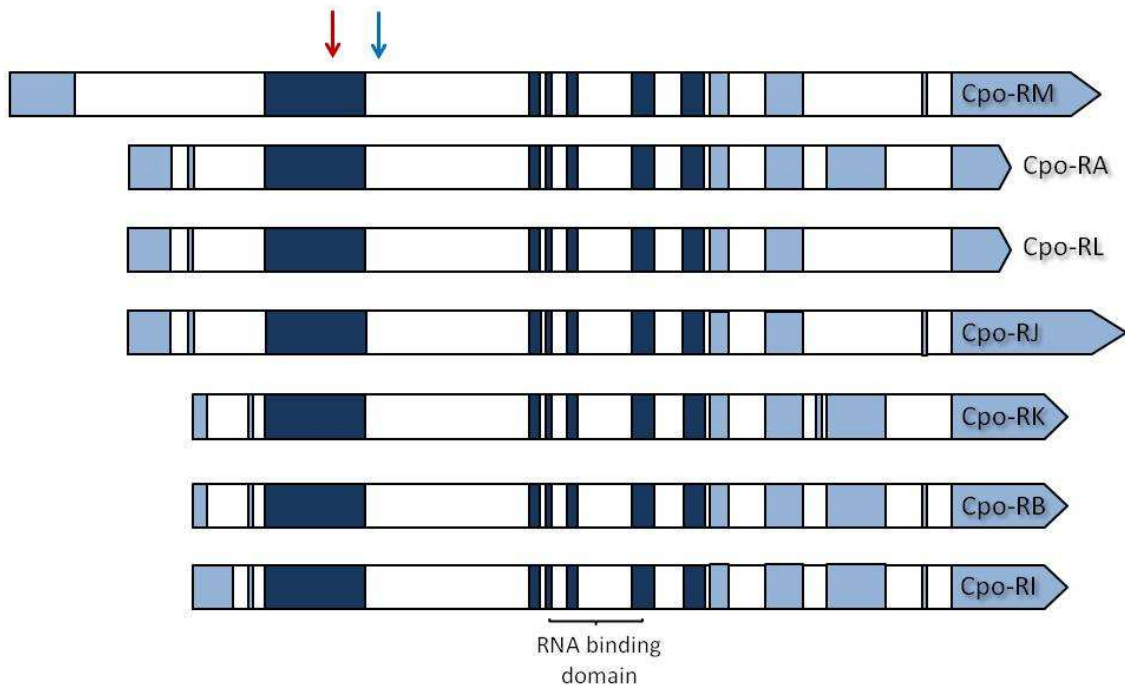
Figure 4.10
Schematic representation of the seven *cpo* splicing variants. Blue boxes encode the protein whereas light blue ones represent the 5' and 3' UTR regions. The red and blue arrows represent respectively SNP A356V and SNP I462K. The figure is not to scale and was redrawn from the database Flybase as it appeared at the time this thesis was written (February 2012).

4.3.8 BEHAVIOURAL STUDY

Given the lack of cline in *cpo* SNP I462K in European natural flies, the probable lack of expression of isoform *cpo*-RH and the results obtained by Lee and colleagues in Australian flies (Lee *et al.*, 2011), it was decided to create different fly lines characterised by different combinations of the *cpo* SNPs A356V and I462K. These lines were created using as a starting point a population obtained by combining equal numbers of non-virgin female flies from 35 isofemale lines collected in Treviso (Lat 45.71°N, Long 12.26°E). This location was chosen because it is characterised by an allelic frequency of ~50% for both SNPs. Table 4.12 lists the genotypes of the 4 lines created.

FLY LINE	*tim* ALLELE	SNP A356V	SNP I462K
sTA	s/s-tim	T/T cpo^{356Val}	A/A cpo^{462Lys}
sTT	s/s-tim	T/T cpo^{356Val}	T/T cpo^{462Ile}
sCA	s/s-tim	C/C cpo^{356Ala}	A/A cpo^{462Lys}
sCT	s/s-tim	C/C cpo^{356Ala}	T/T cpo^{462Ile}

Table 4.12
tim and *cpo* genotypes of the four lines created using as a starter point a population from Treviso, Italy.

The populations were then expanded and their phenotype was characterised in terms of diapause incidence. The experiment was performed at two different photoperiods (LD 8:16 and LD 16:8) and flies were dissected at two different time points: after 12 and 28 days. 5/6 replicates were performed for each condition, together with two DD (constant dark) controls. These were performed like normal replicates except they were wrapped in aluminum foil, in order to expose them only to the small temperature fluctuations caused by the different photoperiods (~0.5°). A total of 20-30 female flies were dissected and scored for each replicate, leading to a total of 3901 flies analysed in this experiment. The results are shown in Figure 4.11 (the results regarding the DD controls are shown in Appendix 8.3). Diapause was also assessed in the whole Treviso population, from which the other lines were created (referred to as 'TOT' in Figure 4.11, gray bars). Nevertheless, this population was not included in the statistical analysis. The diapause incidence values were then subjected to arcsine square root transformation, and used for statistical analysis. The result of the 3-way ANOVAs are shown in Table 4.13.

A) As for the <u>12 days</u> data (Table 4.13 A), the statistical analysis revealed that SNP A356V influences diapause levels ($F_{1,39}$=42.68, p<10^{-7}), and not surprisingly so does the photoperiod at which flies are exposed ($F_{1,39}$=21.74, p<10^{-3}). Flies characterised by *cpo*^{356Ala} seem to respond more to photoperiod than flies characterised by the other allele (SNP X Photoperiod interaction $F_{1,39}$=4.45, p=0.04). Interestingly the SNP A356V effect persists also in the DD controls ($F_{1,8}$=7.98, p=0.02).

B) After <u>28 days</u> of exposure to diapause conditions (Table 4.13 B), SNP A356V still has a significant impact on the diapause level, in both LD and DD conditions ($F_{1,39}=401.47$, $p<10^{-21}$ and $F_{1,8}=6.12$, p=0.04 respectively). At 28 days the photoperiodic effect is not significant anymore, the flies responding similarly to the two conditions.

C) When data collected at <u>LD 8:16</u> were considered (Table 4.13 C), again SNP A356V significantly influenced diapause levels, in both LD and DD conditions ($F_{1,40}=113.59$, $p<10^{-12}$ and $F_{1,8}=6.32$, p=0.04 respectively). A significant fall in diapause levels is observed after 28 days ($F_{1,40}=89.40$, $p<10^{-12}$), especially in the genotype cpo^{356Ala} (SNP X Days interaction $F_{1,40}=37.73$, $p<10^{-7}$).

D) A very similar pattern was observed when <u>LD 16:8</u> data were considered (Table 4.13 D): a significant effect of SNP A356V in both LD and DD conditions ($F_{1,39}=232.96$, $p<10^{-17}$ and $F_{1,8}=6.56$, p=0.03 respectively), a decrease in diapause levels ($F_{1,39}=40.89$, $p<10^{-6}$), especially regarding the genotype cpo^{356Ala} (SNP 356 X Days interaction $F_{1,39}=25.71$, $p<10^{-4}$).

Overall *cpo* SNP A365V seems to influence not only the diapause levels, but also photoperiodism and the level of diapause decline over time. In particular having the derived variant cpo^{356Val} seems to increase diapause levels and maintain them relatively high also in long photoperiods, and for a considerably long time. Surprisingly, SNP I462K does not seem to influence diapause levels nor the capability of flies to respond to different photoperiods, in any of the conditions tested.
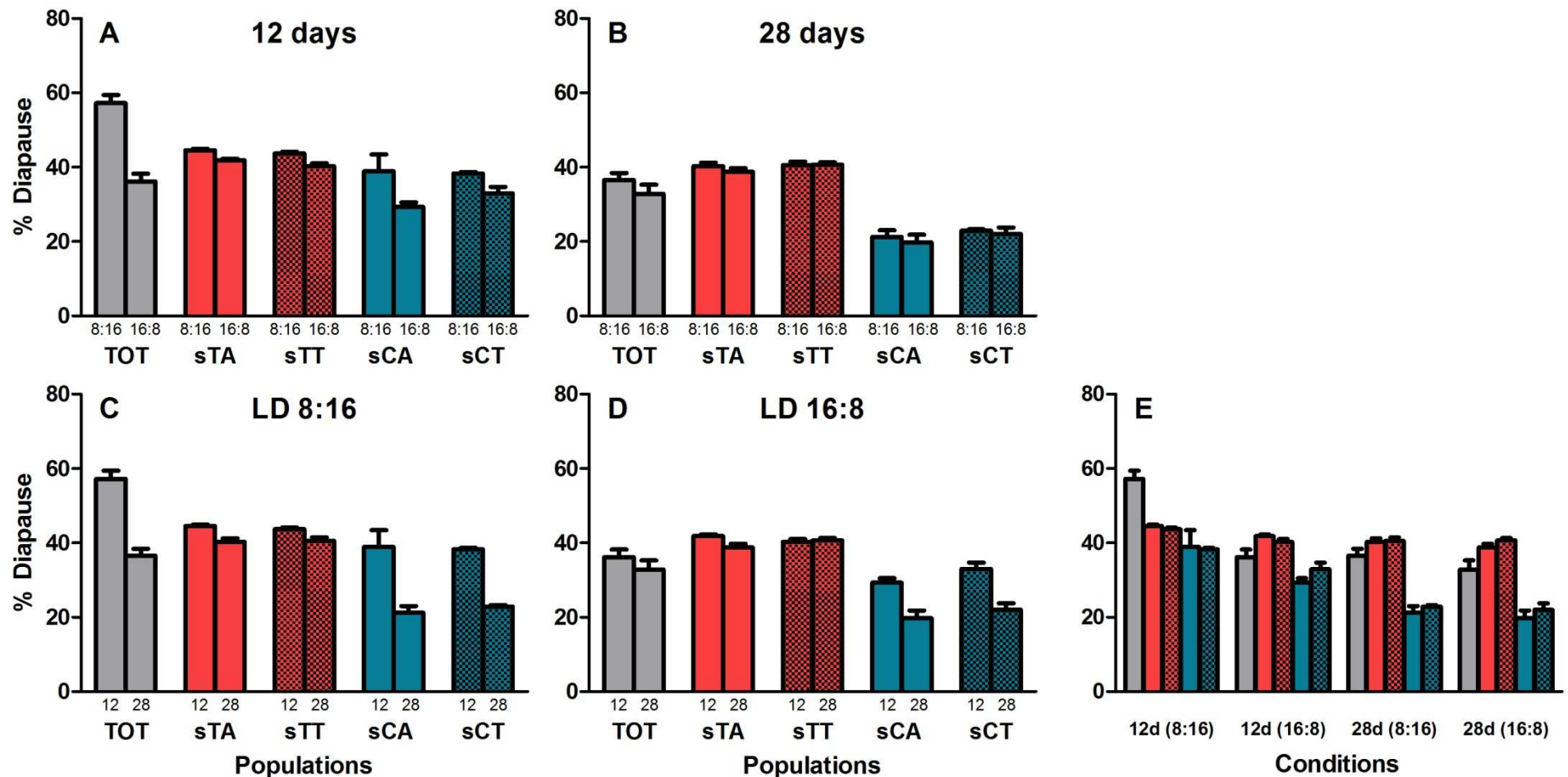
Figure 4.12

Diapause incidence in the Treviso population (TOT) and in the four lines created (sTA, sTT, sCA and sCT, see Table 4.11). Y axis: percentage of diapause averaged between 5/6 replicates (±SEM). The same data is clustered in different ways for clarity. A) Diapause assessed after 12 days, at two different photoperiods (LD 8:16 and LD 16:8). B) Diapause assessed after 28 days, at the two different photoperiods. C) Diapause tested at LD 8:16. The two time points are compared (12 and 28 days). D) Diapause levels at LD 16:8, comparison between the two time points. E) Diapause levels in the four conditions tested. Gray bars: Treviso population. Different bar colours indicate different nucleotides in *cpo* SNP A356V (red: T; blue: C), whereas different bar patterns indicate differences in *cpo* SNP I462K (plain: A; dotted: T).

Tables A and B — ALL EFFECTS (12 days) and ALL EFFECTS (28 days)

	df Effect	MS Effect	df Error	MS Error	F	P value	MS Effect	Df Error	MS Error	F	p value
	A ALL EFFECTS (12 days)						**B** ALL EFFECTS (28 days)				
SNP 356	1	0.073	39	0.002	42.678	$<10^{-7}$	0.491	39	0.001	401.469	$<10^{-21}$
	1	0.043	8	0.005	7.978	0.022	0.088	8	0.014	6.123	0.038
SNP 462	1	$<10^{-5}$	39	0.002	0.004	0.949	0.004	39	0.001	3.089	0.087
	1	0.013	8	0.005	2.356	0.163	0.023	8	0.014	1.613	0.240
Photoperiod	1	0.037	39	0.002	21.743	$<10^{-3}$	0.001	39	0.001	1.118	0.297
	1	0.002	8	0.005	0.418	0.536	0.003	8	0.014	0.227	0.646
SNP 356 X SNP 462	1	0.002	39	0.002	1.116	0.297	0.001	39	0.001	0.453	0.505
	1	0.012	8	0.005	2.242	0.173	0.021	8	0.014	1.437	0.265
SNP 356 X Photoperiod	1	0.008	39	0.002	4.449	0.041	$<10^{-3}$	39	0.001	0.141	0.709
	1	$<10^{-3}$	8	0.005	0.080	0.784	0.002	8	0.014	0.121	0.737
SNP 462 X Photoperiod	1	0.001	39	0.002	0.814	0.373	$<10^{-3}$	39	0.001	0.335	0.566
	1	0.001	8	0.005	0.112	0.746	0.003	8	0.014	0.184	0.679
SNP 356 X SNP 462 X Photoperiod	1	0.003	39	0.002	1.482	0.231	$<10^{-3}$	39	0.001	0.085	0.772
	1	$<10^{-3}$	8	0.005	0.060	0.813	0.004	8	0.014	0.245	0.634

Tables C and D — ALL EFFECTS (LD 8:16) and ALL EFFECTS (LD 16:8)

	df Effect	MS Effect	df Error	MS Error	F	P value	MS Effect	Df Error	MS Error	F	p value
	C ALL EFFECTS (LD 8:16)						**D** ALL EFFECTS (LD 16:8)				
SNP 356	1	0.190	40	0.002	113.58	$<10^{-12}$	0.287	39	0.001	232.96	$<10^{-17}$
	1	0.069	8	0.011	6.325	0.036	0.058	8	0.009	6.561	0.034
SNP 462	1	$<10^{-4}$	40	0.002	0.024	0.878	0.004	39	0.001	3.006	0.091
	1	0.021	8	0.011	1.955	0.200	0.014	8	0.009	1.592	0.243
Days	1	0.150	40	0.002	89.401	$<10^{-11}$	0.050	39	0.001	40.893	$<10^{-6}$
	1	0.050	8	0.011	4.610	0.064	0.014	8	0.009	1.608	0.240
SNP 356 X SNP 462	1	$<10^{-2}$	40	0.002	0.068	0.796	0.003	39	0.001	2.343	0.134
	1	0.022	8	0.011	1.998	0.195	0.011	8	0.009	1.265	0.293
SNP 356 X Days	1	0.063	40	0.002	37.734	$<10^{-6}$	0.032	39	0.001	25.705	$<10^{-4}$
	1	0.006	8	0.011	0.536	0.485	$<10^{-2}$	8	0.009	0.022	0.887
SNP 462 X Days	1	0.002	40	0.002	1.018	0.319	$<10^{-2}$	39	0.001	0.352	0.556
	1	0.003	8	0.011	0.310	0.593	$<10^{-2}$	8	0.009	0.036	0.854
SNP 356 X SNP 462 X Days	1	$<10^{-2}$	40	0.002	0.180	0.674	0.002	39	0.001	1.325	0.257
	1	0.003	8	0.011	0.286	0.607	$<10^{-2}$	8	0.009	0.052	0.825

Table 4.13

Results of 3-way ANOVAs performed on the 12 day, 28 day, LD 8:18 and LD 16:8 datasets (Table A, B, C and D respectively). Shaded cells show the results regarding the DD controls. Significant values are indicated in red.

4.4 DISCUSSION

cpo is a gene whose expression has recently been associated to diapause incidence not only in *D. melanogaster*, but in many other insects. The direction of the change in its expression levels is not uniform among different species or developmental stages. *cpo* levels are low in diapausing *Sarcophaga crassipalpis* pupae (Ragland *et al.*, 2010), but *cpo* expression was found to be upregulated in diapausing *D. melanogaster* (Schmidt *et al.*, 2008), *D. montana* flies (Kankare *et al.*, 2010) and in the mosquito *Culex pipiens* (Zhang & Denlinger, 2011).

The link between *cpo* and diapause incidence is also suggested by the domains present in its sequence: other than an RNA recognition motif (RRM), *cpo* also contains a conserved domain which shows similarity to type 1 antifreeze proteins (AFP) and/or Kv 1.4 voltage-gated potassium channels (Kankare *et al.*, 2011). These proteins have been reported to be connected to cold tolerance in insects and spiders (Duman *et al.*, 2004) and in cold hardening in an Antarctic midge (Teets *et al.*, 2008). Furthermore *cpo* contains ecdysone response elements, suggesting that its influence on diapause might be mediated by ecdysteroids (Emerson *et al.*, 2009a).

Schmidt and coworkers identified a region at the 3' end of *cpo* exon 5 to be responsible for the diapause variation in American *D. melanogaster* lines. This region contains two non-synonymous polymorphisms, *cpo* SNP A356V and SNP I462K, the latter alone accounting for all the variance in diapause incidence in the sampled sequences (Schmidt *et al.*, 2008). Given the strong linkage disequilibrium between the two sites, the former was used as a marker for SNP I462K to genotype 11 populations captured along the US eastern coast. A significant latitudinal cline was detected for SNP A356V, with the derived allele *cpo*[356Val] increasing in frequency with latitude. Based only on the frequencies found in the two most extreme latitudes, a steeper cline was inferred for SNP I462K, which is considered to concur

with other (known and unknown) genetic factors in generating the steep cline in diapause incidence observed in American populations (Schmidt *et al.*, 2005a). The fact that the derived alleles (*cpo*[356Val] and *cpo*[462Lys]) both increase in frequency with latitude is another indication that the new alleles confer evolutionary advantages at temperate latitudes, and is consistent with and Afrotropical origin of *D. melanogaster* (David & Capy, 1988), with selection favouring new mutations advantageous in northern habitats (Hoffmann & Weeks, 2007; Sezgin *et al.*, 2004; Paaby *et al.*, 2010).

One way to confirm that a cline in allele frequency is due to adaptation to the environment is to evaluate the presence of the cline at different times, and correlate environmental changes with potential changes in the steepness of the allele frequency cline (Umina *et al.*, 2005). Alternatively, the presence of the same cline in different continents also supports the possibility that the polymorphism under study has an adaptive role. Many natural polymorphisms in *D. melanogaster* show a similar trend in frequency in different continents, to name a few, the indels in the *Insuline-like-Receptor* gene (Paaby *et al.*, 2010), chromosomal inversions (Knibb, 1982), the number of Thr-Gly repeats in the *period* gene (Sawyer *et al.*, 2006; Costa *et al.*, 1992), and the very well studied polymorphisms in the *Alcohol dehydrogenase* gene (Oakeshott *et al.*, 1982).

In order to confirm the evolutionary advantage conferred by *cpo*[462Lys] and its implication in temperate habitat adaptation, the frequency of the derived allele has been recently checked in Australian natural populations. Surprisingly it has been shown to present a latitudinal cline exclusively driven by the Payne chromosomal inversion, and not to be involved in the regulation of diapause incidence in the Australian continent (Lee *et al.*, 2011).

As for European flies, *cpo* SNP A356V shows a significant yet quite dispersed cline in allele frequency. We first hypothesised that such a *scenario* emerged because a marker SNP was being assessed, rather than the polymorphism directly implicated in regulating the

diapause phenotype. A not-so-strong linkage disequilibrium between the two SNPs in European populations would be compatible with the bottle-neck faced by *D. melanogaster* populations upon their recent colonisation of North America, and could be responsible for different geographical distributions of two quite closely located SNPs. The amount of linkage disequilibrium between the two SNPs is in fact lower in Europe than in the USA, but surprisingly this leads to a flat geographical distribution of the alleles at SNP I462K, rather than to a stronger cline. Overall the situation in Europe resembles more the one in Australia where, when the frequency of the chromosomal inversion was taken into account, no cline in SNP I462K was detected.

Interestingly many of the populations analysed are not in Hardy Weinberg Equilibrium (5 out of 19 and 3 out of 17 for SNP 356 and 462 respectively), and they are characterised by fewer heterozygotes than expected. A lack of heterozygotes can be attributed to subpopulation structure (Wahlund effect). Nevertheless this phenomenon can be excluded here, since the same populations appear to be in equilibrium when other loci are considered (see Chapter 3). Alternatively, lack of heterozygotes can arise when "Disruptive selection" (also referred to as "Diversifying selection"), is acting on the population. As a consequence of the different selective pressures of a non-homogenous environment, disruptive selection acts against the heterozygotes, favouring the extreme values of a trait. Schmidt and co-workers showed that flies characterised by one high- and one low-diapausing allele show a phenotype intermediate between the one of the two homozygous (Schmidt *et al.*, 2008). If *cpo*^{356Val} and *cpo*^{356Ala} are being selected for in the North and in the South respectively, this could lead to a general decrease of heterozygous whose phenotype is intermediate therefore slightly disadvantageous at any latitude.

From a phenotypic perspective, European flies behave differently from American and Australian natural lines. Whereas in America and Australia diapause incidence decreases as

the distance from the equator grows smaller (Schmidt *et al.*, 2005a; Lee *et al.*, 2011), such a correlation is not present in European flies, at least when diapause levels are checked after 28 days of diapause inducing conditions (see Chapter 6, Pegoraro, Zonato *et al.*, manuscript in preparation). This situation points to a *scenario* still compatible with the role of SNP I462K proposed by Schmidt and colleagues, and where selection at this SNP is relaxed in Europe at both allelic frequency and phenotypic levels.

Nevertheless a subsequent set of experiments raised some doubts about this model, indicating that the European situation might be slightly more complex. Our results from the Tajima's and Fu and Li's tests suggest that the *cpo* locus is under directional selection, and the non-significant results obtained for an upstream region showed that this phenomenon is peculiar to the *cpo* locus, rather than involving the whole right arm of the third chromosome. The HKA test returned non-significant results, but it cannot be ruled out that this occurred because we were unable to include indel polymorphism in our analysis. However the attempt to amplify the *cpo* isoform RH, the only one expressing SNP I462K, was unsuccessful at any of the temperatures/time points analysed. This result raised some doubts on the accuracy of *cpo* structure annotation, which was recently updated to exclude this isoform. Alternatively, *cpo*-RH might be expressed in conditions which were not tested in our experiment (different photoperiods and/or temperatures), only during development, or only in a small subset of neurons, thus making its amplification difficult.

Furthermore a high level of interspecies variability already reported for the 3' region of *cpo* exon 5 (Kankare *et al.*, 2011) was confirmed at SNP I462K, when *D. melanogaster* sequences were compared to sequences obtained from *D. simulans* flies collected in Africa. A high level of variability could be an indication of low structural and/or functional constraints at this particular residue. From a phenotypic perspective SNP I462K did not seem to influence the level of diapause in flies dissected after 12 nor 28 days, again in line with the results obtained

by Lee and co-workers in Australian natural flies (Lee *et al.*, 2011). Nevertheless replicate lines need to be created and analysed in order to confirm the results obtained in the phenotypic association study.

However a latitudinal cline was found in SNP A356V, independent on the *In(3R)P* chromosomal inversion, which was confirmed to be present at very low frequencies in the latitudinal range considered here. The correlation between allele frequency and latitude is significant, even though not as strong as the one found in American flies (Schmidt *et al.*, 2008). Our behavioural experiment cannot rule out the contribution of SNPs in linkage disequilibrium with SNP A356V, but it shows that haplotypes characterised by cpo^{356Val} have a higher diapause levels at both photoperiods and time points considered.

Taken as a whole the results described in this chapter confirm that *cpo* is involved in regulating the diapause phenotype in *D. melanogaster* flies but raise some doubts as to which SNP is responsible for such regulation. The results can be summarised as follows:

- *cpo* SNP A356V shows a significant latitudinal cline in Europe, although it is more dispersed than the American one. The derived allele increases in frequency with latitude. This cline is not driven by the linkage disequilibrium with the *In(3R)P* chromosomal inversion which, as expected, is present at very low frequencies in our dataset.

- *cpo* SNP I462K does not vary significantly with latitude in Europe, unlike American populations.

- The linkage disequilibrium between the two SNPs is not as strong in Europe as it is in America.

- Isoform *cpo*-RH, the only one containing SNP I462K, does not seem to be expressed, at least in the conditions here considered. Furthermore considerable interspecies variability is present at this site.

- *cpo* seems to be under directional selection

- SNP A356V (or some other SNPs in linkage disequilibrium with it) influences diapause levels. Haplotypes characterised by *cpo*^{356Val} show a significantly higher diapause incidence. In contrast, SNP I462K does not seem to influence diapause levels, consistent with the results obtained by Lee and colleagues in Australian populations.

5. INSULIN-LIKE RECEPTOR

5.1 INTRODUCTION

An inverse correlation between reproductive success and life span is present across different taxa (Partridge *et al.*, 2005). The causes of this correlation are still unclear: reproduction itself might damage the body thus affecting lifespan, or reproduction might compete in resources with somatic maintenance processes. Either way reproduction speeds up the aging process. The link between lifespan, reproduction and genetic factors was first suggested by a study by Friedman and Johnson (Friedman & Johnson, 1988). Both environmental and genetic factors influence this trade off: dietary restriction considerably extends lifespan and decreases (or abolishes) fecundity in rodents, *C. elegans* and *D. melanogaster* (Partridge *et al.*, 2005). In *C. elegans* lifespan extension segregated with a single gene, *age-1*, which was later found to be a phosphatidylinositol-3-kinase involved in the insulin signaling pathway (*Dp110* in *D. melanogaster*) (Morris *et al.*, 1996).

The Insulin signaling pathway is very well conserved across taxa and mutations at different levels of the signaling cascade have all been shown to lead to longer lifespan, compromised reproductive success and increased stress tolerance in mice, *C. elegans* and flies (Giannakou & Partridge, 2007; Partridge & Gems, 2002; Kenyon, 2005).

Figure 5.1 shows the Insulin-like signaling cascade in *D. melanogaster*. Mutations which have been shown to increase life span include those that lead to ablation of the cells producing Insulin-like peptides, ILPs (Broughton *et al.*, 2005), mutations in the Insulin-like Receptor, INR (Tatar *et al.*, 2001b), mutations in CHICO (Tu *et al.*, 2002a) and overexpression of the transcriptor factor FOXO (Hwangbo *et al.*, 2004). Not only do these mutations significantly increase life span, most of them also compromise growth and fecundity and the mutants are characterised by a higher resistance to acute oxidative stress (Giannakou & Partridge, 2007). A few mutants though have an uncompromised body size, fecundity and/or resistance to stress, showing that lifespan and the other life history traits can be uncoupled (Giannakou & Partridge, 2007). Female flies carrying a mutation in the *InR* gene emerge with immature ovaries, and in adults the egg chambers remain previtellogenic (Tatar *et al.*, 2001b). This, and the other phenotypes conferred by mutations in this pathway, show remarkable similarities with the diapause response.
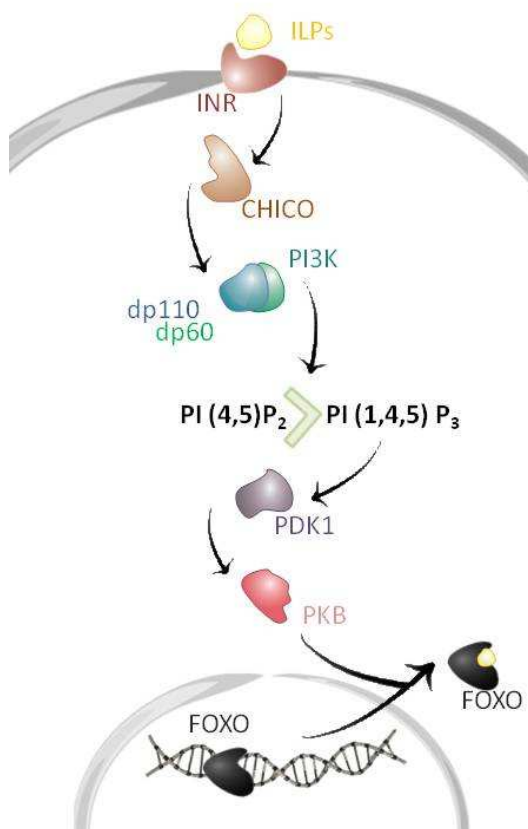

Figure 5.1
Insulin-like signalling pathway in *D. melanogaster*. Seven genomic sequences encode Insulin-like peptides in flies (ILPs). They bind to the Insulin-like Receptor (INR), which transduces the signal to its molecular partner CHICO. CHICO activates the PI3-kinase (which consists of two subunits, a catalytic and a regulatory one, dp110 and dp60 respectively), which converts phosphatidylinositol (4,5)-biphosphate to phosphatidylinositol (1,4,5)-triphosphate. This in turns activates a series of kinases which eventually phosphorilate the transcription factor FOXO, leading to its inactivation and translocation to the cytoplasm.

Several genes in this cascade have been studied in order to find potential clines in allele frequency which might contribute to the variability in diapause. Williams and colleagues (Williams *et al.*, 2006) identified *Dp110* when performing a segregation analysis aimed at identifying genetic difference between two fly strains characterised by different levels of diapause incidence. They showed that overexpression of the PI3K catalytic subunit in the fly brain leads to a decrease in diapause levels, and they identified 20 nucleotide differences between the two strains, in the genomic region including *Dp110*. Nevertheless none of these polymorphisms leads to an amino acid substitution in the Dp110 subunit. It has been suggested that differences in diapause levels are caused by different *Dp110* expression levels, regulated by one or a combination of the polymorphisms found in the *Dp110* non-coding region.

More recently Paaby and co-workers focused their attention on *InR* and its ligand *chico* (Paaby *et al.*, 2010). As for *chico*, no signature of selection was identified when the McDonald-Kreitman, Tajima's and Fu and Li's tests were performed. Furthermore several polymorphisms were identified when the gene was sequenced from 27 fly strains collected in North America: 25 were synonymous SNPs, 12 lead to an amino acid substitution and one was an indel polymorphism. Nevertheless none of them shows a significant latitudinal trend in allele frequency (Paaby *et al.*, 2010). A more intriguing pattern emerged when *InR* was analysed. Analysis of 41 *InR* sequences showed an excess of fixed replacement changes between *D. melanogaster* and *D. simulans*, suggesting that adaptive evolution had occurred at this locus. However the Tajima's and the Fu and Li's test statistics did not reach significance when applied to these *D. melanogaster* sequences. A considerable number of polymorphisms was found in the 41 *InR* sequences obtained from North American lines: 88 synonymous SNPs, 15 non-synonymous and 15 indels. 13 of these (of which 11 are synonymous or in introns) show significant latitudinal trends, but when the analysis was extended to a bigger sample size, including Australian fly lines, only one polymorphism was characterised by the same

trend in the two different continents. This involves an indel polymorphism in the first exon of the gene, that interrupts a series or Glutamine and Histidine repeats. The six alleles present in both continents are shown in Figure 5.2, together with the sequences in 11 other *Drosophila* species.


```
D. melanogaster    RRRRQHQQQHHH-YQH----------HHQQHLQRQQANVSYT-KK    236
                   RRRRQHQQQHHH-YQH------HQHHHQQHLQRQQANVSYT-KF    245
                   RRRRQHQQQHHHHYQHH------QHHHQQHHQRQQANVSYT-KF    248
                   RRRRQHQQQHHHHYQHHH-----QQHHQQHHQRQQANVSYT-KF    251
                   RRRRQHQQQHHH-YQHHQH---HQHHHQQHLQRQQANVSYT-KF    254
                   RRRRQHQQQHHH-YQHHQHHQHHQHHHQQHLQRQQANVSYT-KF    263

D. simulans        RRRRQHQQQQHHHHQHHP-----QHHHQQ-----QAHVSYT-KF
D. sechellia       RRRRQHQQQHHHNHQHHP-----QHHHQQHHQRQQAHVSYT-KF
D. yakuba          RRRRQHQHKQQ--------------------QQQQVHVNYT-KF
D. erecta          RRRRQQQQQQQQQ------------------YHQQQAHVNYT-KF
D. ananassae       RRRR--EQRQQ--------------------QQQ--DFNFT-KF
D. pseudoobscura   RRRRQQQQQRQE------------------PHQKQKFVNYT-KF
D. persimilis      RRRRQQQQQRQE------------------PHQKQKFVNYT-KF
D. willistoni      SHHHQHQQQQQH--------------QHRHHHRRNQICIN-DI
D. virilis         QQQRQHPPHQQT--------------VQHCRLQQQHAAPA-TC
D. mojavensis      QQQQQQQQYKY-------------------QQEQRASST-TF
D. grimshawi       QQQTPKQQQRES-------------LRQFMKDKNPRRAYTYTF
```
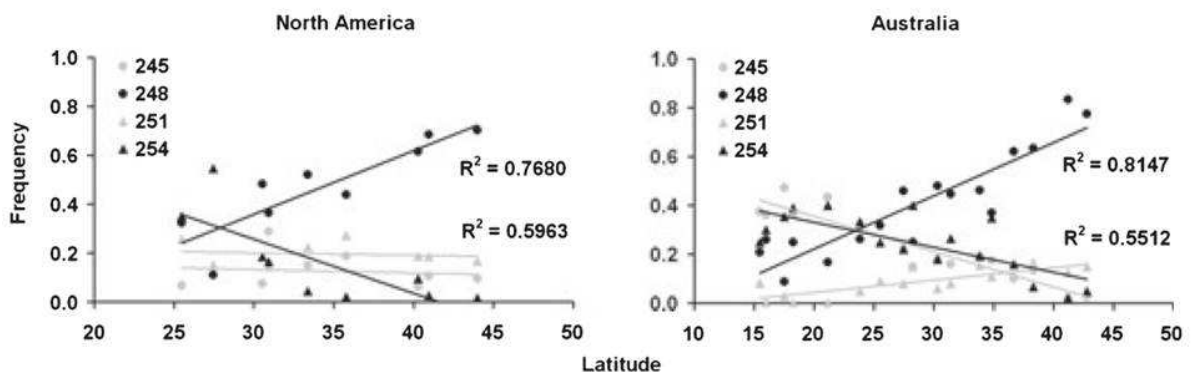
Figure 5.2
Top Panel: six indel alleles present in the first exon of *InR*. Four other variants were identified but at very low frequencies. The six alleles shown here were named after the assumed length in bp of the PCR fragment generated upon their amplification. The two most common alleles, showing latitudinal clines in America and Australia, are shown in black. Mid panel: corresponding region including the polymorphism in 11 other *Drosophila* species. Bottom panel: frequency of the six *InR* alleles plotted *versus* latitude in North America (left) and Australia (right). Figure from Paaby *et al*., 2010.

Two alleles at this site present latitudinal clines in frequency in both North America and Australia, namely the alleles *248* and *254*. The direction and steepness of the cline are very similar in the two continents. The contribution of the chromosomal inversion *In(3R)Payne* (whose frequency is clinal in both North America and Australia) (Knibb, 1982) in driving the cline in *InR* was rejected, given that a significant cline persists for these two alleles in North America when only standard chromosomes are considered, and in Australia when only inverted chromosomes are included in the analysis (Paaby *et al.*, 2010).

From a phenotypic perspective, the two alleles were shown to confer phenotypes consistent with their geographical distribution: lines homozygous for the allele *248*, present at high frequencies at high latitudes, were shown to recover faster from oxidative stress and chill coma, but to lay fewer eggs than flies homozygous for the allele *254* (Paaby *et al.*, 2010). This result is consistent with a *scenario* where in *248* flies, the insulin signaling pathway is downregulated, and this is of evolutionary advantage at northern latitudes.

In this chapter I analysed the frequency of these *InR* alleles in our set of European *D. melanogaster* lines. While working on this project, following the genotyping protocol used by Paaby and coworkers, a discrepancy was identified in the names attributed to the alleles. Because of technical reasons, our amplicon lengths and the authors' do not match, and none of those correspond to the actual fragment size. In order not to generate confusion to the reader, Table 5.1 shows the reported and the real fragment sizes, together with the phenotypes conferred by the two alleles. Hereafter only the "real sizes" will be used in the text.

REPORTED LENGTH	REAL LENGTH	RESISTANCE TO STRESS	EGGS LAYED	F (SOUTH)	F (NORTH)
248	253	High	Few	~0.1-0.2	~0.7-0.8
254	259	Low	Many	~0.4-0.5	~0

Table 5.1

Details regarding the two clinal alleles in *InR*. 'Reported length' refers to the fragment length in bp reported in Paaby *et al.*, 2010. 'Real length' refers to the actual fragment length, and it is the measure that will be used hereafter. F indicates the frequency of the two alleles in North American and Australian populations.

5.2 MATERIALS AND METHODS

5.2.1 FLY LINES

As for the Spanish populations used in this chapter, methods of collections and husbandry are reported in Chapter 2. Where possible, males collected directly from the wild and kept in EtOH 100% at -20˚C were used. In some cases, in order to increase the sample size, one male per isofemale line was also included in the analysis.

All the other flies used in this chapter were already available in the laboratory, details regarding location and date of collection are listed in Chapter 2.

5.2.2 *InR* GENOTYPING

InR was genotyped following the protocol used by Paaby and colleagues. The region under study was amplified using the 6-FAM-fluorescent-tagged forward primer InR-F* 5'-*CAATATCTTTAGCAACTGTCAC*-3' and the reverse oligo InR-R 5'-*TTTAGGGCTTAAACTCAGTC*-3' (nucleotide coordinates 17405775..54 and 17405520..39 respectively). The following thermal cycle was used for the amplification: 40 cycles with 92°C for 35 sec, 52°C for 30 sec, 72°C for 30 sec. After having used part of the samples to verify the success of the amplification in an

agarose gel, the samples were subjected to A-tailing and analysed by the PNACL (Protein Nucleic Acid Chemistry Laboratory) facilities of the University of Leicester, as described in Paragraph 11 of Chapter 2.

5.3 RESULTS

5.3.1 *InR* INDEL IN EUROPEAN POPULATIONS

12 European and one African fly line were genotyped for the indel polymorphism in *InR*. To allow discrimination among alleles which differ by as little as 3 nucleotides, the region under interest was amplified with a couple of primers of which one was fluorescently tagged. The samples were then subjected to A-tailing (Figure 5.3), to guarantee that all the amplified molecules were characterised by the overhanging *A* nucleotide added by the Taq polymerase at the 3' of the amplicons. This step allows the assumption that differences in length are due to indels in the sequence.

The samples were then run in an Applied Biosystems 3730 sequencer and finally analysed with the Peak Scanner Software v1.0. Figure 5.4 shows an example of outcome. Some of the samples were also sequenced to confirm the allele sequence and its length. Nevertheless it is not guaranteed that all the variants with the same length are characterised by the same nucleotide sequence.

Table 5.2 shows the details regarding the fly populations used in this study and the frequency of each allele in the dataset. The yellow shaded rows indicate the two alleles which were found to be clinal in North America and Australia. The last column shows the result of the Hardy-Weinberg test, which was performed with the statistical program R (http://www.r-

project.org) and the package "Genetics". All the populations analysed are in Hardy-Weinberg Equilibrium.
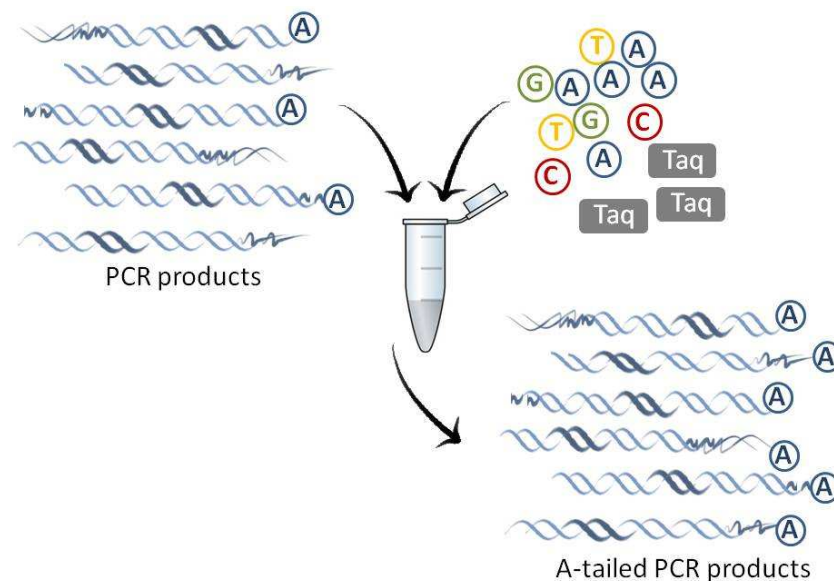

Figure 5.3
A-tailing strategy. The PCR products were incubated at 72°C for 45 minutes, with Taq Polymerase and PCR buffer containing dNTPs. At the end of the reaction all the DNA molecules are characterised by an overhanging A at their 3' end.

As shown in Table 5.2, 10 alleles were identified in the European dataset. Nine of them had previously been reported by Paaby and colleagues. Alleles *262* is unique to the European dataset. Allele *280* (referred to as *275* in Paaby *et al*., 2010), which is present at low frequencies in Australian populations, was not found in our dataset

Figure 5.5 A and B shows how the most rare and most frequent alleles vary with latitude of collection. None of the rare alleles shows a significant correlation with latitude (Figure 5.5 A). Surprisingly not even the two variants identified by Paaby and colleagues vary significantly with latitude in the European dataset (Figure 5.5 B). The only allele to show a significant latitudinal cline in frequency is allele *256* (R^2=0.452; p=0.02), which also appears to be present at slightly higher frequencies in Europe than in North America and Australia (Figure 5.2, where the variant is referred to as *251*). Nevertheless this correlation ceases to be significant if Bonferroni correction for multiple testing is applied.
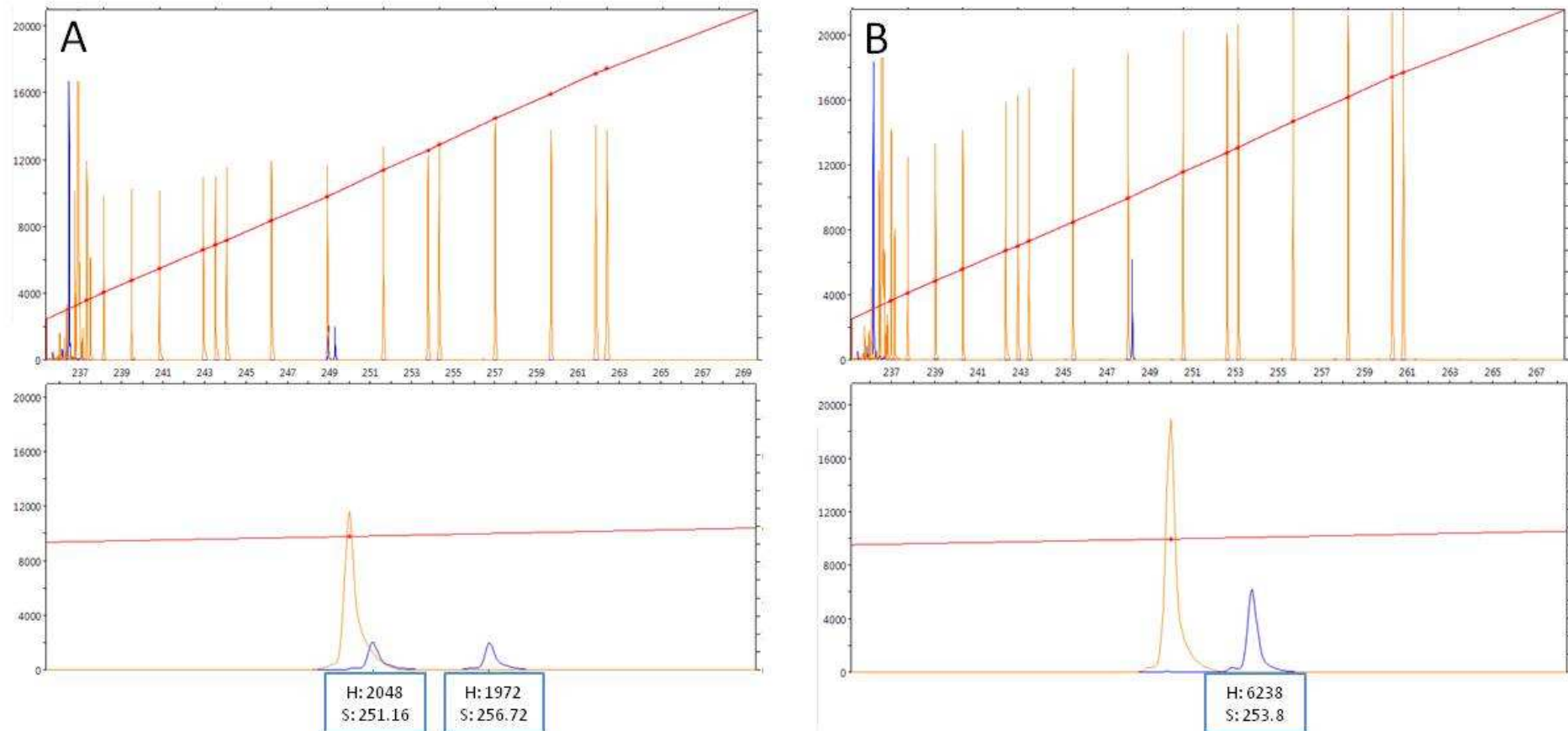
Figure 5.4

Example of output from the Peak Scanner Software v1.0. Two samples are shown here, A and B. The top panels show an overview of the two runs. X-axis: length in bp. Y-axis: height of the peaks, corresponding to the intensity of the signal detected. The orange peaks correspond to the marker bands, the blue ones designate the amplified fragment, which was detected thanks to the fluorophore with which the forward primer was tagged. The bottom panels show a magnification of the region including the amplified fragment peak. H: height. S: size. A) The sample belongs to a heterozygous individual. The two alleles are reported to be 251 and 257 bp long. B) Homozygous fly. The amplified fragment appears to be 254 bp long. The fluorophore attached to the forward primers influences the fragment migration through the capillary, and as a consequence the sizes returned by the software do not correspond to the actual fragment length, which is 2 bp longer. The genotypes of the samples shown here are therefore A: heterozygote *253/259* and B: homozygote *256/256*.

LINE CODE	LOCATION	COUNTRY	N	LAT	LONG	ALT	238	241	244	247	250	253	256	259	262	268	χ²	Df
LINES DETAILS							**FREQUENCIES**										χ²	Df
SP-22	Nijar	Spain	42	**36.97**	-2.21	345	0	0.1	0	0	0.17	**0.29**	0.02	**0.36**	0	0.07	17.8	21
SP-28	Jumilla	Spain	54	**38.48**	-1.32	520	0	0.04	0.02	0.02	0.06	**0.63**	0.09	**0.09**	0.06	0	38.9	36
SP-35	Fontanares	Spain	54	**38.78**	-0.79	630	0	0.02	0.02	0	0.04	**0.63**	0.17	**0.13**	0	0	6.6	21
SP-41;42	Libros	Spain	48	**40.16**	-1.23	770	0	0.02	0	0	0.06	**0.73**	0.17	**0.02**	0	0	17.4	15
SAL	Salice	Italy	58	**40.38**	17.38	48	0	0.03	0.02	0	0	**0.69**	0.26	**0**	0	0	29.4	10
BIT	Bitetto	Italy	58	**41.02**	16.75	149	0	0	0.03	0	0.02	**0.71**	0.22	**0.02**	0	0	3.6	15
CAV	Cavarzere	Italy	90	**45.13**	12.08	3	0	0	0	0	0.01	**0.70**	0.20	**0.07**	0	0.02	48.3	15
HU	Houten	Holland	86	**52.03**	5.17	3	0	0	0	0	0.07	**0.66**	0.20	**0.07**	0	0	18.5	10
MAR	Market Harborough	England	96	**52.48**	-0.92	82	0	0	0.03	0	0.05	**0.72**	0.17	**0.03**	0	0	8.8	15
KIL	Kilworth	England	74	**52.53**	0.98	94	0	0	0	0	0.01	**0.66**	0.27	**0**	0.05	0	1.1	15
HФJ	Hфjbjerg	Denmark	90	**56.11**	10.21	n.a.	0.21	0	0	0	0	**0.37**	0.37	**0**	0	0.06	23.1	15
KOR	Korpilahti	Finland	90	**62.02**	25.55	104	0	0.02	0	0	0.04	**0.40**	0.28	**0.26**	0	0	4.3	21
KN	Nyahururu	Kenya	56	**0.03**	36.37	2360	0	0.02	0	0	0.48	**0.38**	0.04	**0.02**	0.02	0.05	46.9	28

Table 5.2

Details of the flies used in the *InR* SNP indel study and frequency of the 10 alleles identified. N: number of alleles analysed. Lat: latitude in degrees North. Long: longitude in degrees (Negative values: West; Positive values: East). Alt: altitude in meters above sea level. The allele names refer to the actual fragment length. The yellow shaded coloumns indicate the alleles found to be clinal in North America and Australia (Paaby *et al.*, 2010). The second last column shows the results of the Hardy Weinberg test. Df: Degrees of freedom. None of the values is significant.
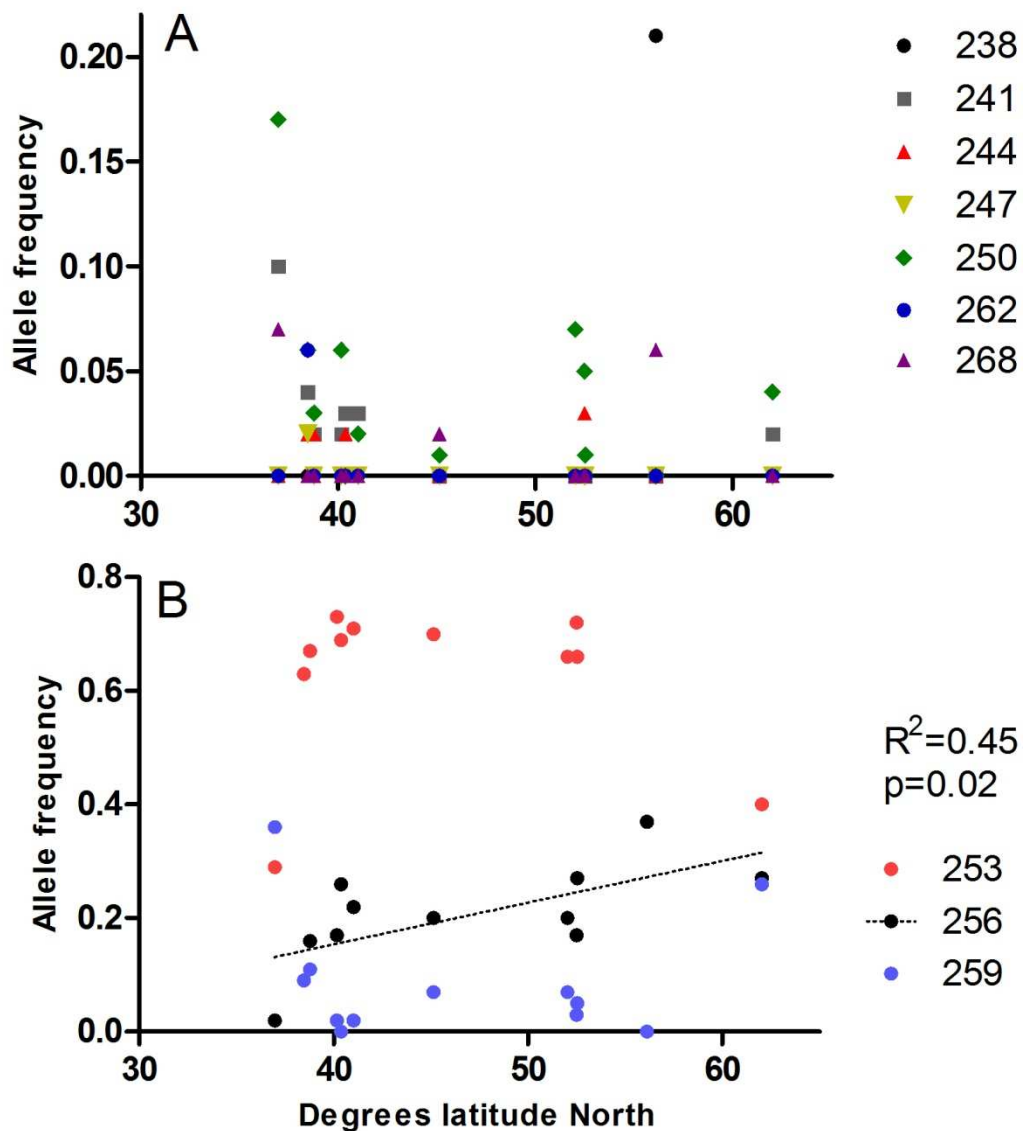
Figure 5.5

Geographical distribution of the *InR* indel alleles in *D. melanogaster* European populations. A) Frequency of the most rare alleles plotted against latitude of collection. None of them changes significantly with latitude. B) Frequency of the most common alleles identified in this study, plotted against latitude. Only the frequency of allele *256* changes significantly with latitude (R^2=0.452; p=0.02). This correlation ceases to be significant if Bonferroni correction is applied. The data regarding the African population (see Table 5.2), were not included in these plots.

5.3.2 *InR* EXPECTED HETEROZYGOSITY

Heterozygosity (H) is a way to measure the level of genetic variability in a population.

It depends on the number of alleles present at a specific locus, and on their frequency. The

value of H ranges between 0 and nearly 1. Zero indicates absence of genetic variability (no

heterozygosity, monomorphic site), whereas values close to one are obtained when many

alleles are present, and all of them are present at the same frequency (1 is only reached when a locus is characterised by infinite alleles).

H is calculated with the following formula:

$$H = 1 - \sum_{i=1}^{k} p_i^2$$

where p_i represents the frequency of the ith out of k alleles.

The results obtained for our populations are shown in Table 5.3. In Figure 5.6 the data is plotted over latitude of collection. The dotted line shows the expected heterozygosity level calculated for the African population analysed.

LINES DETAILS								
LINE CODE	LOCATION	COUNTRY	N	LAT	LONG	ALT	A	H
SP-22	Nijar	Spain	42	**36.97**	-2.21	345	6	0.77
SP-28	Jumilla	Spain	54	**38.48**	-1.32	520	8	0.59
SP-35	Fontanares	Spain	54	**38.78**	-0.79	630	6	0.57
SP-41;42	Libros	Spain	48	**40.16**	-1.23	770	5	0.44
SAL	Salice	Italy	58	**40.38**	17.38	48	4	0.46
BIT	Bitetto	Italy	58	**41.02**	16.75	149	5	0.46
CAV	Cavarzere	Italy	90	**45.13**	12.08	3	5	0.47
HU	Houten	Holland	86	**52.03**	5.17	3	4	0.52
MAR	Market Harborough	England	96	**52.48**	-0.92	82	5	0.46
KIL	Kilworth	England	74	**52.53**	0.98	94	4	0.49
HØJ	Højbjerg	Denmark	90	**56.11**	10.21	n.a.	4	0.69
KOR	Korpilahti	Finland	90	**62.02**	25.55	104	5	0.71
KN	Nyahururu	Kenya	56	**0.03**	36.37	2360	7	0.63

Table 5.3
Number of alleles (A) and Heterozygosity (H) in the populations used in the *InR* study. N: number of alleles analysed. Lat: latitude in degrees North. Long: longitude in degrees (Negative values: West; Positive values: East). Alt: altitude in meters above sea level.
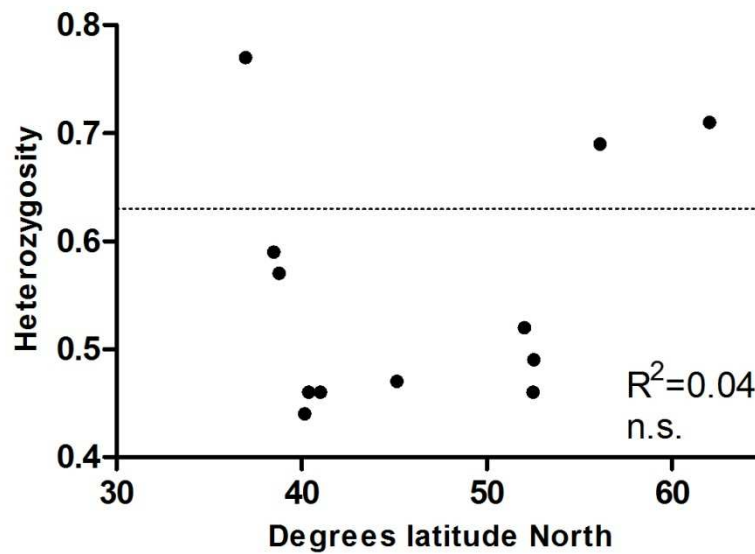
Table 5.3
Heterozygosity (H) of the populations used in the *InR* study. The dotted line shows the Heterozygosity level found in the African population.

Surprisingly the African population does not show the highest level of heterozygosity. Furthermore all the alleles present in Africa are also present in Europe, in Australia and North America. This suggests that the high level of variability present at this locus has ancient origins, rather than having appeared after *D. melanogaster* colonised the Old World. Alternatively, the low level of heterozygosity detected in the Kenya flies could simply be due to the high level of non-African genetic admixture identified in this population (*Drosophila* Population Genomics Project, DPGP2-African Survey, available at http://www.dpgp.org).

Interestingly, the heterozygosity in European populations seems to increase at the extremes of our latitudinal range (latitude below 40° North and above 55° North), settling to values around 0.45-0.50 at all intermediate latitudes.

5.4 DISCUSSION

Mutations affecting the Insulin-like pathway have been shown to increase life span and reduce reproduction not only in *D. melanogaster*, but also in *C. elegans* and rodents (Partridge *et al.*, 2005; Kenyon, 2005). Different genes in this pathway have been studied in flies, such as the PI3K catalytic subunit *Dp110* (Williams *et al.*, 2006), the *Insulin-like Receptor* and its molecular partner *chico* (Paaby *et al.*, 2010), in an attempt to identify variations which could (at least partly) account for differences in diapause levels in *D. melanogaster* natural populations collected at different latitudes (Paaby *et al.*, 2010; Schmidt *et al.*, 2005a). Out of these three genes analysed, only *InR* seemed to be carrying the signature of natural selection. Furthermore an indel polymorphism was identified in the first exon of the gene, characterised by two alleles which not only show latitudinal clines in frequency in Australia and North America, but also influence life history traits linked to diapause, in the direction expected given their geographical distribution (Paaby *et al.*, 2010).

In this chapter we specifically looked at this polymorphic site in natural populations collected in Europe, and in one collected in Kenya. 10 length variants were identified, only one of which (*262*) is specific to Europe and was not identified in the previous studies. Such a result comes as a surprise, since the bottle-neck faced by flies upon their colonisation of North America and Australia was expected to have affected the variability at this locus. An explanation might be that the repetitive nucleotide sequence at this locus increases the chances of mispairing between the template DNA and the strand being synthesised during DNA replication (Streisinger *et al.*, 1966; Rosato *et al.*, 1997a), thus generating new length alleles at a much higher rate than expected. In such a *scenario* the number of alleles could have been drastically reduced when flies first colonised the New World and Australia, but would have increased rapidly due to the repetitive nature of the sequence. It would be possible to discriminate between these two hypotheses by sequencing a number of

haplotypes from Africa, Europe, America and Australia, in order to determine if a specific allele appears on the same haplotype in all the continents. Since in this study alleles were scored based on their size rather than on their sequence, the possibility that the Australian and American allele originated independently cannot be ruled out. Furthermore, all the alleles identified in the African population are also present in Europe, suggesting that the high variability in *InR* has ancient origins, and appeared before *D. melanogaster* colonised Europe.

Interestingly, the two alleles previously reported to be clinal in North America and Australia, do not show any significant latitudinal distribution in Europe, once again demonstrating a *scenario* in which the situation in Europe unexpectedly differs from the one in the other two continents. *D. melanogaster* flies have been in Europe for the past 10-15 thousand years, whereas they probably colonised North America and Australia only 2-300 years ago (David & Capy, 1988). It seems therefore unlikely that they successfully adapted to the environmental conditions there, but not yet in the Old World. Two theories could explain the differences in the distribution of allele frequencies among these continents. The clines observed in North America and Australia could have arisen merely because of demographic processes and genetic drift rather than being driven by natural selection. Nevertheless the similarity in allele frequency and cline steepness between the two continents, together with the phenotypes conferred by the two alleles, severely challenge this hypothesis. Alternatively, the climatic conditions in Europe might not provide a selective pressure as strong as in the other two continents, for a cline to generate. The term "selective pressure" here includes all those environmental cues which change with latitude and might be perceived by flies to anticipate the change of the seasons. Whereas photoperiod changes predictably with latitude, such a correlation is not obvious for other cues such as temperature. Using latitude as a proxy for these environmental cues might therefore be too much of an over-simplification when the selective strength is not photoperiod alone. A similar situation was described by David and coworkers (David *et al.*, 1989) who analysed the frequency of the *Adh-F* allele in the *Alcohol*

dehydrogenase (*Adh*) gene in European, African, Australian and American populations. The overall frequency of the *Adh-F* allele increases significantly with latitude, but in a non-linear fashion, highlighting three latitudinal ranges probably corresponding to three distinct climatic conditions. Between 30° and 40° latitude, the steepness of the cline increases dramatically underlining a high variability in *Adh-F* frequency in populations from these latitudes. Although flies collected in this range of latitudes in Europe, Australia and along the American West coast look alike in terms of allele frequency, this differs greatly between populations from northern Europe and populations collected at the same latitudes along the American East coast. Furthermore in populations from southern Europe (Spain and Greece) the *Adh-F* frequency was found to be unstable, and to vary significantly within short distances or throughout the year. Temperature is considered one of the potential factors affecting the distribution of *Adh-F*. The patchy southern European environment, contrary to the environmental continuity between subtropical and cold conditions found along the eastern American coast is believed to be responsible for the differences observed (David *et al.*, 1989).

The heterozygosity was calculated in our populations, as a measure of genetic variability at this locus. Surprisingly the highest heterozygosity was detected in populations at the extremities of the latitudinal range considered here (below 40° and above 55° latitude). Most of these populations were characterised by values even higher than the one found in the African population (>0.63), maybe because of the high level of non-African genetic admixture identified in this population (~40%, *Drosophila* Population Genomics Project, DPGP2-African Survey, available at http://www.dpgp.org). In 1974 Lewontin proposed a model, later reported by Knibb (Knibb *et al.*, 1981) to explain the significant decrease in the frequencies of chromosomal inversions as the distance from the Equator grows larger. Marginal habitas are characterised by temporally unstable environmental conditions, therefore selection pressure is strong but temporally variable. The absence of inversions allows relatively free recombination which facilitates rapid selection responses. A similar mechanism could be at

work here. Marginal environments are affected by more extreme conditions, in terms of photoperiod (at northern latitudes) and temperature changes throughout the year (at southern latitudes), and having a vast pool of alleles in the population might increase the speed of adaptation, hence the chances of survival.

Overall, the results presented in this chapter can be summarised as follows:

- The same indel alleles at the *InR* locus were identified in Europe, North America, Australia and in one African *D. melanogaster* population, except for one allele which was found to be present only in the European dataset, and one only present in Australian populations. This suggests that:
  - the high variability present at this site has an ancient origin.
  - either the colonisation of Australia and North America did not affect the *InR* indel allele pool, or alleles lost during the bottle-neck reappeared quickly because of the mutagenic nature of this repetitive sequence.
- The alleles reported to vary significantly with latitude in North America and Australia (*253* and *259*) do not show any cline in Europe.
- Allele *256* is present at slightly higher frequencies in Europe and increases significantly with latitude.
- Heterozygosity is at its highest in populations collected at the margins of our latitudinal range. High variability might be actively maintained here by natural selection in order to facilitate rapid adaptation to changing environmental conditions.

6. DIAPAUSE

6.1 INTRODUCTION

In the previous result chapters I compared the geographical distribution of diapause-regulating alleles in Europe to the corresponding ones in the USA and, when available, in Australia. But what about the phenotype itself? How does the diapause incidence of flies collected along the US and Australian East coast compare to the phenotype of European flies? Do the discrepancies found at a genetic level match differences in diapause levels?

In natural flies from North America, diapause incidence was first studied by Schimdt and coworkers (Schmidt *et al.*, 2005a). Flies from different locations were kept for four weeks in the diapause inducing conditions of 12°C and short photoperiods (LD 10:14). As shown in Figure 6.1A diapause incidence predictably increases with latitude of collection. Other than the absolute level of diapause detected in a fly line, another phenotype which is informative in terms of fly adaptation to environmental conditions is photoperiodism, namely the capability of a fly line to distinguish different photoperiods, and to behave accordingly. Two American fly lines (collected in Maine and Florida) were tested for photoperiodism and surprisingly they did not show any difference in diapause incidence when exposed for 25-28 days to two different photoperiods ("summer" LD 18:6 and "winter" LD 10:14), regardless of the temperature at which they were kept (Figure 6.1 B and C) (Emerson *et al.*, 2009b).
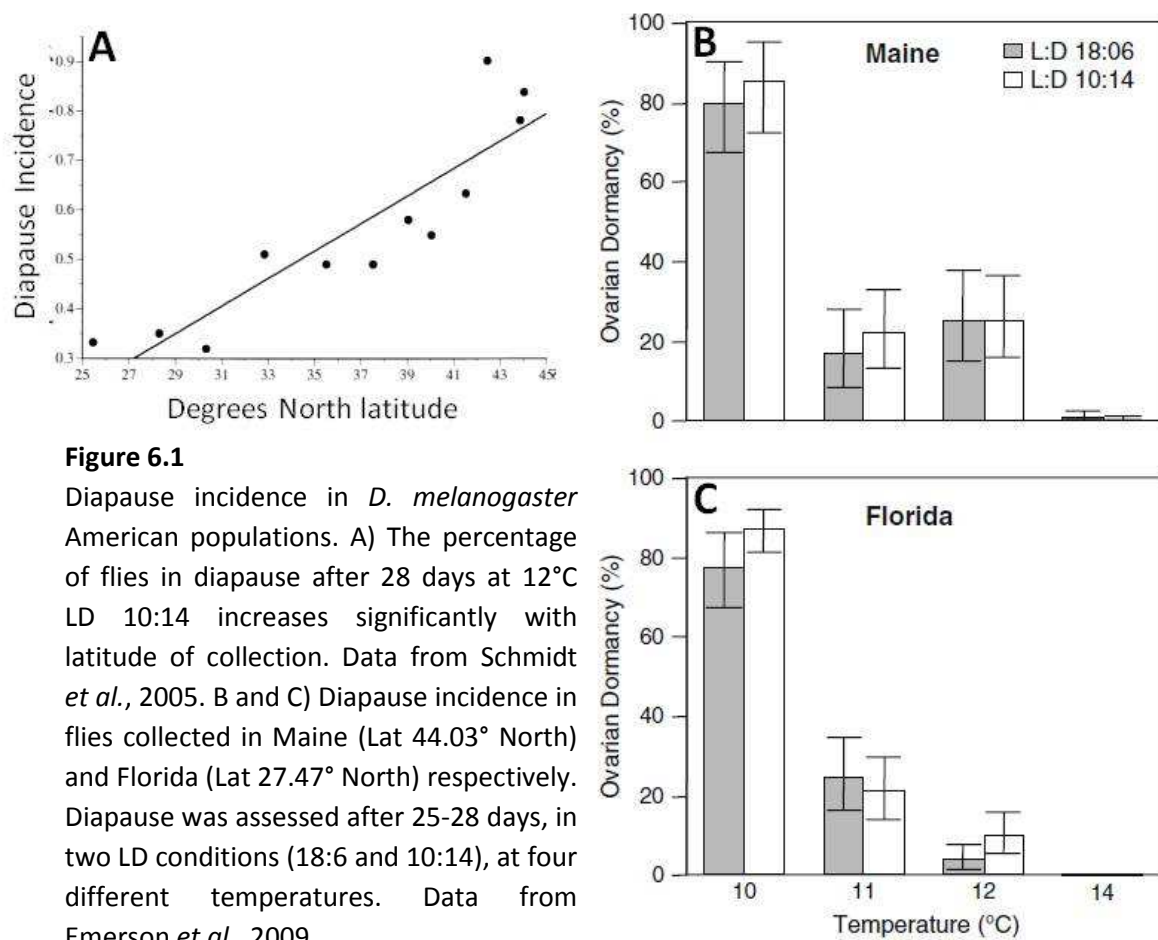
Figure 6.1

Diapause incidence in *D. melanogaster* American populations. A) The percentage of flies in diapause after 28 days at 12°C LD 10:14 increases significantly with latitude of collection. Data from Schmidt *et al.*, 2005. B and C) Diapause incidence in flies collected in Maine (Lat 44.03° North) and Florida (Lat 27.47° North) respectively. Diapause was assessed after 25-28 days, in two LD conditions (18:6 and 10:14), at four different temperatures. Data from Emerson *et al.*, 2009.

As for Australian populations, again a clear correlation between diapause incidence and latitude of collection was identified (Lee *et al.*, 2011). Flies were kept for 28 days at 12°C LD 10:14. Nevertheless, given the low level of diapause detected, the authors raised the usual cut off used to discriminate between diapausing and non-diapausing flies, from developmental stage 8 to stage 11. Interestingly the correlation is linear and resembles the American cline in terms of steepness between latitude 25 and 43° South. However, when populations originally closer to the Equator were included, diapause incidence was found to increase, revealing an interesting non-linear relationship between the two parameters (Figure 6.2).
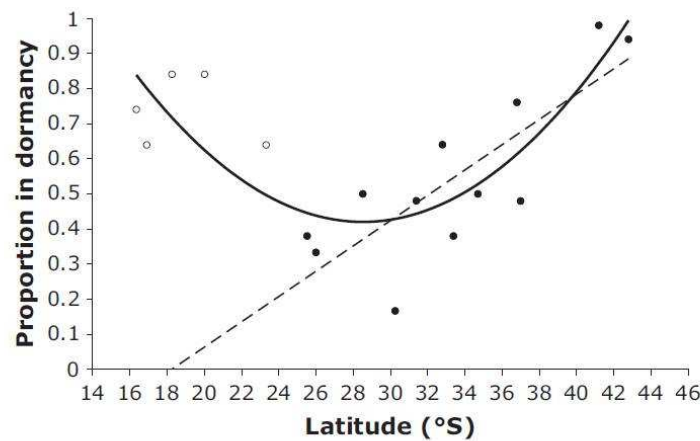
Figure 6.2
Diapause incidence in natural flies collected along the Australian eastern coast. The (dotted) linear regression line was fitted only through the northern dataset (latitudes 25-43° South, full dots). The (solid) quadratic regression line was fitted though all the data points, including the one from latitudes lower than 25°. Figure from Lee *et al.*, 2011.

Finally, in Europe the incidence of diapause was assessed in 6 populations collected at different latitudes. The photoperiodic curve was determined for each population by keeping flies at 12°C at 6 different photoperiods (from LD 8:16 to LD 18:6). After 12 days in these conditions diapause levels were assessed. As shown in figure 6.3 A diapause levels fall as the day gets longer (Pegoraro, Zonato *et al.*, manuscript in preparation). Figure 6.3 B shows the incidence of diapause at photoperiod LD 10:14, plotted against latitude. Only at this particular photoperiod is the correlation significant (R^2=0.80, p=0.02). When the experiment was performed at 28 days (again, at photoperiod LD 10:14), the correlation failed to be significant (green data points in Figure 6.3 B, Pegoraro, Zonato *et al.*, manuscript in preparation).

Table 6.1 sums up all the phenotypes described above, and highlights a situation where direct comparison of data from the three continents is rendered difficult by different groups not using exactly the same experimental approach. Emerson and colleagues have underlined the differences in diapause levels and photoperiodism between American and European lines (Emerson *et al.*, 2009b) but one wonders if these differences are due to the different duration of diapause-inducing conditions that the flies were exposed to in the different studies.
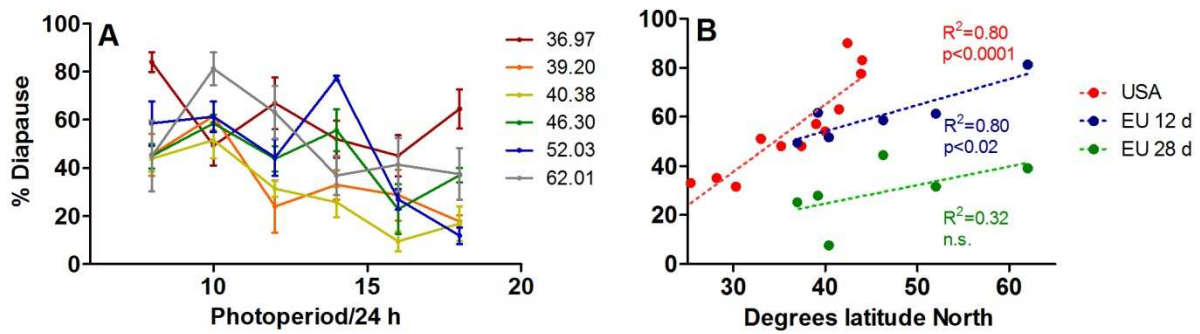
Figure 6.3

Diapause incidence in European *D. melanogaster* natural populations. A) Photoperiodic curves of 6 populations. The legend shows the latitudes of collection in °N. B) Mean diapause at LD 14:10, when the experiment was performed for 12 (blue) or 28 (green) days. Data regarding the 12 days experiment are taken from the plot in A. Red dots represent the values in North American lines (redrawn from Schmidt *et al.*, 2005). Data from Pegoraro, Zonato *et al.*, manuscript in preparation.

CONTINENT	STUDY	DAYS	CUT OFF	CLINE	PHOTOPERIODISM
USA	Schmidt *et al.*, 2005 Emerson *et al.*, 2009	28	<8	YES	NO
Australia	Lee *et al.*, 2011	25-28	<11	YES*	?
Europe	Pegoraro, Zonato *et al.*, manuscript in prep.	12 28	<8	YES** NO	YES ?

Table 6.1

Summary of diapause phenotypes in natural lines collected in 3 different continents. "Days" refers to how long the flies were kept in diapause conditions in the different studies considered here. "Cut off" refers to the developmental stage of the most mature oocyte used to define an individual as in diapause or not. * there is a cline in diapause when only the range 25-43°N is considered. Otherwise the correlation diapause levels/latitude is not linear. ** The cline in significant, but only at one particular photoperiod (LD 10:14). At all the other LDs considered the correlation fails to reach significance.

We therefore decided to analyse the diapause levels of European and American lines, exposing them to the same experimental conditions. We analysed the phenotype after exposing flies to diapause inducing conditions for 12 and 28 days. Furthermore, in order to detect their ability to react to different photoperiods, the experiment was performed at LD 8:16 and LD 16:8.

In the second part of this chapter the evolutionary age of this phenotype is addressed. *D. simulans* flies and *D. melanogaster* flies from Africa have been suggested not to show any diapause (Schmidt *et al.*, 2005a), and therefore this phenotype has been considered a recent adaptation to temperate environments. The fact that these observations were performed using the 28-days-approach, together with the lack of published results on these fly lines, led us to include these populations in our experiments.

6.2 MATERIALS AND METHODS

6.2.1 FLY LINES

Methods of fly husbandry are reported in Chapter 2. Fly lines from Spain were collected as described in 2.1.1, while flies from Treviso (Italy), Korpilahti (Finland), Rende (Italy, both the *D. melanogaster* and the *D. simulans* lines) were already available in the laboratory. The two American fly lines were kindly provided by P. Schmidt, whereas the two African strains (*D. melanogaster* and *D. simulans*), were given to us by C. Schlotterer. The details regarding the collection sites of all the lines used in this chapter are listed in Table 6.2.

			LINE DETAILS				
LINE	SPECIES	LOCATION	COUTRY	N	LAT	LONG	ALT
SP-22	*D. melanogaster*	Nijar	Spain	24	36.97	-2.21	345
SP-35	*D. melanogaster*	Fontanares	Spain	32	38.78	-0.79	630
SP-53	*D. melanogaster*	Manresa	Spain	20	41.73	1.82	278
MREN	*D. melanogaster*	Rende (Italy)	Italy	25	39.20	16.11	480
TRV	*D. melanogaster*	Treviso (Italy)	Italy	35	45.67	12.24	40
KOR	*D. melanogaster*	Korpilahti	Finland	20	62.01	25.33	120
FL	*D. melanogaster*	Homestead	Florida (USA)	12	25.47	-80.48	1
MAINE	*D. melanogaster*	Bowdoinham	Maine (USA)	50	44.01	-69.9	1
KN	*D. melanogaster*	Nyahururu	Kenya (Africa)	30	0.04	36.37	2360
SREN	*D. simulans*	Rende	Italy	21	39.20	16.11	480
simAFR	*D. simulans*	Uganda (5) Tanzania (3) Kenya (5) Malawi (2)	Africa	15	From -15.38 to 0.46	From 30.40 to 36.67	n.a.

Table 6.2

Details of the fly lines used in this chapter. N: number of isofemale lines used to create the populations analysed. Lat: latitude in degrees North. Long: longitude in degrees (Negative values: West; Positive values: East). Alt: altitude in meters above sea level. n.a.: not available. simAFR is a composite line, created pooling together 15 isofemale lines from 4 African locations.

6.2.2 DIAPAUSE EXPERIMENT

Details on how the experiments were carried out are described in paragraph 2.12. Briefly, an equal number of non-virgin females from each available isofemale line was pooled to create a population. This was expanded and female virgins (max 6 h post eclosion) were collected and put at 12°C in fluorescent light boxes, at the required photoperiod. Temperature was recorded by means of a temperature logger (Tinytag, purchased from Gemini Data Loggers) kept inside one of the two light boxes. After 12 or 28 days flies were dissected and

scored as to whether they were in diapause or not. 5-6 replicates per condition were performed, alongside with 2 "Dark-Dark" (DD) controls, which were maintained in the same conditions as the other samples, except they were wrapped in aluminum foil. These were used to make sure that potential phenotypical differences between flies kept in the two photoperiods were due to the different light-dark regimes themselves, other than to the 0.2°C differences in temperature caused by the two photoperiods (see Figure 2.2). A total of 10107 flies were dissected in these experiments.

6.3 RESULTS

6.3.1 DIAPAUSE IN SPANISH LINES

As described in the Introduction, recent work performed in our laboratory investigated the level of diapause incidence in fly lines collected in six different locations in Europe. Figure 6. 3 A shows the photoperiodic curves obtained as a result of this study, and highlights a population collected in the South of Spain as an outlier (SP-22, latitude 36.97° North, red data points). Given its latitude and the low frequency of *ls-tim* in this population (see Chapter 3) the level of diapause there was expected to be rather low, but surprisingly SP-22 is characterised by an extremely high level of diapause at all photoperiods considered (except LD 10:14). We reanalysed the above mentioned population, together with two other populations from Spain (one from intermediate latitudes: SP35, Lat 38.78° North, and one from the North: SP53, Lat 41.73° North, see Table 6.2) in an attempt to verify if this unexpected phenotype was peculiar to the SP-22 population or rather reflected the whole Spanish peninsula. The experiment was performed by inducing diapause at 12°C for 12 days, at two photoperiods, LD 16:8 and LD 8:16. Table 8.2 in Appendix 8.4.1 shows the mean levels of
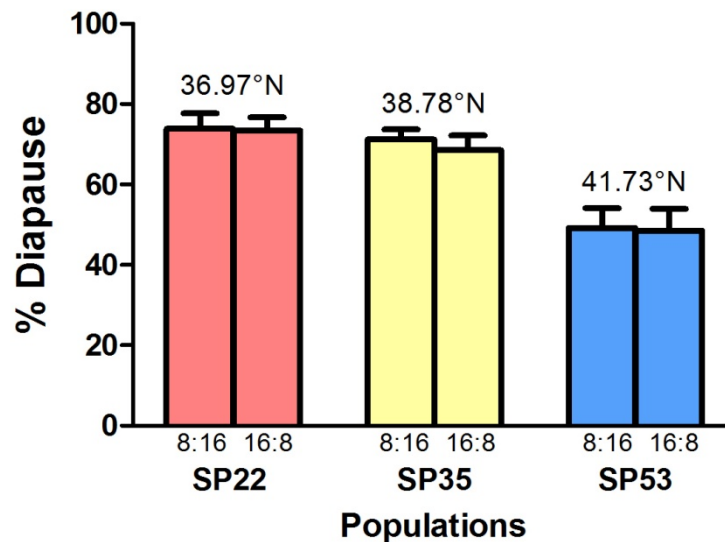
diapause identified.

Figure 6.4
Results of the diapause experiment performed on the three Spanish lines. Diapause was assessed at two photoperiods, LD 8:16 and LD 16:8. Y axis: averaged percentage of diapause (±SEM).

The results of this experiment are shown in Figure 6.4. Data regarding the DD controls can be found in Table 8.2 and Figure 8.2 of Appendix 8.4.1. The southernmost and the central population (SP22 and SP35 respectively) show a remarkably high level of diapause, thus confirming the previous results obtained in the laboratory and suggesting that the whole South of Spain is likely to be characterised by an unexpectedly high diapause incidence, rather than SP22 being an outlier.

As for the northernmost population (SP53), it is characterised by a lower diapause incidence. As shown in table 6.3 A, a 2-way ANOVA highlighted a population effect ($F_{2,30}=21.46$, $p<10^{-5}$) due to the lower diapause incidence in the SP53 population. This effect disappears when flies are kept in constant darkness. Surprisingly the three populations analysed do not seem to be able to discriminate between the two photoperiods (Photoperiod effect $F_{1,30}=0.15$, $p>0.70$).

A	df Effect	MS Effect	F	P value
Population	2	0.24	21.46	$<10^{-5}$
Photoperiod	1	$<10^{-2}$	0.15	0.70
Population X Photoperiod	2	$<10^{-3}$	0.04	0.96
Errors	30	0.01		
B	df Effect	MS Effect	F	P value
Population	2	0.03	1.44	0.31
Photoperiod	1	0.04	2.48	0.17
Population X Photoperiod	2	$<10^{-2}$	0.21	0.82
Error	6	0.02		

Table 6.3
A) Results of 2-way ANOVAs performed on the diapause experiment on the three Spanish populations. B) Results regarding the DD controls. Significant values are indicated in red.

6.3.2 EUROPE *VS* USA

As mentioned in the introduction, previous studies highlighted the difference in diapause-related phenotypes between American and European populations (Emerson *et al.*, 2009b), even though the results were collected under different conditions. Here we used American and European populations and compared their phenotypes using the same experimental protocol.

Flies were kept at 12°C, at two different photoperiods (LD 8:16 and LD 16:8) and were dissected after 12 and 28 days. The details regarding the populations are listed in Table 6.2, whereas Figure 6.5 shows the result of the experiment. The same data are plotted in different ways for clarity. The raw data, and the data regarding DD controls are shown in Table 8.3 and Figure 8.3 of Appendix 8.4.2.

Statistical analysis was performed on the 12 day, 28 day, LD 8:16 and LD 16:8 datasets. As for the 12 and 28 day categories (Table 6.4 A and B respectively), two sets of 2-way ANOVAs were carried out. The first set evaluated differences between lines and photoperiods, in the second set the different lines were clustered according to continent of origin, so that the effect of continent and photoperiod could be assessed. The same strategy was used for

the latter two categories (LD 8:16 and 16:8, Table 6.5 B and C respectively). In this case though, the effect of lines/continent of origin was tested together with the effect due to the length of exposure to diapause conditions (12 or 28 days).
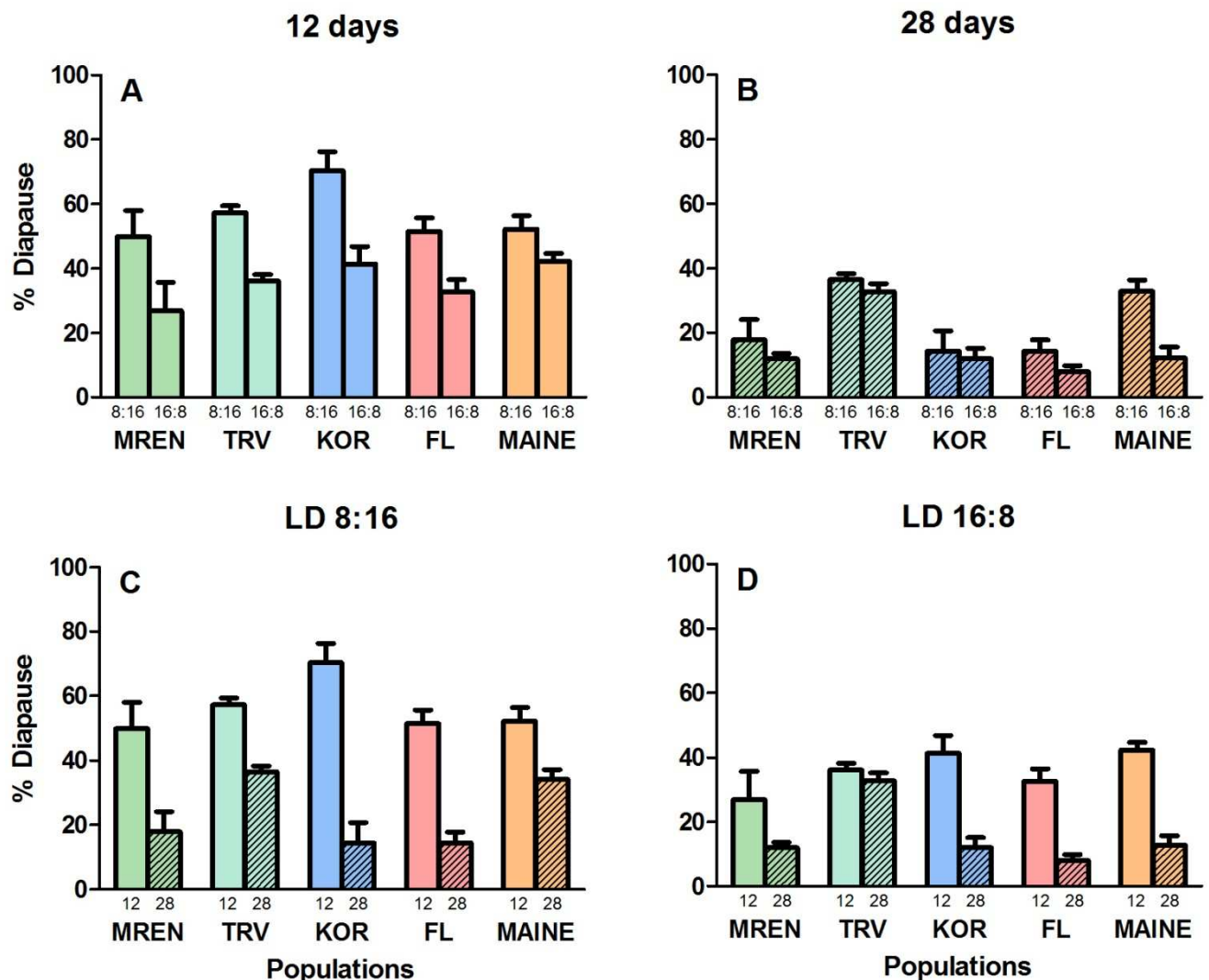

Figure 6.5

Diapause incidence in 3 European and 2 American lines (see Table 6.4 for details regarding the populations). Y axis: mean percentage of diapause averaged between 5/6 replicates (±SEM). The same data is clustered in different ways for clarity. A) Diapause assessed after 12 days, at two different photoperiods (LD 8:16 and LD 16:8). B) Diapause assessed after 28 days, at the two different photoperiods. C) Diapause tested at LD 8:16. The two time points are compared (12 and 28 days). D) Diapause levels at LD 16:8, comparison between the two time points. Striped bars represent 28-day data.

A	ALL EFFECTS (12 days)			B	ALL EFFECTS (28 days)		
	df Effect	MS Effect	F	p value	df Effect	MS Effect	F	p value
Line	4	0.08	2.61	0.04	4	0.18	12.13	$<10^{-4}$
Photoperiod	1	0.77	24.70	$<10^{-4}$	1	0.12	7.98	0.01
Line X Photoperiod	4	0.22	0.71	0.59	4	0.03	2.09	0.10
Error	54	0.03			47	0.01		
Continent	1	$<10^{-2}$	0.02	0.88	1	0.02	0.85	0.36
Photoperiod	1	0.60	17.65	$<10^{-3}$	1	0.18	6.65	0.01
Continent X Photoperiod	1	0.4	1.24	0.27	1	0.05	1.99	0.16
Error	60	0.03			53	0.03		

12 days DD controls				28 days DD controls			
Line	4	0.04	1.40	0.28	4	0.17	9.93	$<10^{-2}$
Photoperiod	1	0.18	5.73	0.03	1	0.09	5.51	0.04
Line X Photoperiod	4	0.09	2.92	<0.05	4	0.05	2.96	0.07
Error	17	0.03			10	0.02		
Continent	1	0.01	0.20	0.66	1	0.18	3.42	0.08
Photoperiod	1	0.12	3.13	0.09	1	0.08	1.49	0.24
Continent X Photoperiod	1	0.06	1.42	0.25	1	0.01	0.13	0.72
Error	23	0.04			16	0.05		

C	ALL EFFECTS (8:16)			D	ALL EFFECTS (16:8)		
	df Effect	MS Effect	F	p value	df Effect	MS Effect	F	p value
Line	4	0.06	2.10	0.09	4	0.07	4.33	$<10^{-2}$
Days	1	2.06	69.45	$<10^{-6}$	1	0086	49.91	$<10^{-6}$
Line X Days	4	0.11	3.87	0.01	4	0.06	3.49	0.01
Error	52	0.03			49	0.02		
Continent	1	0.01	0.14	0.71	1	0.01	0.51	0.48
Days	1	1.78	46.22	$<10^{-6}$	1	0.99	42.68	$<10^{-6}$
Continent X Days	1	0.02	0.64	0.43	1	0.08	3.46	0.07
Error	58	0.04			55	0.02		

8:16 DD controls				16:8 DD controls			
Line	4	0.07	3.08	<0.05	1	0.21	7.14	$<10^{-2}$
Days	1	0.40	16.84	$<10^{-2}$	1	0.28	9.60	$<10^{-2}$
Line X Days	4	0.03	1.26	0.33	1	0.27	0.95	0.47
Error	15	0.02			12	0.03		
Continent	1	0.18	6.84	0.02	1	0.01	0.17	0.69
Days	1	0.42	15.51	$<10^{-2}$	1	0.34	5.09	0.04
Continent X Days	1	0.01	0.45	0.51	1	0.06	0.91	0.35
Error	21	0.03			18	0.07		

Table 6.4

Results of 2-way ANOVAs performed on the 12 day, 28 day, LD 8:18 and LD 16:8 datasets (Table A, B, C and D respectively). Significant values are indicated in red.

A) At <u>12 days</u> (Table 6.4 A) a marginally significant line effect was detected ($F_{4,54}$=2.61, p=0.04). The photoperiod effect was highly significant ($F_{1,54}$=24.70, p<10^{-4}), although a marginal effect due to temperature differences between the two conditions is observed in DD controls ($F_{4,17}$=5.73, p=0.03). Importantly, when the analysis was performed by clustering lines from the same continent, photoperiod effect was still significant ($F_{1,60}$=17.65, p<10^{-3}) but not under the DD control conditions, and the continent of origin did not seem to influence diapause levels nor the way the populations respond to photoperiod.

B) A similar situation was identified when flies were dissected at <u>28 days</u> (Table 6.4 B): lines respond generally differently ($F_{4,47}$=12.13, p<10^{-4}), but in this case the difference persists in constant darkness ($F_{4,10}$=9.93, p<10^{-2}). Also after 28 days flies behave differently when exposed to the two photoperiods, and again there might be a contribution from the marginal temperature oscillations in the two conditions ($F_{1,47}$=7.98, p=0.01; $F_{1,10}$=5.51, p=0.04 in LD and DD respectively). When the data were clustered according to continent of origin, lines responded to photoperiod ($F_{1,53}$=6.65, p=0.01) but no Line X Photoperiod interaction was detected.

C) When the <u>LD 8:16</u> data were analysed (Table 6.4 C), a marginally significant line effect was detected only in the samples kept in DD ($F_{4,52}$=3.08, p<0.05). The time at which flies were dissected affected diapause levels in both LD and DD samples ($F_{1,15}$=69.45, p<10^{-6}; $F_{1,15}$=16.84, p<0.01 respectively). The lines seem to respond differently to the time of dissection, leading to a Line X Days interaction ($F_{4,52}$=3.87, p=0.01). When data from the same continent are grouped together, the continent of collection influences diapause levels only in DD ($F_{1,21}$=6.84, p=0.02). The time of dissection exerts an effect in both LD and DD samples ($F_{1,58}$=46.22, p<10^{-6} and $F_{1,21}$=15.51, p<0.01 respectively).

D) A similar *scenario* was observed for the <u>LD 16:8</u> results (Table 6.4 D). A line effect is present in both LD and DD samples ($F_{4,49}$=4.33, $p<10^{-2}$ and $F_{4,12}$=7.14, $p<10^{-2}$ respectively). Also the time of dissection has an effect in both conditions ($F_{1,49}$=49.91, $p<10^{-6}$ and $F_{1,12}$=9.60, $p<0.01$ in LD and DD respectively). Nevertheless the way the different lines respond to the two dissection time points is not homogeneous, leading to a significant Line X Days interaction ($F_{4,49}$=3.49, $p=0.01$). However lines originally from different continents behave similarly.

Taken together these data suggest that these lines are characterised by overall different levels of diapause incidence, which do not depend on the continent the lines originated from. The lines generally behave differently if exposed to two different photoperiods, but this difference can persist, albeit weakly, in DD controls. There is a general decrease in diapause levels between 12 and 28 days of exposure to diapause inducing conditions. The extent of this decrease is not homogenous between the lines, and it does not depend on the continent of origin. Despite the significant decrease in diapause levels, after 28 days flies maintain their capability of discriminating between the two photoperiods.

6.3.3 DIAPAUSE EVOLUTIONARY AGE

In the third and last part of this chapter we tackled the question of the evolutionary age of the diapause phenotype. Schmidt and co-workers suggested that this phenotype originated upon *D. melanogaster* colonisation of temperate climates (Schmidt *et al.*, 2005a). Plausibly, a seasonal adaptation response was unnecessary in the warm, constant and therefore very predictable environmental conditions of sub-Saharan Africa, from where *D. melanogaster* originated. Flies were subject to selective pressure only when they reached higher latitudes and were faced with challenging and variable temperatures and photoperiods. In support of this theory Schmidt *et al* (2005a) report not being able to detect any diapause in *D. melanogaster* African flies, and in the sibling species *D. simulans*. Given that these

experiments were performed under the 28-days protocol, and in light of our findings regarding the drop in diapause incidence with time, we decided to investigate diapause levels in African *D. melanogaster* flies *and D. simulans* flies collected in Europe and Africa, using the double approach of 12 and 28 days.

Table 6.2 reports the details regarding the populations used. The experiment was performed as above: diapause levels were tested after 12 and 28 days in diapause inducing conditions (12°C, at two photoperiods: LD 8:16 and LD 16:8). The results of the experiment are shown in figure 6.6, whereas Table 8.4 and Figure 8.4 of Appendix 8.4.3 report the raw data and show the results of the single populations plotted with the constant darkness controls. In Figure 6.6 the data regarding MREN are the same as those described before (Figure 6.5 and 8.3, and Table 8.3). They are reported here again in order to allow easy comparison with the other lines.

Table 6.5 reports the result of the 3-way ANOVAs performed on the data set. The effects of Species, Continent of origin, Photoperiods and Days in diapause conditions were assessed.
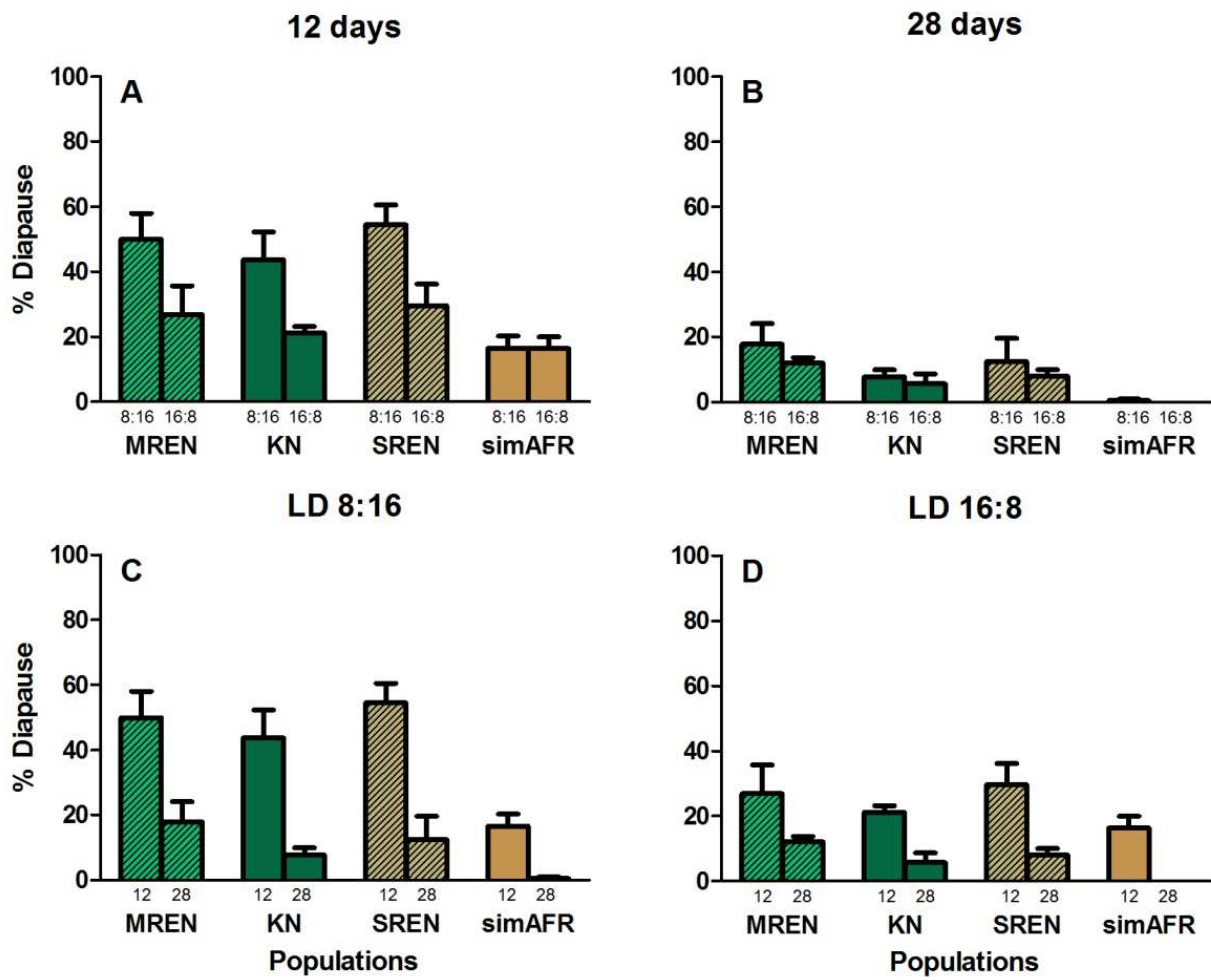
Figure 6.6

Diapause incidence in *D. melanogaster* (green bars) and *D. simulans* (brown bars) lines, collected in Europe and Africa (striped and plain bars respectively). See Table 6.6 for details regarding the populations. Y axis: percentage of diapause averaged between 5/6 replicates (±SEM). The same data is clustered in different ways for clarity. A) Diapause assessed after 12 days, at two different photoperiods (LD 8:16 and LD 16:8). B) Diapause assessed after 28 days, at the two different photoperiods. C) Diapause tested at LD 8:16. The two time points are compared (12 and 28 days). D) Diapause levels at LD 16:8, comparison between the two time points.

A	ALL EFFECTS (12 days)			B	ALL EFFECTS (28 days)		
	df Effect	MS Effect	F	p value	df Effect	MS Effect	F	P value
Continent	1	0.33	6.85	0.01	1	0.47	25.80	$<10^{-4}$
Species	1	0.07	1.35	0.25	1	0.27	15.21	$<10^{-3}$
Photoperiod	1	0.56	11.49	$<10^{-2}$	1	0.02	1.16	0.29
Continent X Species	1	0.18	3.62	0.06	1	0.03	1.65	0.21
Continent X Photoperiod	1	0.10	2.06	0.16	1	$<10^{-3}$	0.01	0.93
Species X Photoperiod	1	0.04	0.80	0.38	1	$<10^{-2}$	0.14	0.71
Continent X Species X Photoperiod	1	0.05	1.07	0.31	1	$<10^{-4}$	$<10^{-2}$	0.95
Errors	45	0.05			36	0.02		

12 days DD controls				28 days DD controls			
Continent	1	0.22	4.69	0.06	1	0.12	0.11	0.07
Species	1	0.11	2.48	0.14	1	$<10^{-2}$	0.11	0.75
Photoperiod	1	0.11	2.42	0.15	1	0.11	3.72	0.09
Continent X Species	1	0.07	1.60	0.23	1	0.09	3.18	0.11
Continent X Photoperiod	1	0.16	3.52	0.09	1	0.05	1.61	0.24
Species X Photoperiod	1	0.07	1.59	0.23	1	0.11	3.74	0.09
Continent X Species X Photoperiod	1	0.08	1.83	0.20	1	0.01	0.25	0.63
Errors	11	0.05			8	0.03		

C	ALL EFFECTS (8:16 days)			D	ALL EFFECTS (16:8 days)		
	df Effect	MS Effect	F	p value	df Effect	MS Effect	F	P value
Continent	1	0.60	14.72	$<10^{-3}$	1	0.23	8.04	0.01
Species	1	0.26	6.31	0.02	1	0.08	2.70	0.11
Days	1	2.17	52.86	$<10^{-6}$	1	0.90	31.03	$<10^{-5}$
Continent X Species	1	0.15	3.69	0.06	1	0.04	1.23	0.27
Continent X Days	1	$<10^{-2}$	0.09	0.76	1	0.05	1.83	0.18
Species X Days	1	0.01	0.17	0.68	1	0.05	1.72	0.20
Continent X Species X Days	1	0.05	1.17	0.28	1	$<10^{-4}$	$<10^{-3}$	0.98
Errors	41	0.04			40	0.03		

8:16 DD controls				16:8 DD controls			
Continent	1	0.53	16.15	$<10^{-2}$	1	0.01	0.21	0.66
Species	1	0.03	0.83	0.38	1	0.18	3.91	0.08
Days	1	0.62	18.79	$<10^{-2}$	1	0.57	12.21	0.01
Continent X Species	1	0.04	1.33	0.27	1	0.03	0.56	0.47
Continent X Days	1	0.02	0.61	0.45	1	$<10^{-2}$	0.03	0.86
Species X Days	1	0.05	1.64	0.23	1	0.03	0.54	0.48
Continent X Species X Days	1	0.04	1.10	0.32	1	0.14	0.07	0.12
Errors	11	0.03			8	0.05		

Table 6.5
Results of 3-way ANOVAs performed on the 12 day, 28 day, LD 8:18 and LD 16:8 datasets (Table A, B, C and D respectively). Significant values are indicated in red.

A) After <u>12 days</u> in diapause inducing conditions (Table 6.5 A) a significant difference between African and European flies was detected ($F_{1,45}$=6.85, p=0.01), with the African flies being characterised by slightly lower diapause incidence. Also the photoperiod effect was significant ($F_{1,45}$=11.49, p<10^{-2}). A marginally non-significant interaction between continent and species was identified ($F_{1,45}$=3.62, p=0.06), indicating that flies belonging to the same species but from two different continents may react slightly differently when exposed to diapause inducing conditions.

B) When flies were dissected after <u>28 days</u> (Table 6.5 B), both continent of origin and species affected the level of diapause ($F_{1,36}$=25.80, p<10^{-4} and $F_{1,36}$=15.21, p<10^{-3} respectively), with *simulans* flies, and flies from Africa being characterised by lower diapause incidence.

C) As for the <u>LD 8:16</u> data (Table 6.5 C), a continent effect was identified, in both LD and DD conditions ($F_{1,41}$=14.72, p<10^{-3} and $F_{1,11}$=16.15, p<10^{-2} respectively) as well as an effect due to species ($F_{1,41}$=6.31, 0.02). Diapause levels fall significantly with days of maintenance at 12°C, in both LD and DD conditions ($F_{1,41}$=52.86, p<10^{-6} and $F_{1,11}$=18.79, p<10^{-2} respectively). Although only marginally statistically significant, the analysis highlighted a possible interaction between continent and species ($F_{1,41}$=3.69, p=0.06), suggesting that flies belonging to the same species, but originally from different continents, might behave differently.

D) Finally when the <u>LD 16:8</u> data was analysed (Table 6.5 D) again a significant continent effect was identified ($F_{1,40}$=8.04, p=0.01). Also the time at which the samples were dissected played a significant role in determining diapause incidence, in both LD and DD samples ($F_{1,40}$=31.03, p<10^{-5} and $F_{1,8}$=12.21, p=0.01).

Taken as a whole these results highlight a difference between the species, but only at 28 days and in the short photoperiod LD 8:16. A much stronger effect on diapause incidence is played by the continent of origin of the fly line used in the experiment. Interestingly our data suggest that in some conditions (12 days and LD 8:16) flies from the same species behave differently, depending on which continent they come from. Not surprisingly photoperiod and time of dissection were also found to play an important role in determining diapause incidence.

6.4 Dɪscussɪoɴ

Previous studies and experiments carried out in our laboratory investigated diapause incidence in natural fly lines, and their capability to react when exposed to different photoperiods. Schmidt and co-workers analysed flies collected in a latitudinal cline along the US East coast, and showed that, if flies are dissected after 28 days at 12°C, diapause incidence predictably increases with latitude (Schmidt *et al.*, 2005b). A similar result was found in Australian lines, except here diapause surprisingly increases at very low latitudes (<24° South; (Lee *et al.*, 2011)). Results regarding the European dataset are in contrast with the studies cited here, and they point to a *scenario* where, after 28 days of diapause-inducing conditions, diapause levels do not correlate with latitude of collection (Pegoraro, Zonato *et al*., manuscript in preparation). The reason for this difference could be attributed to different intensities of climatic selection such as temperature, which shows a steeper cline on the eastern coast of America than in mainland Europe (see Final discussion, Chapter 7). Furthermore, as described in Chapter 3, the frequency distribution of the phylogenetically recent *ls-tim* variant which enhances diapause levels, is probably playing a prominent role in Europe. The (counter-intuitive) *ls-tim* high frequencies in southern Europe would be expected to flatten any

diapause cline in Europe. In contrast, the increase of *ls-tim* frequency with latitude in the USA will be contributing to the observed phenotypic cline. Unfortunately no data is available about the geographical distribution of *ls-tim* in Australia.

Photoperiodic curves were obtained for 6 European natural lines, whose level of diapause was evaluated after 12 days in diapause conditions. In these conditions a population from southern Spain was identified as an outlier, since its diapause incidence was found to be unexpectedly high at all photoperiods tested. We analysed three populations from the Spanish peninsula and our results suggest that the higher-than-expected diapause incidence is a phenomenon which is not peculiar to the southern population, but likely affects the populations from the Eastern side of Spain. Furthermore none of the three Spanish populations analysed seem to respond differently when exposed to the two photoperiods, unlike other European populations ((Tauber *et al.*, 2007), Pegoraro, Zonato *et al*., manuscript in preparation). The frequency of the allele *ls-tim* is rather low in the Spanish peninsula (Chapter 3), and therefore cannot be responsible for the high levels of diapause. Lee and colleagues (2011), who identified a U-shaped correlation between diapause and latitude in Australian populations, suggested that genetic factors affecting the diapause phenotype could also influence other behaviours which are favoured in the tropics. Nevertheless, the latitudinal range considered here (36-41° North) is far from the tropics, and it is very similar to the latitude in southern Italy, where flies are characterised by a lower diapause incidence (Tauber *et al.*, 2007). On the other hand, environmental factors other than photoperiod do not always correlate with latitude: average temperature and temperature fluctuations throughout the seasons, precipitations, drought and humidity can be drastically different in environments characterised by the same latitude. For example, even though the average temperature across the seasons is very similar in Spain and southern Italy, the diurnal temperature range (the difference in maximum and minimum temperature throughout the day) is significantly larger in southern Spain than in southern Italy (Figure 6.7). Diapause is a widespread phenotype,

which extends to tropical insects (Denlinger, 1986). They use diapause as a protection from the alternating dry/wet season, which leads to cycling presence of predator species and availability of food resources. This suggests that diapause can be triggered by cues other than low temperature and short photoperiods, which are never experienced by tropical insects. Therefore the high diapause observed in Spanish and tropical Australian flies could be manifested to cope with other stressful environmental conditions they are exposed to, rather than simply for overwintering.
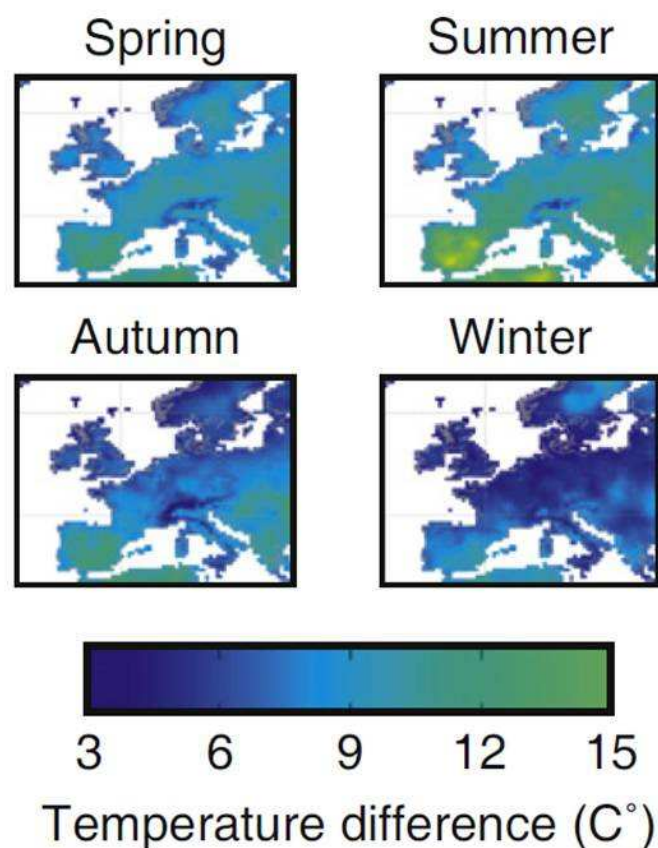
Figure 6.7
Diurnal temperature range (difference between maximum and minimum temperature throughout the day) across the seasons in Europe. Figure from (McClung & Davis, 2010).

As for photoperiodism, American flies do not behave differently when exposed to two different photoperiods after 28 days at 12°C (Emerson *et al.*, 2009b), unlike European flies, whose diapause levels, after 12 days at 12°C, decrease significantly as the day gets longer ((Tauber *et al.*, 2007); Pegoraro, Zonato *et al.*, manuscript in preparation)). Nevertheless the

different times at which flies were dissected in the American and European study might have had an effect on the diapause levels identified. In this chapter we analysed three European and two American natural *D. melanogaster* lines. The experiment was performed at two different photoperiods, and flies were dissected at both 12 and 28 days. Overall the results suggest that the lines do behave differently, but such difference does not depend on the continent of collection. A difference in diapause incidence between the two photoperiods was observed at 12 days and was maintained after 28 days, but the lack of interaction between Line/Continent and photoperiod suggests that all the lines react in similar ways when exposed to two different day lengths. Not surprisingly time of dissection plays a very important role in determining diapause levels. As previously observed by Saunders and colleagues in the fly laboratory strain Canton-S (Saunders *et al.*, 1989), the percentage of vitellogenic (non-diapausing) females kept at 12°C under a long photoperiod regime (LD 16:8) increases progressively reaching 100% after roughly a month. As for a shorter photoperiod (LD 10:14), spontaneous diapause remission is observed after 6-7 weeks. The reduced levels of diapause observed in our dataset after 28 days (in both photoperiods, and even in DD controls), therefore does not come as a surprise. Nevertheless this is probably the reason why Emerson and colleagues (Emerson *et al.*, 2009b), as well as Lee and co-workers (Lee *et al.*, 2011), identified a lower diapause incidence in American and Australian lines as compared to the published results regarding the European lines (Tauber *et al.*, 2007), which were dissected only after 12 days in diapause-inducing conditions. It could be argued that reproductive development generally takes longer than 12 days at such low temperature, therefore the observed increase in vitellogenic female flies with time might merely represent a delay in ovarian maturation rather than a spontaneous diapause remission. We cannot rule out the possibility that ovaries scored as "in diapause" after 12 days are actually still developing and have not entered diapause at all, but even so, this slowed-down-development must be under some form of "seasonal control" given the significant effect played by photoperiod even after

12 days. Furthermore even at 12 days ~50% of flies in the short photoperiod (LD 8:16) are vitellogenic, so clearly developmental is possible at 12 days, in about half the females.

It might seem counterproductive for a fly to undertake such profound hormonal and physiological changes, and to go back to normal reproductive state just after a few weeks, when in temperate regions the cold season can last up to several months. Natural populations of *D. melanogaster* could be ephemeral, with adults dying off in winter, and the region being repopulated again by flies from southern regions as spring comes. Nevertheless several experiments performed in the wild, together with the presence of latitudinal clines in gene frequencies, seem to support the idea that adult flies can successfully survive winter (Izquierdo, 1991; Mitrovski & Hoffmann, 2001; Costa *et al.*, 1992). The first thing to consider to solve this apparent contradiction between laboratory remission of diapause and what must occur in natural conditions, is that when winter approaches temperature and photoperiod get progressively lower/shorter, as opposed to the constant conditions used in our experimental protocols. A recent experiment performed in our laboratory clearly showed that when exposed to progressively shortening photoperiods and gradual lowering of temperature for 8 weeks to mimic oncoming winter, diapause incidence remains at constant higher levels than in flies kept in constant diapause-inducing conditions (L. Collins *et al.*, unpublished). Perhaps then it is not surprising that maintaining flies in constant temperatures of 12°C, which is at the threshold for inducing diapause, leads to remission, as flies may require additional environmental feedback to consolidate their response. Failure to supply this "winter signal" may simply allow them to switch back to vitellogenesis. Secondly, one must not forget that diapause is not the only strategy that flies have, when it comes to overwintering. Female flies have been shown to express an egg-retention-phenotype, whereby they delay egg deposition in unfavourable environmental conditions (Boulétreau, 1978). It has been shown that under semi-natural conditions, flies from temperate regions lay most of their eggs later than flies from lower latitudes (Mitrovski & Hoffmann, 2001). As eggs laid in cold weather conditions fail

to develop to adulthood, the delay in egg deposition by flies from temperate regions favours the survival of their offspring. Clearly egg retention would be disadvantageous for flies which are not exposed to seasonal changes, as in the tropics (Mitrovski & Hoffmann, 2001). Egg retention has also been observed in virgin females, who delay egg deposition as a response to the lack of sexual partners (initial retention (Boulétreau-Merle *et al.*, 1992)). Virgin females can postpone oviposition for 5-14 days, with flies from temperate regions being characterised by longer egg delaying times than flies from tropical regions (Boulétreau-Merle *et al.*, 1992). Unlike female flies from temperate regions, which face shortages of food and male partners at the beginning of the year, individuals from very low latitudes hardly face these conditions, and they are not subjected to selective pressure to delay egg deposition. However, and surprisingly perhaps, this does not seem to apply to flies collected along the eastern Australian coast, where no cline in initial egg retention was observed (Sgrò *et al.*, 2006). Finally, while there is no evidence for an overwintering strategy involving earlier developmental stages in *D. melanogaster*, it has been shown that eggs laid in natural conditions do develop into viable adults in 45-60 days, if the ambient temperature is around 12°C (Izquierdo, 1991). We can therefore conclude that the female's initial egg retention and the delay in fertilised egg deposition, together with the extended diapause observed in laboratory-simulated-natural-conditions and the slow egg to adult development observed at 12°C, combine so that flies (adults and embryos) can successfully survive long and quite extreme winter conditions.

Finally, we compared the level of diapause incidence measured at 12 and 28 days in *D. melanogaster* and *D. simulans* lines, captured in Italy and Africa. It has been previously suggested that *D. simulans* and African flies do not exhibit the phenotype, and therefore diapause has been attributed a very recent evolutionary age (Schmidt *et al.*, 2005a). We identified diapause in both *D. simulans* and African flies, especially when flies were dissected after 12 days. Continent of origin does influence diapause levels, after 12 or 28 days, and significant difference between the species only emerge when diapause is assessed after 28

days, in agreement with the observation reported by Schmidt. On the other hand flies behave differently when exposed to the two different photoperiods (but only at 12 days), which again makes it reasonable to assume that the early diapause levels measured at 12 days do not merely represent delay in reproductive development. An almost significant Continent X Species interaction was identified at 12 days, which suggests that in the two species the pheynotype, already present in African populations, may have adapted when flies colonised Europe. When the two photoperiods were analysed separately, the continent of origin continued to play an important role in determining diapause incidence, but the two species only seemed to behave differently in short photoperiods (LD 8:16). In both photoperiods diapause decreases dramatically with time, leading to a significant Days effect, but only in the short photoperiod LD 8:16 does there seem to be a marginally significant interaction between Continent of origin and Species, suggesting once again that the preexisting phenotype adapted to the novel seasonal environment encountered by flies in Europe. The continent effect partly is surprising since the *D. melanogaster* Kenya population used in this study was found to be characterised by high levels of non-African genetic admixture (*Drosophila* Population Genomics Project, DPGP2-African Survey, available at http://www.dpgp.org). The relatively high percentage of "European genes" in this population could in part explain why this African line surprisingly shows diapause levels comparable to the ones observed in European populations. Nevertheless the finding that this phenotype is also observed in the two *D. simulans* lines suggests that diapause is more likely to have ancient origin rather than having appeared after the flies colonisation of Europe.

These results indicate that diapause may be of more ancient origin than previously suggested. The molecular machinery to switch off reproduction was already present in African *D. simulans* flies, and it might have been used as a stress response triggered by unfavourable conditions such as drought (wet *vs* dry season), lack of resources, presence of predators. When flies eventually reached temperate environments such machinery was put under strong

seasonal selection. The preexisting stress response mechanism slowly became sensitive to new environmental cues such as yearly changes in photoperiod and temperature. From an evolutionary perspective the fine tuning and adjustment of a general stress phenotype to new external stimuli is much less costly in terms of resources than designing a novel mechanism to accomplish such a task.

The results shown in this chapter can be summarised as follows:

- The southern Spanish line is not an outlier in terms of diapause incidence, since a very similar phenotype has been observed also in another population from central Spain. The genetic and environmental factors responsible for this unexpectedly high diapause incidence still remain unknown.

- When subjected to the same experimental protocol, no significant difference was observed between American and European *D. melanogaster* natural lines, in terms of diapause levels and photoperiodism.

- A significant decrease in diapause is observed after 28 days in diapause inducing conditions.

- Unlike previous suggestions, diapause appears to be an ancient phenotype, already present in the form of a stress response in African *D. simulans* flies. Nevertheless this behavioural response adapted to the temperate European environmental conditions.

7. FINAL DISCUSSION

The work presented in this thesis focused on diapause, the seasonal adaptation phenotype in *D. melanogaster.* In particular, my experiments aimed at studying the strategy adopted by wild flies in order to adapt to different thermal and photoperiodic environments. What makes wild flies different is not the diapause machinery itself apparently, but its fine tuning which allows natural flies to perceive the same external stimuli but respond to them differently, depending on their geographic origin.

One of the focal points of this project was the identification of latitudinal clines, at both the genetic and phenotypic levels. Latitudinal clines in allele frequencies and/or external traits have long been used to infer adaptation to those environmental conditions which change with latitude, namely temperature and photoperiod. Consequently, the geographical distribution of alleles of three diapause genes, and of the diapause incidence phenotype, have been analysed in Europe, USA and Australia. The overall results are summarised in Table 7.1.

	tim	*Cpo 356*	*Cpo 462*	*InR*	**D**iap
USA	⁄ (green)	⁄ (green)	⁄ (green)	⁄ (green)	⁄ (green)
Europe	✕ (blue)	⁄ (green)	— (blue)	— (blue)	— (blue)
Australia	**?**	**?**	** ⁄ (green)	⁄ (green)	⌣ (blue)

Table 7.1

Summary of the clines in allele frequency and phenotype studied in this work. A green symbol refers to a cline where frequency increases with latitude. Blue symbols are used when a different situation was identified e.g. lack of cline or atypical geographical distribution. Question marks indicate 'not known' situations. *tim* refers to the frequency of the allele *ls-tim* (Chapter 3). *cpo 356* and *462* refer to the frequency of cpo^{356Val} and cpo^{462Lys} respectively (Chapter 4). *InR* refers to the frequency of the allele *248* in *InR* (Chapter 5), whereas the column indicated by "Diap" contains the results regarding the diapause phenotype. ** Cline driven by the chromosomal inversion 3(*In(3R)P*). Data regarding the clines of *cpo* in the USA and Australia can be found in Schmidt *et al*., 2008 and Lee *et al*., 2011 respectively. Schmidt and co-workers (2005a) and Lee *et al*. (2011), analysed the diapause incidence of American and Australian natural flies.

7.1 CONTINENT COMPARISON: DNA LEVEL

A first look at Table 7.1 highlights a complex situation, where none of the characters studied shows the same distribution in the three continents considered here. Surprisingly enough the odd one out appears to be Europe, where only cpo^{356Val} shows a significant latitudinal cline in frequency. All other allele frequencies and the phenotype itself do not vary significantly with latitude in the Old World. Such a situation was unexpected, given the demographic history of *D. melanogaster*. *D. melanogaster* is thought to have originated in tropical Africa, and to have colonised the rest of the African continent, Eurasia and the Far East due to its dispersal capacities, around 15 kya (Li & Stephan, 2006) (blue arrows in Figure 7.1). America and Australia were probably colonised much more recently, around 300 and 150 years ago respectively, through human-associated transportation (David & Capy, 1988) (red arrows in Figure 7.1).
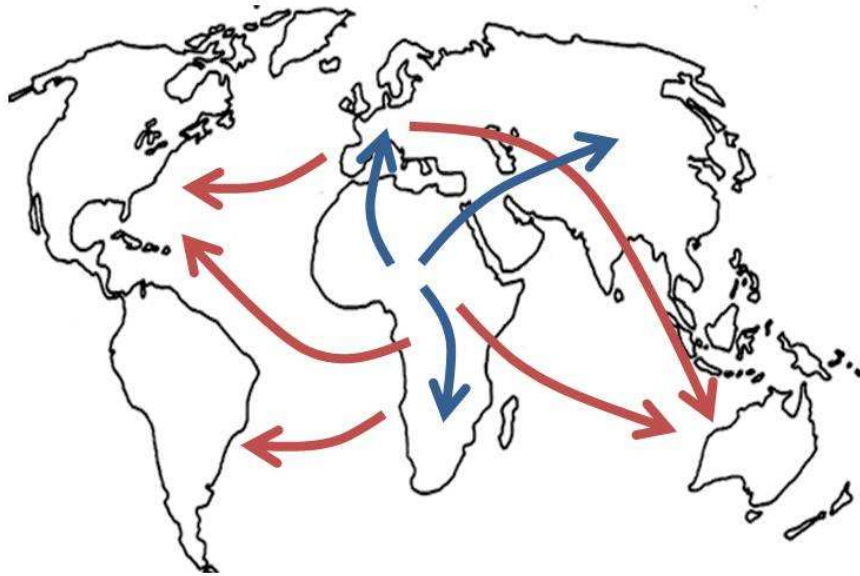
Figure 7.1
D. melanogaster colonisation dynamics. Thanks to their dispersal abilities flies reached the Old World and the Far East from sub-Saharan Africa flies around 15 Kya (blue arrows). Only with human migration did they reach the New World and Australia (150-300 years ago, red arrows). Figure redrawn from (David & Capy, 1988).

The time required for the establishment of a new cline depends on a variety of factors, such as the strength of the selective pressure, the extent of the benefit conferred by the new phenotype, the generation time and dispersal ability of the organism under study. Given the large differences in terms of photoperiod between northern and southern Europe, the compromised fitness of flies with poor overwintering capacities and the short *D. melanogaster* generation time, one would assume 15 Kya to be long enough for the establishment of a diapause-related cline in the Old World. The presence of a well characterised adaptive temperature-related cline for different alleles of the *period* gene in the Old World supports this idea (Costa *et al.*, 1992).

Nevertheless not only does Europe not seem to be characterised by the presence of adaptive diapause-related clines, but these seem to be present in other continents. In North America, the frequency of the three alleles *ls-tim*, *cpo*[356Val], *cpo*[462Lys] and *248* in *InR* all increase with latitude of collection. All these genes have been shown to be involved in regulating the

diapause phenotype (Tauber *et al.*, 2007; Schmidt *et al.*, 2008; Paaby *et al.*, 2010) and in particular these alleles enhance diapause incidence, so that flies with these alleles enter diapause earlier in the year than flies with different genotypes. Their increase in frequency with latitude therefore does not come as a surprise, since flies from the more seasonally challenged environments would be expected to benefit from being able to enter diapause relatively early.

If 15K years were not long enough for the establishment of a cline in the Old continent, how can 300 years have been enough, in the USA? One of the main reasons for this apparently contradicting situation is to be sought for in the environmental conditions characterising the two continents. In our studies we used latitude as a proxy for the selective pressure, and whereas latitude correlates perfectly with photoperiod, this does not hold true for temperature.

Figure 7.2 shows how the average yearly minimum and maximum temperature changes with latitude in North America and in Europe. The temperature data collected along the USA East coast have much smaller standard deviations than the data regarding the minimum temperatures in Europe. As each point represents many years of data, this means that the temperature must fluctuate in a rather unpredictable way in Europe. The statistically significantly steeper temperature-latitude cline for the USA data suggests that in North America compared to Europe, any differential selective pressure at the latitudinal extremes would be more extensive. This may contribute to why we do not observe in Europe the latitudinal clines present in the USA. Furthermore if in Europe the temperature fluctuates considerably year after year, having a less structured and rigid diapause machinery may be advantageous. The phenotype might not be perfectly tuned to every year's weather conditions, but it allows the species to be flexible in a wide range of conditions. Overall, the lack of clines in Europe is not an indication that flies in the Old World are not adapted to their

surroundings. On the contrary, they are perfectly adapted to an environment which is space-wise more homogeneous but time-wise more variable than in the USA.
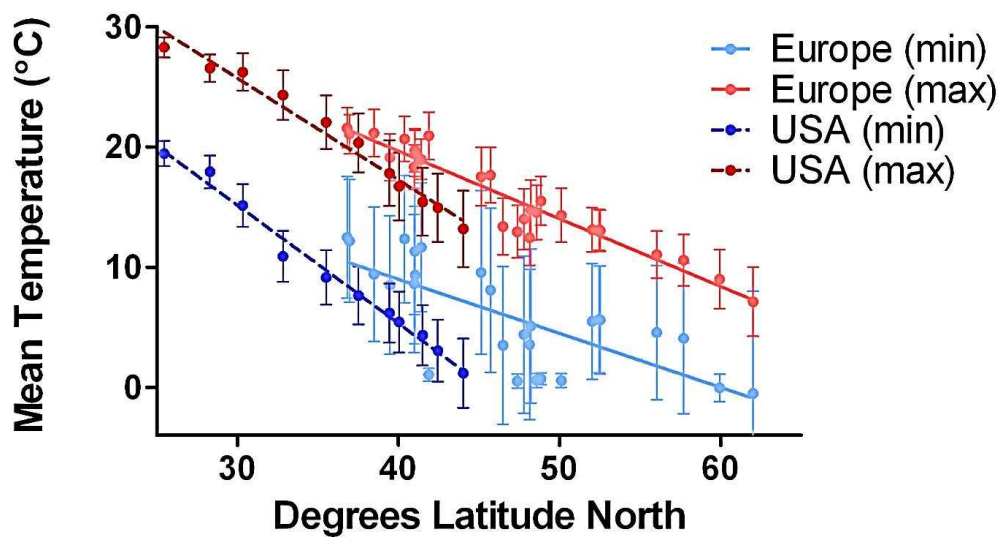

Figure 7.2
Correlations between minimum and maximum temperatures and latitude in USA and Europe. The temperatures were obtained from the Worldclim database (www.worldclim.org), representing observations from 1950 to 2000. R^2(Europe min)=0.52, p<10^{-3}; R^2 (Europe max)=0.80, p<10^{-3}; R^2(USA min)= 0.99, p<10^{-3}; R^2 (USA max)=0.97, p<10^{-3}. The differences between the slopes are significant (Europe and USA max: $F_{1,452}$=5.20, p=0.02; Europe and USA min: $F_{1,452}$=24.73, p<10^{-3}). The data corresponds to the locations of North American populations collected by Schmidt et al., (2010) and for European locations in our study, plus some additional European sites including Spain, France, Austria, Germany, Denmark, UK, Sweden and Norway. Average daily temperature during the month of maximum and minimum temperature is presented for each location (Mean and SD are given).

The different selective strength probably plays the major role in causing the observed differences between continents, but surely it is not the only factor contributing to the process. As described in Chapter 3 the evolutionarily young allele *ls-tim* shows a peculiar distribution in Europe because the putative site of origin of this mutation. Had the single nucleotide insertion occurred in the North of Europe, a predictable and traditional latitudinal cline in frequency might have been found in Europe as well as in the USA. Instead, the data gathered by Tauber and colleagues (2007) and the data shown in this work, suggest that the new allele first appeared in the South of Italy and its dispersion, probably aided by human migration, generated the atypical distance-from-site-of-origin cline that we observe today. The fact that

at the same distance from Novoli, the frequency of the new allele is higher at higher latitudes makes it reasonable to assume that a differential selection is applied to the allele in different European locations. We can speculate that in the future the *ls-tim* cline might disappear in Europe as *ls-tim* goes to fixation.

Another factor which might have contributed to creating the complex situation observed in Europe is population migration, which could affect the LD between the two SNPs in *cpo*. Whereas the LD is very strong in the two American populations analysed by Schmidt *et. al.*, (2008), it does not seem to be the same in Europe (Chapter 4). This difference might be due to the founder effects faced by European and African populations upon their colonisation of the USA, around 300 years ago (David & Capy, 1988). The stronger LD means that a latitudinal cline in frequency in one locus is more likely to drive the geographical distribution of a flanking haplotype over a longer genomic distance compared to flies characterised by a weaker LD. Hence we have the complicated situation regarding the two *cpo* SNPs in USA and Europe. What is interesting is that Schmidt's study (2008) and our work indicate different SNPs as responsible for modulating the diapause phenotype. Recent work carried out on Australian populations (Lee *et al.*, 2011) seems to reach conclusions in agreement with that on European data, given that *cpo* SNP I462K was found to be clinal, but only due to a LD effect, and not to its influence on diapause levels. We suggest that *cpo* exerts its influence in modulating diapause levels through SNP A356V, which shows a strong latitudinal cline in the USA, and a weaker but still significant one in Europe, probably because of the weaker European selective pressure.

As for the *InR*, as described in Chapter 5, a diapause-relevant cline was observed in an indel polymorphism, both in Australia and USA, but not in Europe. Given that the same allele pool was found in the three continents (and in Africa) a lack of cline in Europe is probably due to the different environmental conditions affecting this continent.

One might wonder why, despite the presumably weaker selective strength, *cpo* SNP A356V shows a significant (although less robust) cline in Europe, whereas *InR* diapause-relevant indels are completely flat in the Old World. As explained in Paragraph 1.5 and 1.6 of the Introduction, these two genes act at two different steps of the intricate signaling cascade which eventually leads to diapause expression. It is therefore very likely that they respond differently to the same selective pressure.

7.2 CONTINENT COMPARISON: PHENOTYPE LEVEL

When this project was initiated, European and American natural flies appeared to show diapause phenotypes which differed in at least three characteristics: unlike European ones, American lines appeared to be non-photoperiodic, to show a rather low diapause level, and to be characterised by a clear significant correlation between latitude of collection and diapause incidence (Schmidt *et al.*, 2005; Emerson *et al.*, 2009b) (Pegoraro, Zonato, *et al.*, manuscript in preparation).

These discordances turned out to be due to a combination of technical and biological reasons. First of all, leaving flies in diapause conditions for as long as 28 days leads a much lower observed level of diapause, which hides the flies capability to discriminate between the two photoperiods at which they were exposed. When European and North American flies were tested using the same experimental approach, no difference was detected due to their origin (Chapter 6). All the fly lines tested, regardless of continent of collection, show the same diapause levels and the same photoperiodism, which in general decreases as a function of time spent in constant diapause inducing conditions, as originally reported by Saunders and colleagues (1989). The 12 *vs* 28 day protocol also helped shed some light into the evolutionary age of the diapause phenotype. Significant differences in behaviour were identified between African and European lines, but at least after 12 days in winter condition *D. simulans* flies are

indistinguishable from *D. melanogaster* flies, indicating that, despite previous suggestions (Schmidt *et al.*, 2005a), diapause represents an ancient survival strategy.

If the SNPs responsible for tuning the diapause machinery to the gradually changing environment do not change over latitude as impressively as along the US East coast, then consequently the diapause phenotype itself will not show a gradual and smooth change in incidence over latitude. As for Australia, a cline in diapause incidence has recently been identified (Lee *et al.*, 2011), which nevertheless becomes non-linear when populations from close to the Equator are included in the study (Table 7.1). An unexpected and very interesting diapause increase in tropical flies confirms the hypothesis that the phenotype can be triggered by inputs other than winter: probably an ancestral, underlying capability which unveils the broader biological relevance of the ancient phenotype.

Overall, we observed profound difference at both DNA and phenotype level between these three continents, which likely reflect different environmental conditions and population dynamics affecting these locations.

7.3 FUTURE DIRECTIONS

Are there still central jigsaw pieces missing from this picture? Australian populations need to be analysed in terms of their *ls-tim* and *cpo* A356V allele frequencies, and that would give us a deeper insight into the differences between continents. Data regarding new locations also need to be gathered: Asia would be a good candidate since it is characterised by ancient populations of *D. melanogaster* (David & Capy, 1988), thus allowing us to assume that enough time has passed for the establishment of any diapause-related cline. Furthermore it is characterised by a colonisation history and climatic conditions which clearly differ from Europe.

It would also be interesting to check for the presence of altitudinal clines, at both DNA and phenotypic level. That would take photoperiod out of the equation and would allow us to study separately the strength of temperature as a selective pressure to generate diapause-related clines. The comparisons between altitudinal transects collected at different latitudes would then offer a way to integrate the effects of the two main diapause triggers.

As for the diapause phenotype experiments, flies maintained in diapause conditions for only 12 days need to be characterised in terms of their life history traits. They are mostly characterised by previtellogenic ovaries, whose percentage depends on the photoperiod used in the experiment, thus suggesting a diapause phenotype rather than a slower development due to the low temperature. Nevertheless a difference in fitness/stress resistance/longevity between flies maintained in diapause inducing conditions for only 12 days, at two different photoperiods, will further confirm and validate our observations (Schmidt *et al.*, 2005a).

Finally other forms of stresses (physiologically relevant or not, such as oxygen deprivation and drought) could be simulated in the laboratory to test the hypothesis that diapause originated as a way to survive general stress, which then was fine tuned by seasonally challenging environments in order to evolve into an overwintering strategy.

8. APPENDIX

8.1 GENETIC DIFFERENTIATION

8.1.1 VARIABLE DESCRIPTION

The values reported in Table 3.7 can be briefly described as follows:

- χ^2: traditional method to test for genetic differentiation between populations, based on haplotype frequency

$$\chi^2 = \sum_{i=1}^{L} \sum_{j=1}^{K} \frac{(n_{ij} - n_i \widehat{p}_j)^2}{n_i \widehat{p}_J}$$

where L is the number of localities tested, K is the number of haplotypes, n_i is the sample size from locality i, n_{ij} is the observed number of copies of haplotype j from locality i, and $\widehat{p}_J$ is the frequency of haplotype j in the total sample.

Under the null hypothesis of no differentiation, this value should be χ^2 distributed (Df=(L-1)(K-1)). When the frequency of some haplotypes is very low, this statistic is not very accurate.

- H_{ST}: yet another haplotype-based statistic, defined as follows:

$$H_{ST} = 1 - \left(H_S / H_T \right)$$

Where H_S is the weighted average of the estimated haplotype diversity within the subpopulations, and H_T is the estimated haplotype diversity in the total population.

Haplotype diversity measures the uniqueness of a specific haplotype in the population, and is define as:

$$H = \left(N / N - 1 \right) \sum_{i=1}^{N} x_i^2$$

Where N is the sample size and x is the frequency of a particular haplotype in the sample.

- K_{ST}: statistic based on nucleotide differences:

$$K_{ST} = 1 - \left(K_S / K_T \right)$$

Where K_S is the weighted average of the nucleotide differences within the individual populations, and K_T is the average of differences between sequences, regardless of their place of origin.

- $K_{ST}*$: this statistic is calculated as K_{ST} except it does not give much weighting to large nucleotide difference. Instead of $d_{ij,lk}$ (number of difference between the sequence j from location i and the sequence i from location k), used in the K_{ST} statistic, $log\left(1 + d_{ij,lk}\right)$ is used here. This downweights large values of $d_{ij,lk}$.

- Z: it represents a rank statistic. First all the $d_{ij,lk}$ values are ranked according to their values. Z represents the weighted sum of Z_1, Z_2... Z_i, where Z_i is the average of the ranks of all the $d_{ij,lk}$ calculated for pairs of sequences within the same location. The null hypothesis of no differentiation between the locations analysed is rejected if Z is too small.

- Z*: this statistic is calculated as Z, but as for K_{ST}*, the logarithm of 1 plus the rank of the $d_{ij,lk}$ is used.

8.1.2 PERMUTATION BASED TEST

The values described above do not measure the extent of the difference between the populations under study. They rather give an indication as to whether the populations under study are genetically different or not. Therefore a statistical test is required to decide if the null hypothesis (that the populations are not genetically different) can be rejected or not. In order to do so, DNAsp 5.10.01 uses a permutation based test which assigns a statistical significance to the values listed above.

The p-values of the observed value of the statistic is estimated by pooling the data regarding all the populations together, partitioning the pooled sample 1000 times in as many parts as original populations, and calculating the statistic every time a new sample is partitioned. The p-values then represent the probability of obtaining the observed value (or a more extreme one) on the partitioned samples.

8.2 LINKAGE DISEQUILIBRIUM

The three LD estimators computed by the package "Genetics" of the software R, are the following:

- Difference between the observed frequency of an haplotype, and its expected value:

$$D = p(AB) - p(A)p(B)$$

- Normalised D, independent on the values of the allele frequencies. It can only reach values between 0 (complete linkage equilibrium) and 1 (complete linkage disequilibrium):

$$D' = \frac{D}{Dmax}$$

Where $Dmax = min\big(p(A)p(b), p(a)p(B)\big)$ if D>0

and $Dmax = max\big(-p(A)p(B), -p(a)p(b)\big)$ if D<0

- Correlation coefficient between the two loci:

$$r = \frac{-D}{\sqrt{p(A)p(a)p(B)p(b)}}$$

Where:

- $p(A)$ is defined as the observed frequency of allele A for marker 1

- $p(a)$ is defined as the observed frequency of allele a for marker 1

- $p(B)$ is defined as the observed frequency of allele B for marker 2

- $p(b)$ is defined as the observed frequency of allele b for marker 2

- $p(AB)$ is defined as the observed frequency of the haplotype AB

- $p(A)p(B)$ represents the expected of the haplotype AB

For genotype data, the genotype AB/ab cannot be distinguished from aB/Ab. Consequently R estimates $p(AB)$ using maximum likelihood and uses this value in the computations. Alternatively, the analysis was also performed not including data regarding flies heterozygous at both the markers. The values obtained with the two methods were then compared.

8.3 DIAPAUSE PHENOTYPE (Chap. 4)

Raw data (Table 8.1) and plots (Figure 8.1) regarding diapause levels of the individual populations described in Chapter 4. For details about the populations and the experiment see paragraph 4.3.8 of Chapter 4.

	DIAPAUSE RESULTS 12 DAYS			DIAPAUSE RESULTS 28 DAYS		
LINE	FLIES	8:16	16:8	FLIES	8:16	16:8
TOT	309 (6, 6)	57.26 (5.30)	36.15 (5.07)	280 (6, 5)	36.56 (4.51)	32.82 (5.51)
	94 (2, 2)	52.96 (8.90)	55.65 (2.10)	93 (2, 2)	52.92 (10.02)	52.27 (9.64)
sTA	313 (6, 6)	44. 54 (1.05)	41.86 (1.16)	306 (6, 6)	40.32 (2.29)	38.77 (2.36)
	111 (2, 2)	46.61 (2.52)	44.95 (2.09)	89 (2, 2)	39.20 (2.41)	39.56 (0.61)
sTT	297 (6, 6)	43.73 (1.01)	40.30 (1.84)	291 (6, 6)	40.54 (2.39)	40.72 (1.46)
	87 (2, 2)	46.41 (2.00)	45.42 (0.59)	98 (2, 2)	39.23 (1.09)	40.37 (0.52)
sCA	281 (6, 5)	39.66 (9.30)	29.28 (2.70)	268 (5, 6)	21.20 (4.07)	19.79 (4.57)
	93 (2, 2)	33.20 (8.38)	28.69 (15.99)	85 (2, 2)	18.17 (19.55)	24.85 (16.28)
sCT	302 (6, 6)	38.27 (1.06)	32.94 (4.36)	299 (6, 6)	22.91 (0.98)	22.04 (4.16)
	100 (2, 2)	41.74 (2.46)	40.48 (1.68)	105 (2, 2)	32.74 (0.84)	32.18 (1.62)

Table 8.1

Raw data regarding the diapause experiment described in Chapter 4. Flies: number of flies analysed per condition, in brackets the number of replicates performed in the two photoperiods respectively. 8:16 and 16:8 indicate the two photoperiods at which the experiment was performed. Values are averaged percentage of diapause (±SEM). Shaded cells indicate results obtained in DD conditions

Figure 8.1 (See below)

Diapause of the individual populations used in the experiment described in Paragraph 4.3.8. Flies were dissected after 12 days at 12°C. The black data points represent the data regarding DD controls. A and B) Population from Treviso (Italy), at 12 and 28 days. C and D) sCA line at 12 and 28 days. E and F) sCT line at 12 and 28 days. G and H) sTA at 12 and 28 days. I and J) cTT line at 12 and 28 days.
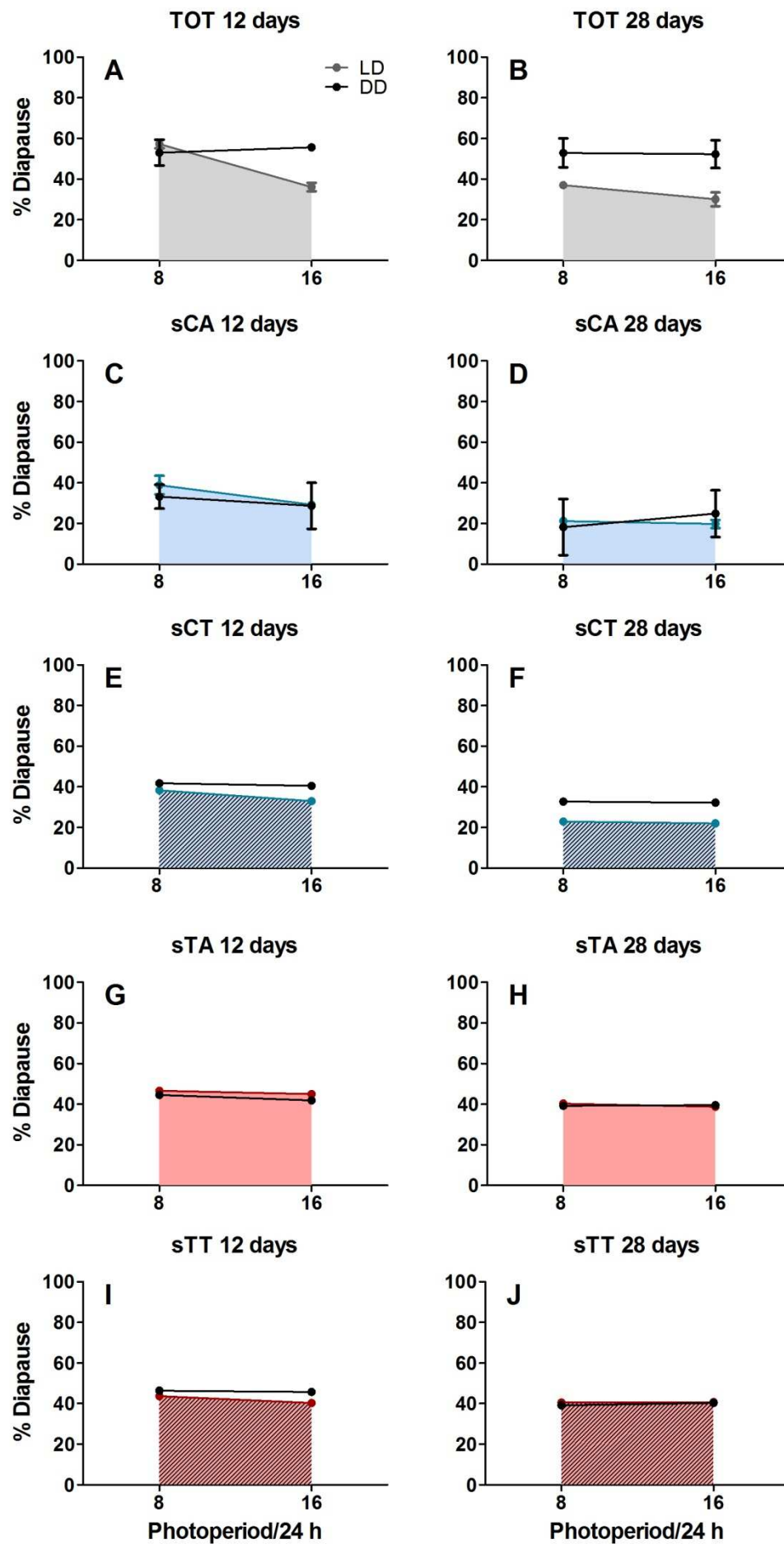
8.4 DIAPAUSE PHENOTYPE (Chap. 6)

8.4.1 DIAPAUSE IN SPANISH LINES

Table 8.2 shows the details of the lines used in Paragraph 6.3.1 of Chapter 6, whereas Figure 8.2 shows the levels of diapause of the individual populations with the data regarding the constant darkness controls (DD).

LINE DETAILS						DIAPAUSE RESULTS		
LINE	LOCATION	N	LAT	LONG	ALT	FLIES	8:16	16:8
SP-22	Nijar	24	36.97	-2.21	345	271 (6, 6)	73.89 (3.81)	73.49 (3.18)
						86 (2, 2)	70.29 (3.62)	57.14 (4.76)
SP-35	Fontanares	32	38.78	-0.79	630	249 (6, 6)	71.17 (2.53)	68.55 (3.65)
						86 (2, 2)	66.36 (6.36)	49.58 (20.42)
SP-53	Manresa	20	41.73	1.82	278	272 (6, 6)	49.19 (5.39)	48.56 (5.92)
						89 (2, 2)	50.89 (3.27)	45.45 (0)

Table 8.2

Details of the Spanish lines used in the diapause experiment. N: number of isofemale lines used to create the three populations analysed. Lat: latitude in degrees north. Long: longitude in degrees (Negative values: west; Positive values: east). Alt: altitude in meters above sea level. Flies: number of flies analysed per condition, in brackets the number of replicates performed in the two photoperiods respectively. 8:16 and 16:8 indicate the two photoperiods at which the experiment was performed. Values are averaged percentage of diapause (±SEM). Shaded cells indicate results obtained in DD conditions.
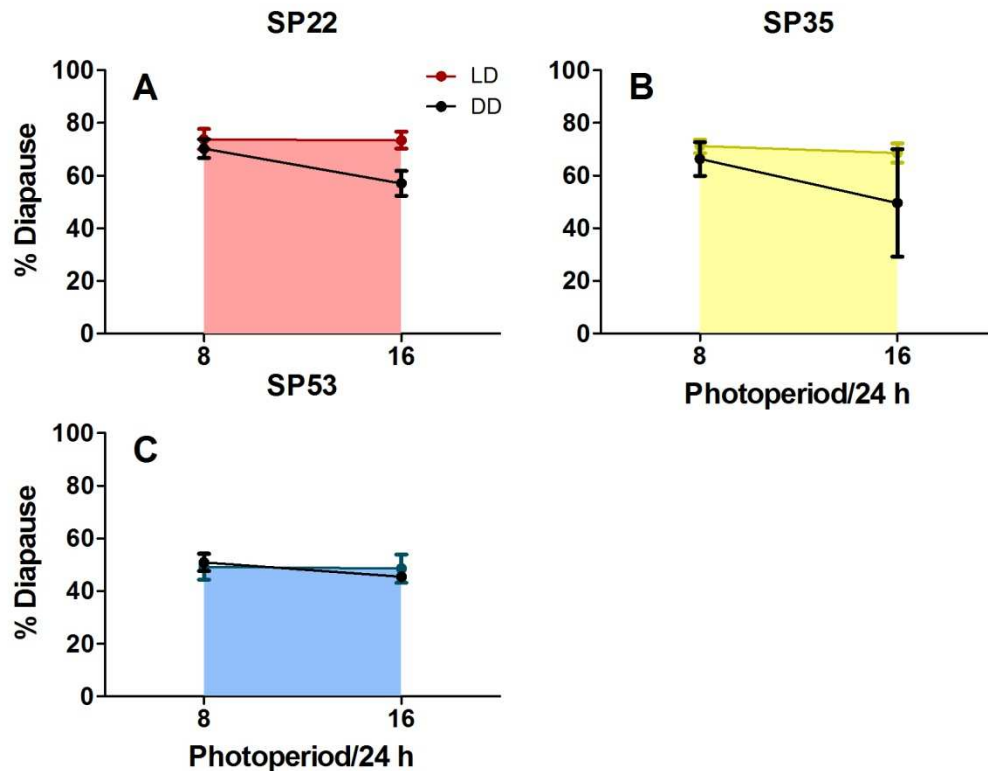
Figure 8.2
Diapause of the individual populations used in the experiment described in Paragraph 6.3.1 of Chapter 6. Flies were dissected after 12 days at 12°C. The black data points represent the data regarding DD controls. A) Population SP22. B) Population SP35. C) Population SP53. See Table 6.2 for details regarding these lines.

8.4.2 EUROPE *VS* USA

Table 8.3 shows the details of the lines used in Paragraph 6.3.2 of Chapter 6, whereas

Figure 8.3 shows the levels of diapause of the individual populations with the data regarding

the constant darkness controls (DD).

	LINE DETAILS					DIAPAUSE RESULTS 12 days			DIAPAUSE RESULTS 28 days		
LINE	LOCATION	N	LAT	LONG	ALT	FLIES	8:16	16:8	FLIES	8:16	16:8
MREN	Rende (Italy)	25	39.20	16.11	480	419 (10, 6)	49.97 (8.09)	26.90 (8.84)	403 (4, 6)	17.88 (6.28)	12.06 (7.67)
						196 (5, 2)	59.63 (9.77)	12.96 (12.96)	134 (2, 2)	36.73 (4.30)	6.00 (6.00)
TRV	Treviso (Italy)	35	45.67	12.24	40	309 (6, 6)	57.26 (2.16)	36.15 (2.07)	411 (6, 5)	36.56 (1.84)	32.82 (2.46)
						140 (2, 2)	52.96 (6.30)	55.66 (1.40)	138 (2, 2)	52.92 (7.08)	52.27 (6.82)
KOR	Korpilahti (Finland)	20	62.01	25.33	120	413 (6, 6)	70.34 (5.98)	41.38 (5.23)	396 (6, 6)	14.36 (6.33)	12.01 (3.21)
						258 (4, 4)	52.57 (7.57)	36.24 (3.41)	127 (2, 2)	14.52 (4.84)	13.33 (6.67)
FL	Homestead (Florida, USA)	12	25.47	-80.48	1	413 (6, 6)	51.55 (4.13)	32.74 (3.84)	396 (6, 6)	14.34 (3.42)	8.07 (1.80)
						258 (2, 2)	40.39 (9.62)	31.88 (13.13)	127 (2, 2)	15.92 (1.22)	4.69 (4.69)
MAINE	Bowdoinham (Maine, USA)	50	44.01	-69.9	1	399 (6, 6)	52.24 (4.22)	42.37 (2.34)	411 (6, 6)	34.09 (3.16)	12.88 (2.83)
						140 (2, 2)	45.06 (0.23)	44.83 (0.00)	128 (2, 2)	12.12 (0.00)	17.50 (3.71)

Table 8.3

Details of the European and American lines used in the diapause experiment. N: number of isofemale lines used to create the populations analysed. Lat: latitude in degrees north. Long: longitude in degrees (Negative values: west; Positive values: east). Alt: altitude in meters above sea level. Flies: number of flies analysed per condition, in brackets the number of replicates performed in the two photoperiods respectively. 8:16 and 16:8 indicate the two photoperiods at which the experiment was performed. Values are averaged percentage of diapause (±SEM). Shaded cells indicate results obtained in DD conditions.
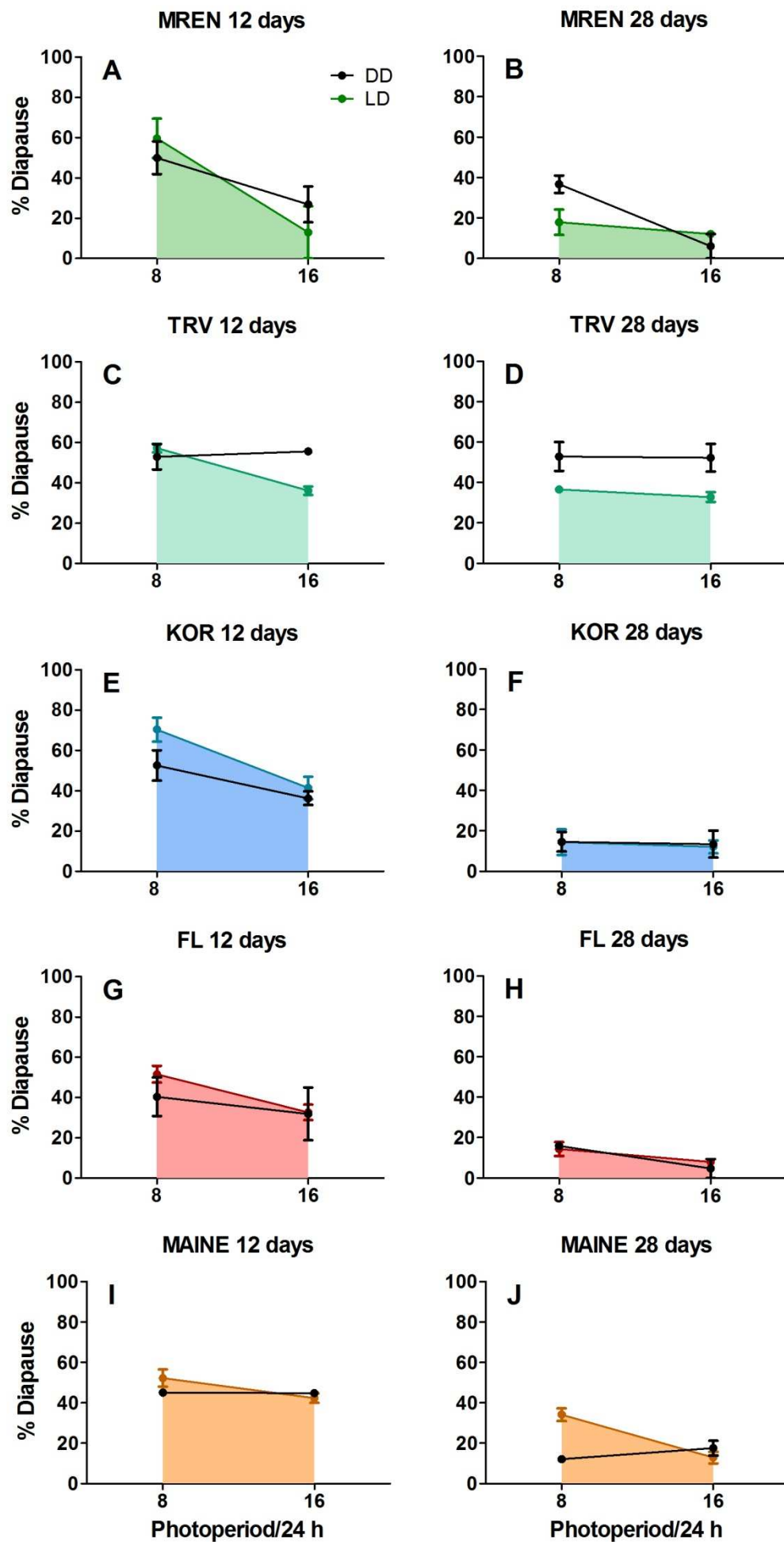
Figure 8.3

Diapause of the individual populations used in the experiment described in Paragraph 3.2 of Chapter 6. Flies were dissected after 12 and 28 days at 12°C. The black data points represent the data regarding DD controls. A and B) Population from Rende (Italy), at 12 and 28 days. C and D) Population from Treviso (Italy) at 12 and 28 days. E and F) Population from Korpilahti (Finland) at 12 and 28 days. G and H) Population from Homestead (Florida) at 12 and 28 days. I and J) Populations from Bowdoinham (Maine) at 12 and 28 days.

8.4.3 DIAPAUSE EVOLUTIONARY AGE

Table 8.4 shows the details of the lines used in Paragraph 6.3.3 of Chapter 6, whereas Figure 8.4 shows the levels of diapause of the individual populations with the data regarding the constant darkness controls (DD).

LINE	SPECIES	LOCATION	N	LAT	LONG	ALT	DIAPAUSE RESULTS 12 days			DIAPAUSE RESULTS 28 days		
							FLIES	8:16	16:8	FLIES	8:16	16:8
MREN	*D. mel*	Rende (Italy)	25	39.20	16.11	480	419 (10, 6)	49.97 (8.09)	26.90 (8.84)	403 (4, 6)	17.88 (6.28)	12.06 (7.67)
							196 (5, 2)	59.63 (9.77)	12.96 (12.96)	134 (2, 2)	36.73 (4.30)	6.00 (6.00)
KN	*D. mel*	Nyahururu (Kenya)	30	0.04	36.37	2360	479 (6, 6)	43.68 (8.64)	21.10 (2.09)	415 (6, 6)	7.78 (2.16)	5.69 (3.06)
							128 (2, 2)	21.50 (11.83)	24.82 (12.32)	132 (2, 2)	3.10 (0.24)	0.00 (0.00)
SREN	*D. sim*	Rende (Italy)	25	39.20	16.11	480	403 (7, 6)	54.57 (6.00)	29.63 (6.58)	350 (4, 6)	12.43 (7.22)	8.05 (1.92)
							121 (2, 2)	62.80 (8.04)	54.67 (1.33)	110 (2, 2)	7.03 (0.97)	7.81 (7.81)
simAFR	*D. sim*	Uganda (5) Tanzania (3) Kenya (5) Malawi (2)	15	-15.38 to 0.46	30.40 to 36.67	n.a.	393 (6, 6)	16.54 (3.75)	16.34 (3.63)	400 (6, 6)	0.46 (0.46)	0.00 (0.00)
							130 (2, 2)	23.74 (4.39)	26.38 (10.76)	147 (2, 2)	6.06 (6.06)	6.41 (3.85)

Table 8.4

Details of the *D. melanogaster* and *D. simulans* lines from Europe and Africa used in the diapause experiment. The data regarding MREN are the same as the ones described before (Figure 6.6 and 8.3, and Figure 8.3). They are reported here too in order to allow easy comparison with the other lines. N: number of isofemale lines used to create the populations analysed. Lat: latitude in degrees north. Long: longitude in degrees (Negative values: west; Positive values: east). Alt: altitude in meters above sea level. n.a.: not available. Flies: number of flies analysed per condition, in brackets the number of replicates performed in the two photoperiods respectively. 8:16 and 16:8 indicate the two photoperiods at which the experiment was performed. Values are averaged percentage of diapause (±SEM). Shaded cells indicate results obtained in DD conditions.
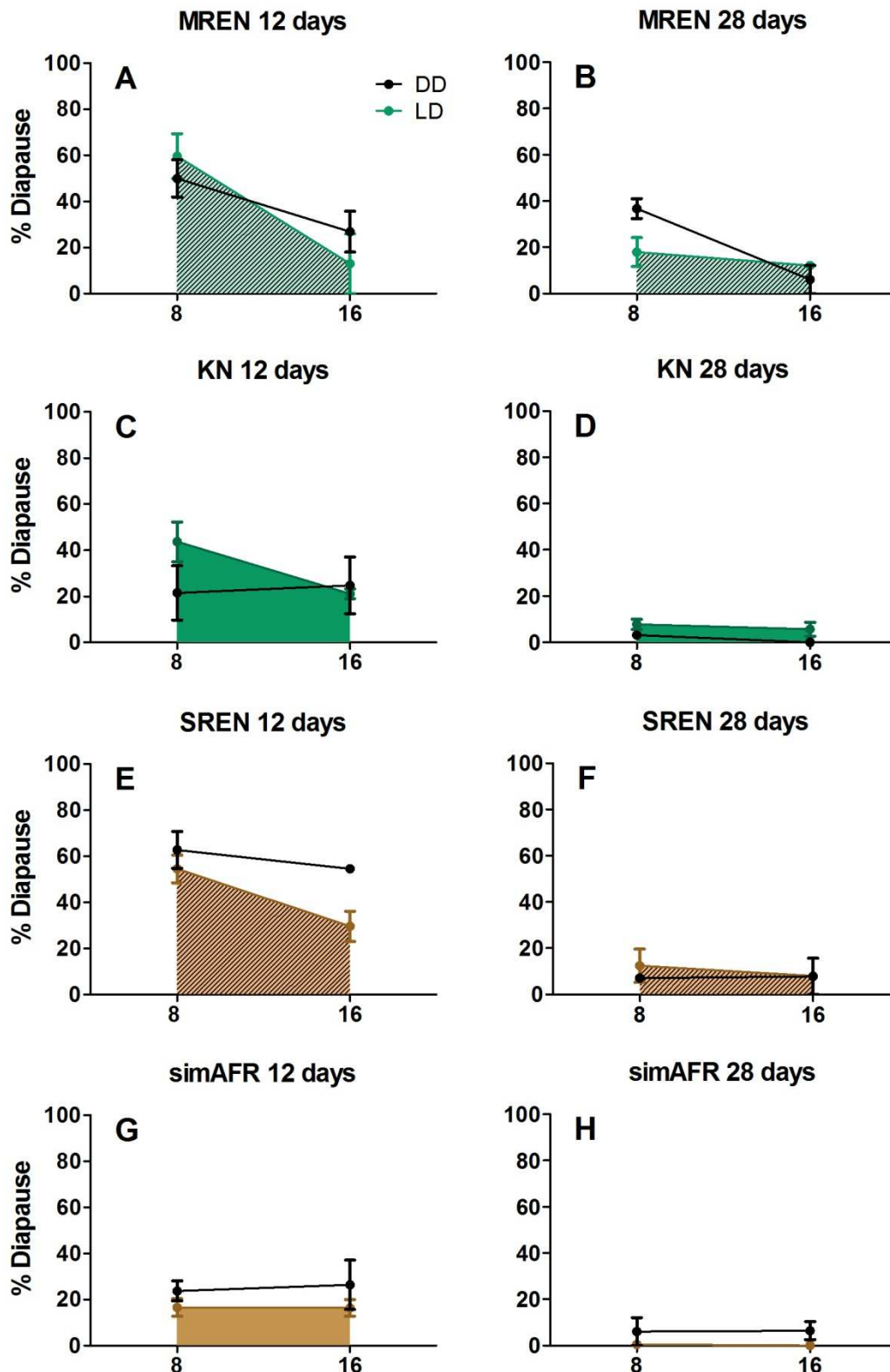
Figure 8.4

Diapause of the individual populations used in the experiment described in Paragraph 3.3 of Chapter 6. Flies were dissected after 12 and 28 days at 12°C. The black data points represent the data regarding DD controls. A and B) *D. melanogaster* population from Rende (Italy), at 12 and 28 days. C and D) *D. melanogaster* population from Kenya at 12 and 28 days. E and F) *D. simulans* population from Rende (Italy) at 12 and 28 days. G and H) composit *D. simulans* population from Africa at 12 and 28 days.

REFERENCES

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

Ashburner, M., Golic, K.G. and Hawley, R.S., 1989. *Drosophila*. Cold spring harbor laboratory press.

Azevedo, R.B.R., French, V., Partridge, L., 1996. Thermal evolution of egg size in *Drosophila melanogaster*. Evolution. 2338-2345.

Belgacem, Y.H. & Martin, J.R., 2006. Disruption of insulin pathways alters trehalose level and abolishes sexual dimorphism in locomotor activity in *Drosophila*. Journal of Neurobiology. 66, 19-32.

Bellen, H.J., Kooyer, S., D'Evelyn, D., Pearlman, J., 1992a. The *Drosophila* couch potato protein is expressed in nuclei of peripheral neuronal precursors and shows homology to RNA-binding proteins. Genes & Development. 6, 2125-2136.

Bellen, H.J., Vaessin, H., Bier, E., Kolodkin, A., D'Evelyn, D., Kooyer, S., Jan, Y.N., 1992b. The *Drosophila couch potato gene*: an essential gene required for normal adult behavior. Genetics. 131, 365-375.

Berry, A. & Kreitman, M., 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. Genetics. 134, 869-893.

Boulétreau, J., 1978. Ovarian activity and reproductive potential in a natural population of *Drosophila melanogaster*. Oecologia. 35, 319-342.

Boulétreau-Merle, J., Fouillet, P., Terrier, O., 1992. Clinal and seasonal variations in initial retention capacity of virgin *Drosophila melanogaster* females as a strategy for fitness. Evolutionary Ecology. 6, 223-242.

Bradshaw, W.E. & Holzapfel, C.M., 2001. Genetic shift in photoperiodic response correlated with global warming. Proceedings of the National Academy of Sciences. 98, 14509.

Broughton, S.J., Piper, M.D.W., Ikeya, T., Bass, T.M., Jacobson, J., Driege, Y., Martinez, P., Hafen, E., Withers, D.J., Leevers, S.J., 2005. Longer lifespan, altered metabolism, and stress resistance in *Drosophila* from ablation of cells making insulin-like ligands. Proceedings of the National Academy of Sciences of the United States of America. 102, 3105.

Cao, C. & Brown, M.R., 2001. Localization of an insulin-like peptide in brains of two flies. Cell and Tissue Research. 304, 317-321.

Caracristi, G. & Schlötterer, C., 2003. Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. Molecular Biology and Evolution. 20, 792.

Carney, G.E. & Bender, M., 2000. The *Drosophila ecdysone receptor* (*EcR*) gene is required maternally for normal oogenesis. Genetics. 154, 1203-1211.

Chown, S.L. & Gaston, K.J., 2009. Body size variation in insects: a macroecological perspective. Biological Reviews. 85, 139-169.

Claeys, I., Simonet, G., Poels, J., Van Loy, T., Vercammen, L., De Loof, A., Vanden Broeck, J., 2002. Insulin-related peptides and their conserved signal transduction pathway. Peptides. 23, 807-816.

Clancy, D.J., Gems, D., Harshman, L.G., Oldham, S., Stocker, H., Hafen, E., Leevers, S.J., Partridge, L., 2001. Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. Science's STKE. 292, 104.

Costa, R., Peixoto, A.A., Barbujani, G., Kyriacou, C.P., 1992. A latitudinal cline in a *Drosophila* clock gene. Proceedings of the Royal Society of London.Series B: Biological Sciences. 250, 43-49.

Coyne, J.A., Boussy, I.A., Prout, T., Bryant, S.H., Jones, J., Moore, J.A., 1982. Long-distance migration of *Drosophila*. The American Naturalist. 119, 589-595.

David, J. & Mercot, H., 1986. Alcohol tolerance and *Adh* gene frequencies in European and African populations of *Drosophila melanogaster*.

David, J., Alonso-Moraga, A., Borai, F., Capy, P., Mercot, H., McEvey, S., Munoz-Serrano, A., Tsakas, S., 1989. Latitudinal variation of *Adh* gene frequencies in *Drosophila melanogaster*: a Mediterranean instability. Heredity. 62, 11-16.

David, J.R. & Capy, P., 1988. Genetic variation of *Drosophila melanogaster* natural populations. Trends in Genetics. 4, 106-111.

Denlinger, D.L., 1972. Induction and termination of pupal diapause in *Sarcophaga* (Diptera: Sarcophagidae). The Biological Bulletin. 142, 11-24.

Denlinger, D.L., 1986. Dormancy in tropical insects. Annual Review of Entomology. 31, 239-264.

Duman, J., Bennett, V., Sformo, T., Hochstrasser, R., Barnes, B., 2004. Antifreeze proteins in Alaskan insects and spiders. Journal of Insect Physiology. 50, 259-266.

Emerson, K.J., Bradshaw, W.E., Holzapfel, C.M., 2009a. Complications of complexity: integrating environmental, genetic and hormonal control of insect diapause. Trends in Genetics. 25, 217-225.

Emerson, K.J., Uyemura, A.M., McDaniel, K.L., Schmidt, P.S., Bradshaw, W.E., Holzapfel, C.M., 2009b. Environmental control of ovarian dormancy in natural populations of *Drosophila melanogaster*. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology. 195, 825-829.

Fiston-Lavier, A.S., Singh, N.D., Lipatov, M., Petrov, D.A., 2010. *Drosophila melanogaster* recombination rate calculator. Gene. 463, 18-20.

Friedman, D.B. & Johnson, T., 1988. A mutation in the *age-1* gene in *Caenorhabditis elegans* lengthens life and reduces hermaphrodite fertility. Genetics. 118, 75.

Frydenberg, J., Hoffmann, A., Loeschcke, V., 2003. DNA sequence variation and latitudinal associations in *hsp23, hsp26* and *hsp27* from natural populations of *Drosophila melanogaster*. Molecular Ecology. 12, 2025-2032.

Fu, Y.X. & Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics. 133, 693.

Garen, A., Kauvar, L., Lepesant, J.A., 1977. Roles of ecdysone in *Drosophila* development. Proceedings of the National Academy of Sciences. 74, 5099.

Garner, W. & Allard, H., 1922. Photoperiodism, the response of the plant to relative length of day and night. Science. 55, 582-583.

Giannakou, M.E. & Partridge, L., 2007. Role of insulin-like signalling in *Drosophila* lifespan. Trends in Biochemical Sciences. 32, 180-188.

Glasscock, E. & Tanouye, M.A., 2005. *Drosophila couch potato* mutants exhibit complex neurological abnormalities including epilepsy phenotypes. Genetics. 169, 2137-2149.

Gloor, G.B., Preston, C.R., Johnson-Schlitz, D.M., Nassif, N.A., Phillis, R.W., Benz, W.K., Robertson, H.M., Engels, W.R., 1993. Type I repressors of P element mobility. Genetics. 135, 81-95.

Gockel, J., Kennington, W.J., Hoffmann, A., Goldstein, D.B., Partridge, L., 2001. Nonclinality of molecular variation implicates selection in maintaining a morphological cline of *Drosophila melanogaster*. Genetics. 158, 319-323.

Griffiths, R. & Tavaré, S., 1999. The ages of mutations in gene trees. Annals of Applied Probability. 567-590.

Hallas, R., Schiffer, M., Hoffmann, A.A., 2002. Clinal variation in *Drosophila serrata* for stress resistance and body size. Genetical Research. 79, 141-148.

Hardin, P.E., 2005. The Circadian Timekeeping System of *Drosophila*. Current Biology. 15, R714-R722.

Harvie, P.D., Filippova, M., Bryant, P.J., 1998. Genes expressed in the ring gland, the major endocrine organ of *Drosophila melanogaster*. Genetics. 149, 217.

Heinstra, P.W.H., 1993. Evolutionary genetics of the *Drosophila alcohol dehydrogenase* gene-enzyme system. Genetica. 92, 1-22.

Hoffmann, A.A. & Weeks, A.R., 2007. Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. Genetica. 129, 133-147.

Hoffmann, A.A., Anderson, A., Hallas, R., 2002. Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. Ecology Letters. 5, 614-618.

Hoffmann, A.A., Hallas, R., Sinclair, C., Mitrovski, P., 2001. Levels of variation in stress resistance in *Drosophila* among strains, local populations, and geographic regions: patterns for desiccation, starvation, cold resistance, and associated traits. Evolution. 55, 1621-1630.

Hudson, R.R., Kreitman, M., Aguade, M., 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116, 153.

Hwangbo, D.S., Gersham, B., Tu, M.P., Palmer, M., Tatar, M., 2004. *Drosophila* dFOXO controls lifespan and regulates insulin signalling in brain and fat body. Nature. 429, 562-566.

Ingram, C.J.E., Mulcare, C.A., Itan, Y., Thomas, M.G., Swallow, D.M., 2009. Lactose digestion and the evolutionary genetics of lactase persistence. Human Genetics. 124, 579-591.

Itan, Y., Powell, A., Beaumont, M.A., Burger, J., Thomas, M.G., 2009. The origins of lactase persistence in Europe. PLoS Computational Biology. 5, e1000491.

Izquierdo, J., 1991. How does *Drosophila melanogaster* overwinter? Entomologia Experimentalis Et Applicata. 59, 51-58.

Jablonski, N.G. & Chaplin, G., 2010. Human skin pigmentation as an adaptation to UV radiation. Proceedings of the National Academy of Sciences. 107, 8962-8968.

James, A.C., Azevedo, R.B.R., Partridge, L., 1997. Genetic and environmental responses to temperature of *Drosophila melanogaster* from a latitudinal cline. Genetics. 146, 881-890.

Kalushkov, P., Hodkova, M., Nedved, O., Hodek, I., 2001. Effect of thermoperiod on diapause intensity in *Pyrrhocoris apterus* (*Heteroptera Pyrrhocoridae*). Journal of Insect Physiology. 47, 55-61.

Kankare, M., Salminen, T., Laiho, A., Vesala, L., Hoikkala, A., 2010. Changes in gene expression linked with adult reproductive diapause in a northern malt fly species: a candidate gene microarray study. BMC Ecology. 10, 3.

Kankare, M., Salminen, T.S., Lampinen, H., Hoikkala, A., 2011. Sequence variation in *couch potato* and its effects on life-history traits in a northern malt fly, *Drosophila montana*. Journal of Insect Physiology.

Keller, A., 2007. *Drosophila melanogaster*'s history as a human commensal. Current Biology. 17, R77-R81.

Kelty, J.D. & Lee Jr, R.E., 2001. Rapid cold-hardening of *Drosophila melanogaster* (Diptera: *Drosophiladae*) during ecologically based thermoperiodic cycles. Journal of Experimental Biology. 204, 1659-1666.

Kenyon, C., 2005. The plasticity of aging: insights from long-lived mutants. Cell. 120, 449-460.

Kevan, S.M., 1980. Perspectives on season of suicide: A review. Social Science & Medicine.Part D: Medical Geography. 14, 369-378.

Kimura, K.D., Tissenbaum, H.A., Liu, Y., Ruvkun, G., 1997. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans.* Science. 277, 942.

Knibb, W., 1982. Chromosome inversion polymorphisms in *Drosophila melanogaster.* II. Geographic clines and climatic associations in Australasia, North America and Asia. Genetica. 58, 213-221.

Knibb, W., Oakeshott, J., Gibson, J., 1981. Chromosome inversion polymorphisms in *Drosophila melanogaster*. I. Latitudinal clines and associations between inversions in Australasian populations. Genetics. 98, 833.

Konopka, R.J. & Benzer, S., 1971. Clock mutants of *Drosophila melanogaster*. Proceedings of the National Academy of Sciences. 68, 2112.

Kostál, V., 2006. Eco-physiological phases of insect diapause. Journal of Insect Physiology. 52, 113-127.

Kovanen, L., Saarikoski, S.T., Aromaa, A., Lönnqvist, J., Partonen, T., 2010. ARNTL (BMAL1) and NPAS2 gene variants contribute to fertility and seasonality. PLoS One. 5, e10007.

Kozlova, T. & Thummel, C.S., 2000. Steroid Regulation of Postembryonic Development and Reproduction in *Drosophila*. Trends in Endocrinology & Metabolism. 11, 276-280.

Kreitman, M. & Hudson, R.R., 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics. 127, 565-582.

Lee, S.F., Sgro, C.M., Shirriffs, J., Wee, C.W., Rako, L., van Heerwaarden, B., Hoffmann, A.A., 2011. Polymorphism in the *couch potato* gene clines in eastern Australia but is not associated with ovarian dormancy in *Drosophila melanogaster*. Molecular Ecology. 20, 2973-2984.

Leonard, D.E., 1968. Diapause in the Gypsy Moth1. Journal of Economic Entomology. 61, 596-598.

Li, H. & Fu, Y., 2004. Neutrality Test, a novel software for performing tests of neutrality. Bioinformatics.

Li, H. & Stephan, W., 2006. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. PLoS Genetics. 2, e166.

Li, H., & Fu, Y. (2004). Neutrality test, a novel software for performing tests of neutrality. Bioinformatics. In preparation.

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

Lin, C. & Shalitin, D., 2003. Cryptochrome structure and signal transduction. Annual Review of Plant Biology. 54, 469-496.

Liu, Q., Thorland, E.C., Heit, J.A., Sommer, S.S., 1997. Overlapping PCR for bidirectional PCR amplification of specific alleles: a rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes. Genome Research. 7, 389-398.

Lumme, J. & Oikarinen, A., 1977. The genetic basis of the geographically variable photoperiodic diapause in *Drosophila littoralis*. Hereditas. 86, 129-141.

Macdonald, S.J., Pastinen, T., Long, A.D., 2005. The effect of polymorphisms in the enhancer of split gene complex on bristle number variation in a large wild-caught cohort of *Drosophila melanogaster*. Genetics. 171, 1741-1756.

Masaki, S., 1980. Summer diapause. Annual Review of Entomology. 25, 1-25.

McClung, C.R. & Davis, S.J., 2010. Ambient thermometers in plants: from physiological outputs towards mechanisms of thermal sensing. Current Biology. 20, R1086-R1092.

Mitrovski, P. & Hoffmann, A.A., 2001. Postponed reproduction as an adaptation to winter conditions in *Drosophila melanogaster*: evidence for clinal variation under semi-natural conditions. Proceedings of the Royal Society of London.Series B: Biological Sciences. 268, 2163-2168.

Miyashita, N.T., Aguadé, M., Langley, C.H., 1993. Linkage disequilibrium in the white locus region of *Drosophila melanogaster*. Genetical Research. 62, 101-101.

Molin, J., Mellerup, E., Bolwig, T., Scheike, T., Dam, H., 1996. The influence of climate on development of winter depression. Journal of Affective Disorders. 37, 151-155.

Morris, J.Z., Tissenbaum, H.A., Ruvkun, G., 1996. A phosphatidylinositol-3-OH kinase family member regulating longevity and diapause in *Caenorhabditis elegans*.

Mullen, L.M. & Hoekstra, H.E., 2008. Natural selection along an environmental gradient: a classic cline in mouse pigmentation. Evolution. 62, 1555-1570.

Myers, M.P., Wager-Smith, K., Wesley, C.S., Young, M.W., Sehgal, A., 1995. Positional cloning and sequence analysis of the *Drosophila* clock gene, *timeless*. Science. 270, 805-808.

Naya, D.E., Bozinovic, F., Karasov, W.H., 2008. Latitudinal trends in digestive flexibility: testing the climatic variability hypothesis with data on the intestinal length of rodents. The American Naturalist. 172, E122-E134.

Nunes, M.D.S., Neumeier, H., Schloetterer, C., 2008. Contrasting patterns of natural variation in global *Drosophila melanogaster* populations. Molecular Ecology. 17, 4470-4479.

Nunes, M.V. & Saunders, D., 1999. Photoperiodic time measurement in insects: a review of clock models. Journal of Biological Rhythms. 14, 84-104.

Oakeshott, J., Gibson, J., Anderson, P., Knibb, W., Anderson, D., Chambers, G., 1982. *Alcohol dehydrogenase* and *glycerol-3-phosphate dehydrogenase* clines in *Drosophila melanogaster* on different continents. Evolution. 86-96.

Ogienko, A., Fedorova, S., Baricheva, E., 2007. Basic aspects of ovarian development in *Drosophila melanogaster*. Russian Journal of Genetics. 43, 1120-1134.

Paaby, A.B., Blancket, M.J., Hoffmann, A.R.Y.A., Schmidt, P.S., 2010. Identification of a candidate adaptive polymorphism for *Drosophila* life history by parallel independent clines on two continents. Molecular Ecology. 19, 760-774.

Partonen, T., Treutlein, J., Alpman, A., Frank, J., Johansson, C., Depner, M., Aron, L., Rietschel, M., Wellek, S., Soronen, P., 2007. Three circadian clock genes *Per2*, *Arntl*, and *Npas2* contribute to winter depression. Annals of Medicine. 39, 229-238.

Partridge, L. & Gems, D., 2002. The evolution of longevity. Curr Biol. 12, R544-R546.

Partridge, L., Gems, D., Withers, D.J., 2005. Sex and death: what is the connection? Cell. 120, 461-472.

Peschel, N., Chen, K.F., Szabo, G., Stanewsky, R., 2009. Light-dependent interactions between the *Drosophila* circadian clock factors *cryptochrome*, *jetlag*, and *timeless*. Current Biology. 19, 241-247.

Pittendrigh, C.S. & Takamura, T., 1987. Temperature dependence and evolutionary adjustment of critical night length in insect photoperiodism. Proceedings of the National Academy of Sciences. 84, 7169.

Postlethwait, J.H. & Weiser, K., 1973. Vitellogenesis induced by juvenile hormone in the female sterile mutant apterous-four in *Drosophila melanogaster*. Nature. 244, 284-285.

Price, J.L., Blau, J., Rothenfluh, A., Abodeely, M., Kloss, B., Young, M.W., 1998. *double-time* is a novel *Drosophila* clock gene that regulates PERIOD Protein Accumulation. Cell. 94, 83-95.

Ragland, G.J., Denlinger, D.L., Hahn, D.A., 2010. Mechanisms of suspended animation are revealed by transcript profiling of diapause in the flesh fly. Proceedings of the National Academy of Sciences. 107, 14909-14914.

Richard, D.S., Jones, J.M., Barbarito, M.R., Cerula, S., Detweiler, J.P., Fisher, S.J., Brannigan, D.M., Scheswohl, D.M., 2001. Vitellogenesis in diapausing and mutant *Drosophila melanogaster*: further evidence for the relative roles of ecdysteroids and juvenile hormones. Journal of Insect Physiology. 47, 905-913.

Richard, D.S., Watkins, N.L., Serafin, R.B., Gilbert, L.I., 1998. Ecdysteroids regulate yolk protein uptake by *Drosophila melanogaster* oocytes. Journal of Insect Physiology. 44, 637-644.

Rosato, E., Peixoto, A.A., Costa, R., Kyriacou, C.P., 1997a. Linkage disequilibrium, mutational analysis and natural selection in the repetitive region of the clock gene, *period*, in *Drosophila melanogaster*. Genetical Research. 69, 89-99.

Rosato, E., Trevisan, A., Sandrelli, F., Zordan, M., Kyriacou, C.P., Costa, R., 1997b. Conceptual translation of timeless reveals alternative initiating methionines in *Drosophila.* Nucleic Acids Research. 25, 455.

Rosen, L.N., Targum, S.D., Terman, M., Bryant, M.J., Hoffman, H., Kasper, S.F., Hamovit, J.R., Docherty, J.P., Welch, B., Rosenthal, N.E., 1990. Prevalence of seasonal affective disorder at four latitudes. Psychiatry Research. 31, 131-144.

Rosenthal, N.E., Sack, D.A., Gillin, J.C., Lewy, A.J., Goodwin, F.K., Davenport, Y., Mueller, P.S., Newsome, D.A., Wehr, T.A., 1984. Seasonal affective disorder: a description of the syndrome and preliminary findings with light therapy. Archives of General Psychiatry. 41, 72.

Rozen, S. & Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132, 365-386.

Russo, C., Takezaki, N., Nei, M., 1995. Molecular phylogeny and divergence times of drosophilid species. Molecular Biology and Evolution. 12, 391-404.

Sandrelli, F., Tauber, E., Pegoraro, M., Mazzotta, G., Cisotto, P., Landskron, J., Stanewsky, R., Piccin, A., Rosato, E., Zordan, M., 2007. A molecular basis for natural selection at the *timeless* locus in *Drosophila melanogaster*. Science. 316, 1898.

Saunders, D., 1973. Thermoperiodic control of diapause in an insect: theory of internal coincidence. Science. 181, 358-360.

Saunders, D., 2005. Erwin Bünning and Tony Lees, two giants of chronobiology, and the problem of time measurement in insect photoperiodism. Journal of Insect Physiology. 51, 599-608.

Saunders, D., Lewis, R., Warman, G., 2004. Photoperiodic induction of diapause: opening the black box. Physiological Entomology. 29, 1-15.

Saunders, D., Richard, D., Applebaum, S., Ma, M., Gilbert, L., 1990. Photoperiodic diapause in *Drosophila melanogaster* involves a block to the juvenile hormone regulation of ovarian maturation. General and Comparative Endocrinology. 79, 174-184.

Saunders, D.S., 2002. Insect clocks. Elsevier Science.

Saunders, D.S., Henrich, V.C., Gilbert, L.I., 1989. Induction of diapause in *Drosophila melanogaster*: photoperiodic regulation and the impact of arrhythmic clock mutations on time measurement. Proceedings of the National Academy of Sciences. 86, 3748.

Sawyer, L.A., Sandrelli, F., Pasetto, C., Peixoto, A.A., Rosato, E., Costa, R., Kyriacou, C.P., 2006. The *period* gene Thr-Gly polymorphism in Australian and African *Drosophila melanogaster* populations: implications for selection. Genetics. 174, 465.

Schmidt, P.S., Matzkin, L., Ippolito, M., Eanes, W.F., 2005a. Geographic variation in diapause incidence, life-history traits, and climatic adaptation in *Drosophila melanogaster*. Evolution. 59, 1721-1732.

Schmidt, P.S., Paaby, A.B., Heschel, M.S., 2005b. Genetic variance for diapause expression and associated life histories in *Drosophila melanogaster*. Evolution. 59, 2616-2625.

Schmidt, P.S., Zhu, C.T., Das, J., Batavia, M., Yang, L., Eanes, W.F., 2008. An amino acid polymorphism in the *couch potato* gene forms the basis for climatic adaptation in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America. 105, 16207-16211.

Sehgal, A., Price, J.L., Man, B., Young, M.W., 1994. Loss of circadian behavioral rhythms and per RNA oscillations in the *Drosophila* mutant *timeless*. Science. 263, 1603-1606.

Sehgal, A., Rothenfluh-Hilfiker, A., Hunter-Ensor, M., Chen, Y., Myers, M.P., Young, M.W., 1995. Rhythmic expression of timeless: a basis for promoting circadian cycles in period gene autoregulation. Science. 270, 808-810.

Sezgin, E., Duvernell, D.D., Matzkin, L.M., Duan, Y., Zhu, C.T., Verrelli, B.C., Eanes, W.F., 2004. Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. Genetics. 168, 923.

Sgrò, C.M., Magiafoglou, A., Faine, L., Hoffmann, A.A., 2006. Absence of clinal variation in virgin retention capacity in Australian *Drosophila melanogaster*. Evolutionary Ecology. 20, 407-413.

Siegmund, T. & Korge, G., 2001. Innervation of the ring gland of *Drosophila melanogaster*. The Journal of Comparative Neurology. 431, 481-491.

Slatkin, M. & Rannala, B., 2000. Estimating allele age. Annual Review of Genomics and Human Genetics. 1, 225-249.

Stewart, K.T., Gaddy, J.R., Byrne, B., Miller, S., Brainard, G.C., 1991. Effects of green or white light for treatment of seasonal depression. Psychiatry Research. 38, 261-270.

Streisinger, G., et al, 1966. Frameshift mutations and the genetic code, Cold Spring Harbor Symposia on Quantitative Biology, 1966, Cold Spring Harbor Laboratory Press pp77-84.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123, 585.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution. 24, 1596-1599.

Tatar, M. & Yin, C.M., 2001. Slow aging during insect reproductive diapause: why butterflies, grasshoppers and flies are like worms. Experimental Gerontology. 36, 723-738.

Tatar, M., Bartke, A., Antebi, A., 2003. The endocrine regulation of aging by insulin-like signals. Science. 299, 1346.

Tatar, M., Chien, S.A., Priest, N.K., 2001a. Negligible senescence during reproductive dormancy in *Drosophila melanogaster*. The American Naturalist. 158, 248-258.

Tatar, M., Kopelman, A., Epstein, D., Tu, M.P., Yin, C.M., Garofalo, R., 2001b. A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. Science. 292, 107.

Tauber, E., Zordan, M., Sandrelli, F., Pegoraro, M., Osterwalder, N., Breda, C., Daga, A., Selmin, A., Monger, K., Benna, C., Rosato, E., Kyriacou, C.P., Costa, R., 2007. Natural selection favors a newly derived *timeless* allele in *Drosophila melanogaster*. Science (New York, N.Y.). 316, 1895-1898.

Tauber, M.J., Tauber, C.A. and Masaki, S., 1986. Seasonal adaptations of insects. Oxford University Press, USA.

Teets, N.M., Elnitsky, M.A., Benoit, J.B., Lopez-Martinez, G., Denlinger, D.L., Lee Jr, R.E., 2008. Rapid cold-hardening in larvae of the Antarctic midge Belgica antarctica: cellular cold-sensing and a role for calcium. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 294, R1938-R1946.

Tu, M.P., Epstein, D., Tatar, M., 2002a. The demography of slow aging in male and female *Drosophila* mutant for the insulin-receptor substrate homologue *chico*. Aging Cell. 1, 75-80.

Tu, M.P., Yin, C.M., Tatar, M., 2002b. Impaired ovarian ecdysone synthesis of *Drosophila melanogaster* insulin receptor mutants. Aging Cell. 1, 158-160.

Umina, P., Weeks, A., Kearney, M., McKechnie, S., Hoffmann, A., 2005. A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. Science. 308, 691.

Umina, P., Weeks, A., Kearney, M., McKechnie, S., Hoffmann, A., 2005. A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. Science. 308, 691.

Vigue, C. & Johnson, F., 1973. Isozyme variability in species of the genus *Drosophila*. VI. Frequency-property-environment relationships of allelic alcohol dehydrogenases in *D. melanogaster*. Biochemical Genetics. 9, 213-227.

Wilczek, A., Burghardt, L., Cobb, A., Cooper, M., Welch, S., Schmitt, J., 2010. Genetic and physiological bases for phenological responses to current and predicted climates. Philosophical Transactions of the Royal Society B: Biological Sciences. 365, 3129-3147.

Williams, K.D. & Sokolowski, M.B., 1993. Diapause in *Drosophila melanogaster* females: a genetic analysis. York University.

Williams, K.D., Busto, M., Suster, M.L., So, A.K.C., Ben-Shahar, Y., Leevers, S.J., Sokolowski, M.B., 2006. Natural variation in *Drosophila melanogaster* diapause due to the insulin-regulated *PI3-kinase.* Proceedings of the National Academy of Sciences. 103, 15911.

Wulbeck, C. & Helfrich-Forster, C., 2007. RNA in situ hybridizations on *Drosophila* whole mounts. Methods in Molecular Biology-Clifton then Totowa-. 362, 495.

Yamashita, O., 1996. Diapause hormone of the silkworm, *Bombyx mori*: structure, gene expression and function. Journal of Insect Physiology. 42, 669-679.

Yenush, L., Fernandez, R., Myers Jr, M.G., Grammer, T.C., Sun, X.J., Blenis, J., Pierce, J.H., Schlessinger, J., White, M.F., 1996. The *Drosophila* insulin receptor activates multiple signaling pathways but requires insulin receptor substrate proteins for DNA synthesis. Molecular and Cellular Biology. 16, 2509-2517.

Zhang, Q. & Denlinger, D.L., 2011. Elevated *couch potato* transcripts associated with adult diapause in the mosquito *Culex pipiens*. Journal of Insect Physiology.