Universidade de Aveiro
2012
Departamento de Biologia

Helder Henrique Ribeiro Rocha

A base genética da longevidade em *Drosophila americana*

The genetic basis of lifespan in *Drosophila americana*

Universidade de Aveiro
2012

Departamento de Biologia

Helder Henrique Ribeiro Rocha

A base genética da longevidade em *Drosophila americana*

The genetic basis of lifespan in *Drosophila americana*

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica do Doutor Jorge Vieira, Investigador Principal no grupo de Evolução Molecular do Instituto de Biologia Molecular e Celular da Universidade do Porto e coorientação do Professor Doutor Manuel Santos, Professor Associado do Departamento de Biologia da Universidade de Aveiro

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

POPH PROGRAMA OPERACIONAL POTENCIAL HUMANO
QREN QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL PORTUGAL 2007.2013
UNIÃO EUROPEIA Fundo Social Europeu

Este trabalho foi financiado pelo FEDER através do Programa Operacional Potencial Humano – POPH e pela FCT – Fundação para a Ciência e a Tecnologia através do projeto: FCOMP-01-0124-FEDER-008916 (PTDC/BIA-BEC/099933/2008).

Em memória do meu Pai, sempre presente…

A prova do êxito da educação não está naquilo que se sabe depois do último exame, quando se deixa a escola, mas sim naquilo que se faz dez anos mais tarde

Sir Robert Baden-Powell

o júri

presidente

Professora Doutora Maria de Lourdes Gomes Pereira
Professora Associada com Agregação, Universidade de Aveiro

Doutor Rui Miguel Macieira de Faria
Estagiário de Pós-Doutoramento, Centro de Investigação em Biodiversidade e Recursos Genéticos

Doutor Jorge Manuel de Sousa Basto Vieira
Investigador Principal, Instituto de Biologia Molecular e Celular

Professor Doutor Manuel António da Silva Santos
Professor Associado, Universidade de Aveiro

agradecimentos

Ao finalizar mais esta etapa do meu percurso não podia deixar de manifestar o meu grande apreço e agradecimento sincero a todos aqueles que tornaram possível este trabalho.

Em primeiro lugar, o meu mais profundo agradecimento ao Doutor Jorge Vieira, meu Orientador, pela oportunidade e pela honra de poder trabalhar no seu laboratório. Muito Obrigado por me ter guiado ao longo deste período, pela confiança, pelos conselhos e pelos ensinamentos transmitidos. Agradeço também a revisão, o auxílio, as críticas e sugestões que contribuíram não só para a escrita mas para a realização desta Dissertação.

Ao Professor Doutor Manuel Santos, por ter aceitado ser meu Coorientador, por todo o apoio e pela sua disponibilidade.

À Doutora Cristina Vieira, ao Ramiro, ao Bruno, à Ana, à Inês, à Mónica e a todas as pessoas que passaram pelo laboratório de Evolução Molecular do IBMC, pela experiência transmitida, pelo companheirismo, pela colaboração e pelo incentivo. Um agradecimento especial ao Micael, pela amizade, por toda a ajuda, pelos conselhos, as sugestões e as críticas. Agradeço-lhe a revisão cuidada do texto e as inúmeras discussões. Aprendi muito contigo!

Aos meus amigos e a todos os meus colegas, pela amizade e pelo companheirismo indispensáveis para o sucesso desta caminhada.

À Vanessa, aquele Obrigado especial…por tudo! Porque estás sempre comigo, apesar da distância, principalmente nos momentos mais complicados em que com uma palavra, um sorriso sempre me incentivaste a continuar. Pelo apoio incondicional, porque nunca deixaste de acreditar em mim, pela força que sempre me transmites e que me empurra a seguir em frente. Muito obrigado pela ajuda incansável, pela revisão cuidada e por todas as sugestões. Sem ti tudo seria muito mais difícil…

A toda a minha família, especialmente à minha mãe e irmãos, porque fizeram de mim aquilo que sou. Pelo apoio infindável, pela confiança infinita e por proporcionarem que tudo fosse possível. À Iara, pela alegria que transmite a cada dia! Sem vocês nada teria sido possível…

A todos, mais uma vez, o meu muito sincero OBRIGADO!

palavras-chave

Drosophila americana, longevidade, genes candidatos, estudo de associação, polimorfismo

resumo

A compreensão da base molecular da variação fenotípica observada dentro e entre espécies, como acontece na longevidade, é um dos principais objetivos da Biologia. Em *Drosophila melanogaster* foram já identificados mais de 50 genes que podem contribuir para a definição do tempo de vida. Contudo, na maioria dos estudos, as conclusões são baseadas em mutações criadas artificialmente nos genes alvo. Desta forma, é improvável que o tipo de variação observada exista em populações naturais, e como tal não se encontra totalmente esclarecido se variação nestes genes é capaz de explicar as diferenças de longevidade observadas na natureza. A única forma de compreender a arquitetura genética da longevidade passa por descobrir quais os *loci* que afetam a variação do tempo de vida na natureza.

A fim de abordar esta questão de forma abrangente, efetuámos cinco estudos de associação F_2 utilizando marcadores para 21 genes candidatos previamente identificados em *D. melanogaster*. Neste trabalho, utilizámos *Drosophila americana* em vez de *D. melanogaster* de forma a testar a generalidade das observações reportadas para esta última espécie.

Foram encontradas associações significativas entre polimorfismos em dez genes (*hep*, *dFOXO*, *filamin*, *dilp2*, *Cat SOD*, *PTEN*, *Dox-A2*, *Ddc* e *Lim3*) e o tempo de vida. Para o cruzamento F_2 envolvendo as duas estirpes (H5 e W11) cujo genoma já se encontra sequenciado, foi possível identificar possíveis diferenças aminoacídicas nas proteínas Lim3 e Hep que poderão ser responsáveis pelas variações observadas na longevidade. Os níveis de expressão relativa de todos os genes mostrando associações significativas foram também caracterizados nas estirpes H5 e W11 aos 0, 10, 30 e 60 dias. Foram observadas diferenças entre indivíduos “mais jovens” e “ mais idosos” em cinco genes (*dFOXO*, *filamin*, *dilp2*, *SOD* e *PTEN*) e, o que acontece em fases iniciais da vida parece influenciar a longevidade mais fortemente do que o que acontece mais tarde, o que é sugerido pelas diferenças observadas nos níveis de expressão dos genes *hep*, *SOD, PTEN* e *Lim3* e pela ausência de diferenças significativas nos níveis de expressão de todos os genes em indivíduos com 60 dias.

keywords

Drosophila americana, lifespan, candidate genes, association study, polymorphism

abstract

Understanding the molecular basis of within and between species phenotypic variation, such as lifespan differences, is one of the main goals of Biology. More than 50 genes have been identified in *Drosophila melanogaster* that may contribute to the setting of lifespan. Nevertheless, in most studies, conclusions are based upon mutations that were artificially induced in the target genes. Therefore, it is unlikely that the reported type of variation ever occurs in natural populations, and thus it is unclear whether naturally occurring variation at these genes is able to explain the observed changes in lifespan. The only way to understand the genetic architecture of longevity is to know what loci affect variation in lifespan in nature.

In order to comprehensively address this issue, we performed five F_2 association experiments using markers for 21 candidate genes previously identified in *D. melanogaster*. In this work we use *Drosophila americana* rather than *D. melanogaster* in order to also test the generality of the findings reported for the latter species.

Significant associations were found between polymorphisms at ten genes (*hep*, *dFOXO*, *filamin*, *dilp2*, *Cat*, *SOD*, *PTEN*, *Dox-A2*, *Ddc* and *Lim3*) and lifespan. For the F_2 cross involving the two strains (H5 and W11) whose genomes are already sequenced, we were able to identify putative amino acid differences at Lim3 and Hep that could be responsible for the observed changes in lifespan. For all genes showing associations with lifespan, their relative gene expression levels were also characterized at days 0, 10, 30 and 60 for strains H5 and W11. We found differences between “younger” and “older” flies in five genes (*dFOXO*, *filamin*, *dilp2*, *SOD* and *PTEN*), and what happens early in adult life may influence lifespan more strongly than what happens late in life as is suggested by the differences observed in *hep*, *SOD*, *PTEN* and *Lim3* expression levels in new-born flies and because of the absence of significant differences in all genes expression levels in 60 days-old flies.

INDEX

FIGURE INDEX

TABLE INDEX

LIST OF SYMBOLS AND ABBREVIATIONS

% – percentage

° C – degree Celsius

4E-BP – Eukaryotic translation initiation factor 4E – binding protein

AP – Antagonistic Pleiotropy theory

AP-1 – Activator protein 1

ATP – Adenosine-5'-triphosphate

BEST – Bayesian Estimation of Species Trees

BLASTn – Nucleotide-nucleotide Basic Local Alignment Search Tool

bp – base pairs

Bsk – Basket

Cat – *Catalase*

Catsup – *Cathecolamines up*

Cct1 – *CTP:phosphocholine cytidylyltransferase 1*

cDNA – complementary Deoxyribonucleic acid

CDS – Gene Coding Sequence

dATP – Deoxyadenosine triphosphate

dCTP – Deoxycytidine triphosphate

Ddc – *Dopa decarboxylase*

dGTP – Deoxyguanosine triphosphate

dilp – *Drosophila insulin-like peptide*

DNA – Deoxyribonucleic acid

dNTPs – deoxyribonucleotides trifosfate

Dox-A2 – *Diphenol oxidase A2*

DR – Dietary restricition

dS6K– *Drosophila* RPS6-p70-protein kinase

dTsc – *Drosophila Tuberous Sclerosis Complex*

DTT – Dithiothreitol

dTTP – Deoxythymidine triphosphate

EDTA – Ethylenediamine tetraacetic acid

EF-1α48D – *Elongation factor 1α48D*

elF4E – elongation factor 4E

F_0 – Parental generation

F_1 First recombinant generation

F_2 – Second recombinant generation

Fig. – Figure

FOXO – forkhead box, sub-group O

fwd *– four wheel drive*

g, µg, ng – Gram, microgram, nanogram

GWAS – Genome-wide Association Studies

h – hour

hep – *hemipterous*

HSP – Heat-shock proteins

IGF – Insulin Growth Factor

IIS – Insulin / Insulin Growth Factor signaling

Indy *– I'm not dead yet*

InR – Insulin-like Receptor

IPC – Insulin Producing Cells

JNK – c-Jun N-terminal kinase

JNKK – JNK-Activated Kinase

L, mL, µL – Liter, milliliter, microliter

LB – Lysogenic Broth medium

M, mM, µM – Molar, millimolar, micromolar

MA – Mutation Accumulation theory

MAPK – Mitogen-activated protein kinase

mei-41 *– meiotic 41*

min – minutes

MKK – Mitogen-activated protein kinase kinase

MNCs - Median Neurosecretory Cells

mNSCs – median Neural Stem Cells

mth **-** *methuselah*

My – Million years

PBS – Phosphate Buffered Saline

PDK1 – 3-Phosphoinositide-dependent protein kinase 1

PCR – Polymerase Chain Reaction

PI 3- Kinase Dp110/p60 – Phosphoinositide 3-kinase Dp110/p60

PKB / Akt – Protein kinase B

PTEN *– Phosphatase and tensin homolog*

puc *– puckered*

qRT-PCR – quantitative Reverse Transcription Polymerase Chain Reaction

QTL – Quantitative Trait Loci

RFLP – Restriction Fragment Length Polymorphism

RNA- Ribonucleic acid

ROL – Rate of Living theory

ROS – Reactive Oxygen Species

Rpl32 *– Ribosomal protein L32*

rpm – Revolutions per minute

Rpn3 *– Regulatory particle non-ATPase 3*

s – seconds

SNP – Single Nucleotide Polymorphism

SOD *– Superoxide Dismutase*

SPSS – Statistical Package for the Social Sciences

SSR – Simple Sequence Repeat

stc *– shuttle craft*

sug *– sugarbabe*

tBLASTn – Protein-nucleotide 6-frame translation Basic Local Alignment Search Tool

TOR – Target of Rapamycin

tup *– tailup*

U – Units

USA – United States of America

UV – ultraviolet

VhaSFD *– Vacuolar H^+-ATPase SFD subunit*

w/v – weight/volume

x g – acceleration relative to gravity (Relative Centrifugal Force - RCF)

1. INTRODUCTION

1.1 Historical perspective

Aging is one of the most complex biological processes determined both by genetic and environmental factors. Understanding the mechanisms underlying how organisms become old has been a topic of scientific interest for more than two millennia and still constitutes one of the greatest frontiers of science.

Already in Ancient Greece, Aristotle suggested that life is shortened as a consequence of sexual activity. He also observed that larger animals live longer and proposed that aging and death are linked to the dehydration process [1].

Only in the XIX century, and with the emergence of the industrial revolution and the observation that things break down the harder they work, Auguste Weismann proposed that aging and the death of the soma was a process analogue to "wear and tear" [1].

In early 1900s it was suggested that the rate of metabolism was increased as a function of body size, and therefore, larger animals also lived longer [2]. Later, on 1920s, Pearl developed the "rate of living" (ROL) theory proposing that life duration is a function of the genetically determined amount of energy consumed during adulthood (metabolic potential) and the rate of energy expenditure (metabolic rate). This theory was supported by the observation that the rate at which human males died after 45 years-old was related to their occupation, when accidents were excluded [3, 4].

However, in the 1950s a new idea was proposed that led into disregard of the ROL theory, despite the fact that there is a link between the two. The free-radical theory of aging advocates that aging and death are the consequence of the accumulation of oxidative damage to the cellular macromolecules [5]. Free-radicals are a natural by-product of metabolism that reacts with the macromolecules in the body causing damage. Animals have a defense system against this damage, however, despite this repair mechanism some damage always evades and leads to a progressive accumulation of oxidation that contributes to the functional decline and increase in mortality observed in late life [6, 7].

Over the last decades genetic and molecular approaches have allowed huge developments on the understanding of some of the greater mysteries of life, such as development and behavior. Yet, mechanisms underlying aging remain largely unknown.

Whether aging is a process or only a by-product or epiphenomenon of life are questions that remain definitively unanswered [8].

Nowadays, aging is being defined as an inevitable consequence of being a multicellular organism; associated with a random and progressive decline in function; leading to a global loss of homeostasis over time; and mortality increasing with age [9]. Present knowledge indicates that aging is a dynamic and malleable process, controlled not only by genes but also by environmental and epigenetic interactions [10].

Currently, it is already known that the aging process, like many other biological processes, is subject to regulation by classical signaling pathways and transcription factors, thus it should follow the normal chemical and physical laws that all life and other complex biological processes obey [11]. Therefore, combining genetic and molecular approaches should be useful to unravel its mechanisms.

1.2 Lifespan Evolution

Longevity is a very heterogeneous phenotype in nature, showing large variation within and between species. Because of the removal of extrinsic hazards, a species' longevity is usually higher in captivity than in the wild. However, there is evidence for the existence of a limit to lifespan in each species which is set by the aging process [12].

The accumulation of damage to macromolecules and cells lead to the loss of function observed with age (senescence), a nearly universal process among multicellular organisms. However, senescence has species-specific features and it is not a simple property of all complex organisms [13].

No genes are known to have evolved specifically to affect lifespan; nonetheless the specific patterns observed in each species suggest that aging does evolve. However, longevity should not be a major role player in the evolutionary process because very few organisms are allowed to age since mortality caused by predation, starvation and other stressors and diseases results in premature death before the action of the aging process [14]. Therefore, the evolution of longevity seems to be a side-effect of the evolution of other fitness related traits [12, 15].

1.2.1 Mutation-accumulation versus pleiotropy

Two main theories have been proposed to explain the evolution of lifespan: the mutation accumulation (MA) and the antagonistic pleiotropy (AP) theories. These theories are not mutually exclusive. In fact they both lay down on the argument that species have evolved senescent life histories because selection is weak against alleles that cause dysfunction only at late ages [16-19].

The non-adaptive MA theory was first described by Peter Medawar [17] and pointed out that any new mutation that arise in a population and has deleterious effects restricted to late ages will be subject to weak selection, and some of these mutations will increase in frequency and will produce declines in individual fitness at late ages. On the other hand, mutations with deleterious effects on early ages will be under stronger purifying selection and will then accumulate in the population to a much lesser extent. Therefore, the frequency of deleterious alleles in populations is a balance between the rate of arrival by recurrent mutation and the rate of elimination by natural selection because of their adverse effects on survival and fecundity. So, aging will evolve as a consequence of the inability of natural selection to maintain survival and fecundity at later ages due to mutation pressure [12]. Species with small effective population sizes, like humans, should be more sensitive to this accumulation of mutations with deleterious effects, and, in contrast, species like *Drosophila* should be more resistant to MA, since they have larger effective population sizes [20].

The adaptive AP scenario assumes the existence of alleles with beneficial effects on the young but deleterious effects at later ages [18]. Positive selection will cause the increase in frequency of such alleles because their early beneficial effects are subject to stronger selection than the late-age effects. Alleles with the opposite effects will then decrease in frequency due to negative selection. The main prediction of this model for within-species variation is a negative correlation between early and late-age fitness traits. It also predicts that genetic variation should be mainly due to alleles that segregate at intermediate frequencies. This theory is also known as the pleiotropy or trade-off theory, and predicts that aging is causally connected to events occurred earlier in life and can be slowed down by the trade-off with the fitness of young adults.

If mutation-accumulation is responsible for aging, it would be expected that inherited genetic variation for mortality and fecundity would increase with age. Because

deleterious mutations are present at higher frequencies, the later the age-class that they affect, the more likely are these mutations to be passed to offspring. However, no clear experimental evidence has supported these predictions, what could mean that these kind of mutations seem to be too rare to cause aging [12]. Furthermore, the MA theory predicts that death rates increase rapidly after the age at which the strength of selection declines to zero in nature, usually after the end of reproductive age. But, despite the exponential increase in mortality with age, there is no sudden increase at very late ages. In fact, at this point death rates tend to decline [12]. Therefore, although it is plausible, the actual effect of late-age deleterious mutations should be too little to make an effect.

In contrast, the pleiotropy theory has received strong experimental support. Measurements of the properties of genetic variation in *Drosophila* natural populations have revealed a correlation between high early fecundity and early death (Fig. 1). At least in *Drosophila*, the trade-off appears to be between high fecundity and slow subsequent aging [12].

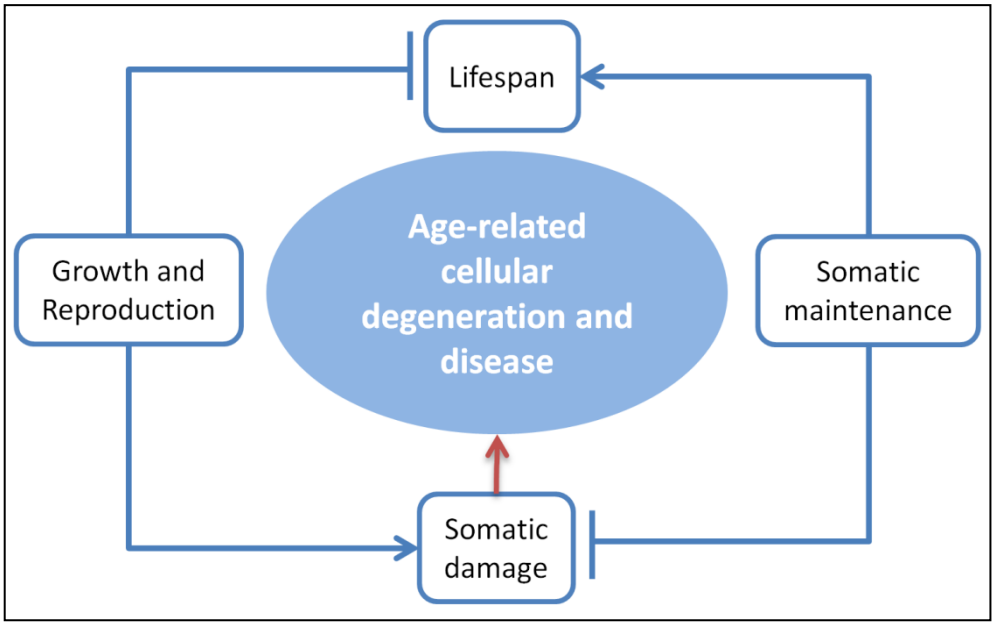

Figure 1 – The trade-off between aging and fecundity. According to the "disposable soma theory", organisms must compromise between energy allocation to growth and reproduction or somatic maintenance and repair. Adapted from Vijg and Campisi [21].

Like many fundamental questions in evolutionary genetics, determining whether aging is primarily a result of adaptive or non-adaptive processes has been a difficult

challenge. Modern system-genetic approaches should be a key tool on the discovery of an unbiased sample of polymorphisms that directly control variation in aging within and between populations. Acquiring more knowledge about the functional significance of these polymorphisms, together with population-genetic tests of selection, should yield more insight into the functional basis of longevity and the molecular mechanisms underlying natural variation [13].

1.2.2 Aging and extrinsic hazard

Broad patterns of variation in natural population are consistent with the idea that longevity evolves in response to hazard. High levels of extrinsic hazard will mean that selection upon early life fitness will be stronger, and selection for slow aging will be weaker. Therefore, life-history trade-offs will favor earlier performance rather than higher lifespan [12]. If the level of extrinsic mortality is high, the average survival period will be short and there will be little selection for a high level of maintenance. Under these conditions, the resources will be channeled towards reproduction, and therefore the individual will live less. In contrast, if the level of extrinsic mortality is lower, selection is likely to act towards the maintenance of the soma [22]. In safe environments aging will evolve to be retarded, whereas in more hazardous environments it should evolve to be more rapid. Adaptations that reduce extrinsic hazard mortality (such as wings and protective shells) are also, in general, linked with an increase in longevity [22].

In *Drosophila melanogaster*, multiple loci analyses have revealed the existence of latitudinal clines for several life-history traits, in which the frequencies of the derived alleles increase in frequency with latitude suggesting a possible adaptation to temperate habitats [23]. This scenario brings up a possible selection regime: in high latitudes cold variation across the year imposes seasonal stresses and favors the most tolerant genotypes. Traits correlated with these genotypes, which may evolve as co-adapted responses or by indirect selection via pleiotropy, will then be characterized by better ability to overcome winter, larger body size, expanded longevity and slower development when in comparison to low latitude populations [24].

1.3 Mechanisms underlying longevity

Understanding the genetic and environmental factors affecting lifespan is an aspect of great interest for evolutionary and applied biology. However, from the evolutionary point of view, has not yet been possible to understand why aging occurs and why there is so much variation in lifespan between and within species [17, 18].

Longevity seems to be governed by multiple mechanisms, such as oxidative stress, cell loss, dietary restriction, nutrient sensing pathways, and several other factors, many of them conserved across species. Due to the high complexity of this phenotype there are no direct methods of measuring aging. Nonetheless, examining the age of death of the individuals could be a powerful tool to identify some important aspects of this process [8]. Several alterations (genetic or environmental) have been used during the years to alter the lifespan of populations in several species. These studies can provide useful information about the aging process and the age-related mechanisms that occur in organisms.

During the last century *Drosophila* has been widely used as a genetic model organism for studying biological complex phenomena. The study of the aging process in *Drosophila* beneficiates mostly with (i) its relatively short lifespan, (ii) ease of maintenance, (iii) environmental and genetic manipulations that alter lifespan, (iv) already available information on aging, (v) availability of stocks containing altered genes, (vi) powerful genetic techniques, (vii) full genome sequence of several species, and (viii) successful utilization in dissecting other complex biological traits such as development [9]. Another important aspect is its life-history, which is clearly divided in two stages, allowing the easy distinction between the growth and development period and the mature adult phase. The mature adult consists almost of postmitotic fully differentiated cells [25]. Thus, aging in *Drosophila* is restricted to a set of cells that are present from the time of adult emergence until its death [26].

Therefore, if the research goal is the understanding of the changes that occur in cells and organs over the time, then organisms such as *Drosophila*, which are almost entirely postmitotic, are excellent model systems for this kind of studies.

1.3.1 Nongenetic factors influencing lifespan

Many mechanisms affecting longevity have already been documented and many of them are conserved across species. In *Drosophila*, changes in ambient temperature,

reproduction and diet can greatly alter lifespan and the slope of the mortality curves. In fact, the aging process is often accompanied by a decline in reproduction, in a well studied trade-off between these two processes (see for review, [8, 27, 28]).

Submitting the organisms to mild non-lethal stressors has been shown to increase lifespan in *Drosophila* [29-32]. Mild stressors, such as periods of heat shock, cold stress or low levels of radiation, can significantly increase lifespan, probably by inducing protective systems that result in short-term beneficial effects [8].

The effect of the exposure to short periods of heat stress on lifespan was first described by Maynard Smith [33]. Ever since, several studies have documented this effect of lifespan extension by inducing high levels of heat-shock proteins (hsp) on diverse organisms (see for example, [34-36]). On the other hand, lifespan extension induced by exposure to non-lethal doses of radiation seems to be a by-product of the effect that radiation has on reproduction rather than a direct effect of this stressor on longevity because, in females, irradiation seems to cause a temporary sterility leading to the lifespan extension observed [37, 38].

One of the most studied factors influencing lifespan in a dramatic manner is temperature. A decrease in ambient temperature towards 18ºC can lead to a more than two-fold increase in lifespan [39]. This effect of temperature on lifespan is probably through a direct effect on metabolic rates. An interesting aspect that corroborates these findings is that, in more severe conditions (like 11ºC) young female flies arrest their egg development at the previtellogenic stage entering into a state of reproductive diapause [40]. Additionally, both males and females can remain at these conditions for at least 11 weeks and, after this period, by putting them on normal temperature conditions they become fully reproductively active and live a completely normal life [41]. Reducing physical activity seems also to influence lifespan through the reduction of metabolic rates. Therefore, flies that move less (e.g. flies living in smaller containers) will live longer [42].

Differences in lifespan have been reported between males and females, however sex-specific effects have been largely ignored throughout time and understanding these differences should lead to a more accurate comprehension of the mechanisms underlying aging in both sexes [43]. It has been proposed that male reproductive strategies are typically associated with high mortality risks. This goes through the idea that males are usually selected to pursue a “live fast, die young” reproductive strategy, characterized by

higher mortality rate and more rapid aging in males than in females [44, 45]. Males usually sacrifice longevity for the possibility of enhanced mating success, whereas females have not this potential benefit because female fitness is limited by the costs inherent to offspring production [46]. Another aspect that could explain differences between sexes is the production of hormones that induce the expression of sexual traits and that can increase mortality rate. Insects do not have testosterone, but males exhibit specific levels of juvenile hormone, insulin-like growth factors and other hormones that stimulate sexual trait expression and affect mortality and aging [47-49]. It seems that males take more risks and trade longevity for enhanced sexual performance, whereas females seem to benefit more by investing in immunity and longevity [50]. Egg production has been associated with an acceleration of the aging process in females [33]. Since then it has been frequently stated that there is a trade-off between reproduction and the aging process [51-54]. The most prevalent theory relating reproduction and lifespan predicts that reproduction is costly, particularly in females, for two main reasons: egg production and sexual harassment of females by males [28].

In several studies it has been observed that there is a decrease in longevity response to reproductive effort in both sexes [33, 55, 56]. It is thought that there is a great amount of energy and other resources that are used in the oogenesis at the cost of other physiological processes. However, this only seems to be true if the energy supply is too limited, e.g. in case of starvation [57]. In contrast with these observations, it has also been shown that the ablation of germ cell line is not enough to promote an extension on lifespan in females, and in males the difference is also inexistent or slightly augmented [58], suggesting that, if there is an association between egg production and longevity it should be the result of the interaction between the oocytes and the surrounding mesodermal cells rather than a direct consequence of the germ cells development.

In 1991, Trevitt and Partridge [59] concluded that receiving sperm, although it has beneficial effects, such as the enhancement of oogenesis and oviposition in normal females, it is also costly because it reduces lifespan. This could be caused by several reasons, some of them being valid only in the case of multiple mating: e.g. injuries caused in the female during copulation attempts [60, 61]. Another aspect contributing to this possible negative effect is that the ejaculate might have some toxic features. In *Drosophila*, the effect of the so called “sex peptide” is well documented to cause mating costs, however

its mechanism of action is still not fully understood [62-64]. Nevertheless, despite some substantial findings that relate the costs of reproduction to lifespan this relationship is not completely cleared, and this is not observed throughout species, with some cases where high fecundity is positively correlated with long-lived females [28].

The availability of food and the nutritional status of individuals have major effects upon reproductive rate and play a central role in lifespan determination [65]. *Drosophila* adult individuals are largely composed of post-mitotic cells and, therefore, their nutritional requirements are mainly to meet the costs associated with reproduction, movement and maintenance of the soma [65]. In the last decades it has been shown that increasing nutrition leads to an elevated reproduction and also resulted in a reduction on lifespan [66, 67]. Therefore, it seems that there is a positive correlation between reproduction and food availability that negatively influences lifespan. In fact, caloric or dietary restriction extends lifespan in a variety of organisms, as is the case of mammals, flies or worms (Table 1) [68-70].

Table 1 – Effects of dietary restriction observed in several organisms. The obtained results differ widely between species and little is known about the mechanisms underlying these effects (adapted from Fontana *et al.* [27]).

	Lifespan increase	**Beneficial health effects**
Yeast	3-fold	Extended reproductive period
Worms	2- to 3-fold	Resistance to misexpressed toxic proteins
Flies	2-fold	None reported
Mice	30-50%	Protection against cancer, diabetes, atherosclerosis, cardiomyopathy, autoimmune, kidney and respiratory diseases; reduced neurodegeneration
Monkeys	Trend noted	Prevention of obesity; protection against diabetes, cancer and cardiovascular disease
Humans	Not determined	Prevention of obesity, diabetes, hypertension. Reduced risk factors for cancer and cardiovascular disease.

In *Drosophila* it has been observed that flies eat several times each hour [65] and, when the access to food is restricted there is an increase in mortality rates during periods of starvation [71]. Then, the most efficient and usual way to study dietary restriction in flies is to dilute the concentration of nutrients in the food, leaving the flies to have free access to

it. With this approach it has been shown that flies do not change their behavior to compensate the low nutrients availability [66, 72, 73] and therefore, lifespan extension due to dietary restriction seems to be a direct consequence of reduced nutrient ingestion.

Manipulations of the food, namely by altering yeast/sugar concentration, cause important physiological responses particularly in females. It has been shown that in dietary restricted (DR) flies egg-laying is rapidly arrested, suggesting that the costs of egg-laying could be an important cause for the differences in lifespan observed between females maintained in low or high food conditions [74]. It has also been shown that dietary restricted flies have a higher relative lipid content compared to control flies [75]. This seems to be the reason why DR flies are more resistant to starvation and live longer [76]. Now, it is already known that this higher lipid levels are the result of the increase in fatty acid anabolism, decreased fatty acid catabolism or a combination of the two [65]. However, and despite over 60 years of studies in dietary restriction, at the present time it is still unknown what are the precise mechanisms that mediate its effects on longevity and whether these mechanisms are conserved across species. This relationship seems to be complex, involving both upregulation of protective mechanisms as well as remodeling metabolism across acute nutritional switches.

1.3.2 Genetic basis of lifespan

It is now largely accepted that genetics plays a major role in the process of aging and in the determination of longevity. However, despite the fact that aging is hereditary, it does not mean that there is a single, fixed genetic pathway that determines its entire process [77]. Studies in model organisms, like *Caenorhabditis elegans* and *D. melanogaster* have revealed dozens of mutations with extended longevity phenotypes. Using different approaches (as mutational and Quantitative Trait Loci analyses) several genes have been identified as candidates to explain lifespan variation in *D. melanogaster* (reviewed by Paaby and Schmidt [24]). Many of these candidate genes are associated to an extension on lifespan both by reduced or extended expression when mutated.

Many of the mutations found to extend lifespan affect the activity of stress-response genes or nutrient-signaling pathways, correlating with the fact that reducing the nutrient intake leads to an extension of lifespan, which is also seen in organisms with an increased stress resistance. The possible justification for these findings is that under harsh conditions,

cells and organisms must be able to enter in a standby mode, turning down some genes whereas others are turned up, resulting in physiological changes that intend to minimize cell division and reproduction to allow energy saving for the systems maintenance [11, 27].

Given the high heterogeneity of the mechanisms affecting lifespan, many of which conserved across species, it is imperative to gain a better understanding of the genetic networks underlying these mechanisms and which are the relationships between them. Therefore, the full characterization of this complex trait can only be achieved on a genome-wide level.

1.3.2.1 Insulin/Igf (insulin-like growth factor) signaling

The insulin/Igf signaling (IIS) pathway seems to have evolved early in history of life, and played probably a key role in the evolution of multicellularity [78]. It has already been shown that this pathway plays an essential role in several physiological mechanisms, such as growth control [79], metabolism [80], stress resistance [81], reproduction [82] and longevity in a great variety of organisms, from worms to mammals [83]. In mammals, all of these mechanisms are controlled by two closely related receptors, the insulin and type 1 IGF receptors, both of them expressed on the surface of most cells. These receptors are members of the tyrosine kinase family and initiate their action through high homologous signaling systems, that triggers the autophosphorylation of the receptor and the further activation of the receptor kinase towards intracellular substrates [14].

The *Drosophila* insulin-like receptor (InR) is homologous to both mammalian receptors, as well as to Daf-2 receptor in *C. elegans* [84, 85]. In contrast to mammals, that often have up to four isoforms of the cellular components of IIS, in *Drosophila* they are encoded by single genes, including the InR, the insulin receptor substrate CHICO, the enzyme phosphatidylinositol 3-kinase (PI 3-kinase Dp110/p60) and the protein kinase B (PKB/Akt). The subsequent activation of these components results in stimulation of glycogen, lipid and protein synthesis, as well as in glucose transporter translocation to the plasma membrane with an increase on glucose transport [86, 87]. PKB/Akt also phosphorylates forkhead transcription factors of the FOXO subfamily leading to their inactivation and retention in cytoplasm (Fig. 2). When insulin is absent FOXO transcription factors are not phosphorylated and remain inside the nucleus, where they are activated and regulate gene expression [88]. These transcription factors are involved in a

series of processes, such as, apoptosis [89], cell cycle [90], DNA repair [91] and oxidative stress [92]. In *Drosophila*, as well as in *C. elegans*, reduced IIS have been shown to extend lifespan probably through an increase in resistance to xenobiotics and other stressors and lipid accumulation [83, 93, 94].

The first IIS mutation that extends lifespan was identified by Clancy *et al.* [95], where *chico*^{1} flies have a null mutation in CHICO. In this study *chico* mutation increases median lifespan in females (up to 31% in heterozygous, and 48% in homozygous) and in heterozygous males (up to 13%). This study also revealed that *chico* controls growth because homozygous *chico*^{1} mutants were dwarf, and that the lifespan extension was not caused by this effect on body size, because long-lived heterozygous were normal sized. Mutation in *InR* has also been shown to significantly increase adult longevity [96]. Females with a heteroallelic hypomorphic genotype are small, infertile and live 85% longer than wild type. Activation of *dFOXO* in the adult pericerebral fat body (equivalent to mammalian white adipose tissue and liver) of *D. melanogaster* increases both male and female lifespan, resistance to oxidative stress and alters lipid metabolism [93, 97]. Downregulation of IIS through overexpression of the *dFOXO* transcription factor reduces insulin-like peptide *dilp2* production in neurons and represses insulin-dependent signaling in peripheral fat body. Additionally, deletion of *dFOXO* has been shown to reduce lifespan [98]. dPTEN is a lipid phosphatase with a function in the organization of the actin cytoskeleton [99]. This protein downregulates IIS and causes nuclear localization of dFOXO by antagonizing the activity of PI 3-kinase (Fig. 2). Overexpression of *dPTEN* in head fat-body was also found to increase lifespan up to 20% [93].

All the 12 *Drosophila* genomes publicly available [100] show seven genes encoding DILPs (insulin like peptides), which are thought to be the ligands of the unique *Drosophila* insulin-like receptor, InR [101]. The only exception is *D. grimshawi* which has a duplication of the gene *dilp2* [102]. In contrast, in *C. elegans* there are 40 different ILP peptides [103], and these proteins are thought to be the equivalent to mammalian insulin [104].

Despite the expression patterns and regulation of the DILPs during development are fully characterized [105], little is known about the role of each DILP in the adult fly. Each *dilp* gene shows a different spatio-temporal expression pattern supporting functional differentiation between these ligands. DILP4 is expressed in the embryonic midgut and

mesoderm [101], while DILP6 is preferentially expressed in the larval and adult fat body, being strongly overexpressed in the larvae-to-pupae transition [106]. DILP7 is expressed in some neurons of the female reproductive tract and, when mutated, females become sterile [107]. DILP1, 2, 3 and 5 are expressed in brain median neurosecretory cells (MNCs), being DILP1 present only in the larval brain [101, 108]. DILP5 is also expressed in the ovary [108]. Ablation of MNCs in early larval stages has been shown to cause developmental delay, growth defects and elevated carbohydrate levels in the hemolymph [109]. When this ablation is made during the final larval stage, it results in lower female fecundity, increase lipids and carbohydrates storage, elevated resistance to starvation and oxidative stress and increases lifespan [108].

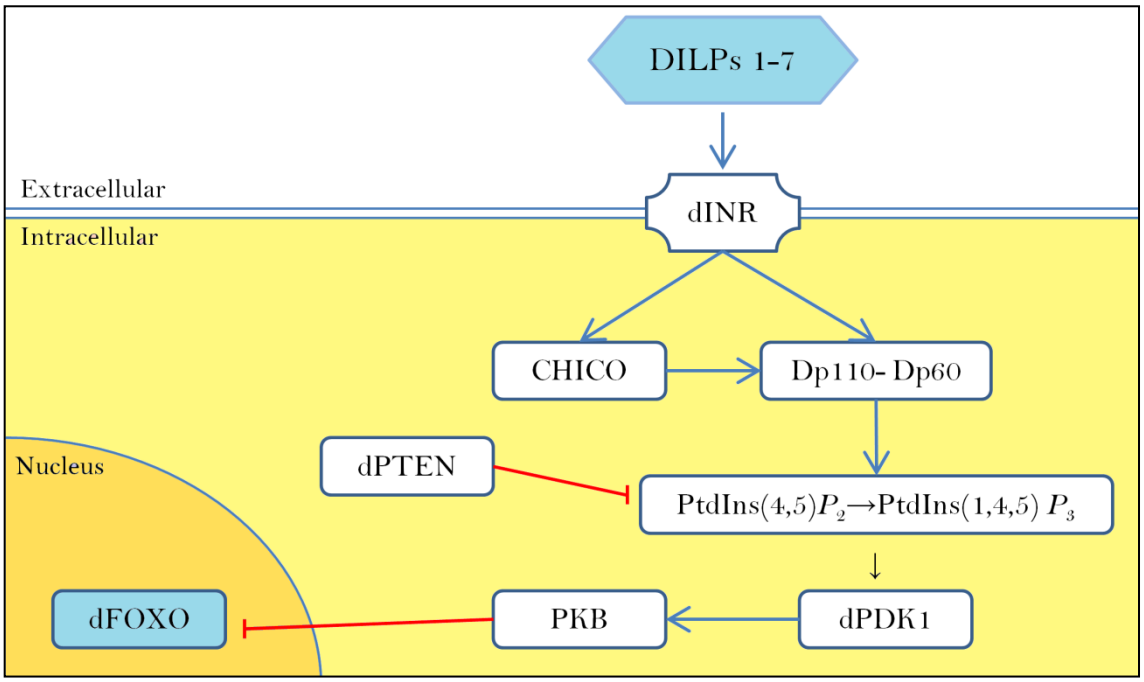

Figure 2 – The *Drosophila* IIS pathway. There is a single *Drosophila* insulin receptor, dINR, which transduces the signal from the DILPs to the lipid PI 3-kinase, either directly or through the single *Drosophila* insulin receptor substrate, CHICO. Dp110 is the catalytic subunit and Dp60 is the regulatory subunit of PI 3-kinase, which converts phosphatidylinositol (4,5)-bisphosphate [PtdIns(4,5)P_2] to phosphatidylinositol (1,4,5)-trisphosphate [PtdIns(1,4,5)P_3]. The action of PI 3-kinase is antagonized by dPTEN, which degrades PtdIns(4,5)P_2 to PtdIns(1,4,5)P_3. The intracellular second messenger PtdIns(1,4,5)P_3 then activates a series of kinases, such as dPDK1 and PKB, which subsequently phosphorylate the transcription factor dFOXO, leading to its inactivation and translocation to the cytoplasm. Adapted from Giannakou and Partridge [110].

All DILPs have the ability to promote growth, but DILP2 is thought to be the most powerful. Overexpression of this ligand also suppresses germline stem cell loss and may also modulate lifespan, since several long-lived mutant lines have this transcript in reduced levels [105, 111]. All seven *dilp* genes have remained present and differentiated from each other in sequence over a period of 40 to 60 million years. Evolutionary conservation of different regions of these peptides suggests that six of them are cleaved like mammalian insulin, while the other (DILP6) remains uncleaved like mammalian IGFs [112]. Despite their different expression patterns and function, some of the DILPs can act redundantly, suggesting that this ability can be evolutionarily advantageous. Some DILPs can act in a synergistic manner while others seem to have compensation mechanisms between them [112, 113].

1.3.2.2 Pathways interacting with IIS

Some pathways have been related to the insulin/insulin-like growth factor signaling pathway and seem to be influencing lifespan, probably through this interaction with the IIS. Among these pathways, the most well studied are the TOR (Target of Rapamycin) signaling [114] and the JNK (Jun NH_2-terminal kinase) pathways (Fig. 3) [115].

The TOR pathway is a highly conserved nutrient-sensing pathway that regulates growth and metabolism to several growth factors, amino acids and various stresses [116]. The role of TOR in lifespan has been addressed by Kapahi *et al.* [117], who showed that reduced TOR signaling through upregulation of the negative regulators of TOR activity, *dTsc1* and *dTsc2*, or using TOR mutants in the downstream kinase dS6K increased *Drosophila* lifespan by up to 24%. The central component of the TOR pathway is TOR kinase, which participates in two different protein complexes: TORC1 and TORC2. TORC1 regulates translation and growth through phosphorylation of the two downstream effectors, dS6K and 4E-BP. Therefore, under diet favorable conditions, TORC1 is active, and the phosphorylation of dS6K functions as a positive mediator of the TOR pathway promoting growth and altering metabolism [118, 119]. On the other hand, phosphorylation of 4E-BP disrupts its association with the translation factor 4E (elF4E), promoting cap-dependent translation. However, under stress conditions, cap-independent translation increases, allowing the synthesis of other proteins, such as heat shock proteins and growth factors that permits the cells to respond to stress and survive [120]. On the contrary,

TORC2 complex is not so well studied and seems to be involved in actin organization and upregulation of IIS through the phosphorylation of Akt, which is the main kinase of IIS pathway [116]. After phosphorylation, Akt can then inactivate dTSC2, a TOR pathway suppressor. In contrast, the downstream component of TORC1, dS6K, inhibits IIS by negative regulation of the insulin receptor substrate at the transcription, degradation and phosphorylation levels (Fig. 3) [119].

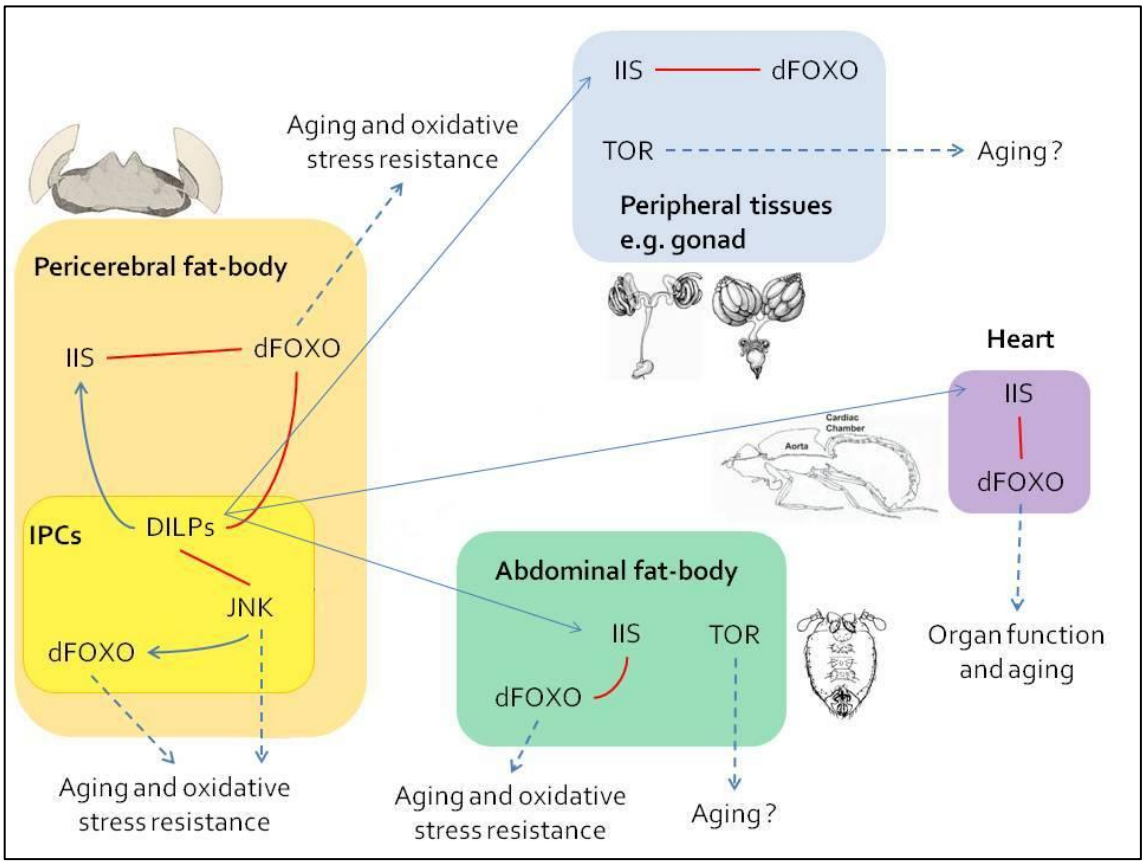

Figure 3 – Model of regulation of adult *Drosophila* lifespan by the IIS and interacting pathways. DILPs are produced by the IPCs in the adult brain and systemically regulate physiological responses and aging through the IIS pathway. dFOXO transcription is inhibited byIIS and, in the pericerebral fat-body, it has been shown to feedback to the IPCs to reduce expression of *dilp2* increasing lifespan and oxidative stress resistance, and to directly regulate these mechanisms. The JNK pathway activates dFOXO in the IPCs and inactivates IIS, by acting on the IPCs to reduce *dilp2* transcript expression also increasing lifespan and oxidative stress resistance. IIS has been also shown to function in an autonomous manner in some tissues to regulate organ function. The TOR pathway also regulates aging in the fat-body, gut, muscle and gonad. Solid blue arrows indicate activation; broken blue arrows indicate possible regulation and red lines represent inhibition. Adapted from Giannakou and Partridge [110].

The JNK (Jun-N-terminal Kinase) signaling is an evolutionarily conserved stress-sensing pathway that is activated by a series of intrinsic and extrinsic hazards (e.g. UV radiation, oxidative stress, DNA damage, heat, infection or inflammation) [115]. JNK phosphorylates numerous transcription factors and enhances their activation potential. In mammals, 20 proteins are known to belong to the JNK kinase family, which are selectively activated by different stimuli and then phophorylate a dual specificity kinase of the MKK family that phosphorylates JNK. JNK itself has a number of cellular targets, mostly transcription factors, including members of the AP-1 family (Jun and Fos) and the transcription factor FOXO (Fig. 3) [121, 122]. In *Drosophila*, the major target gene of AP-1 is *puckered* (*puc*), which encondes a JNK-specific phosphatase that restricts JNK activity [123].

While in vertebrates most of the members of the JNK signaling belongs to large gene families, in *Drosophila* this pathway is much less complex [121]. There is only a single JNK (Basket) and two JNK kinases (Hemipterous, Hep that mediates the most of the JNK effects; and dMKK4, which acts in parallel with Hep in the induction of apoptosis and immune response [124-126]) in *D. melanogaster* genome. However, and despite these differences, the range of effects mediated by JNK is conserved between species and includes apoptosis, morphogenesis, cell migration, cryoprotection and metabolism [121, 127, 128]. In 2003, Wang *et al.* [129] found that flies with mutations that increase JNK signaling specifically in neurons have an increased stress tolerance and extended lifespan. Lately, they also found that JNK is active in the IPCs in the brain, and that *dilp2* expression in these cells is reduced when JNK is repressed [130]. On this way JNK antagonizes IIS, causing nuclear localization of dFOXO leading to the activation of its downstream targets [130]. These findings have shown that JNK may be regulating stress resistance, at least in part, through changes in IIS, and demonstrated that these two pathways seem to be interacting in the regulation of lifespan and stress resistance (Fig. 3) [110].

1.3.2.3 Stress resistance genes

One of the most well accepted theories of aging is the free radical (or oxidative stress) hypothesis, proposed by Harman [5]. According to this hypothesis aging results from the imbalance between the formation and detoxification of reactive oxygen species

(ROS) generated during normal cellular metabolism that leads to the accumulation of oxidative damage to cellular macromolecules, including DNA, proteins and lipids. Through decades several studies have been done that support this theory. For example, the levels of oxidative damage to cellular macromolecules have been reported to increase with age [131]. On the other hand, studies using model species, such as *D. melanogaster*, have shown that increased longevity is associated with reduced oxidative damage or increased resistance to oxidative stress (Fig. 4).

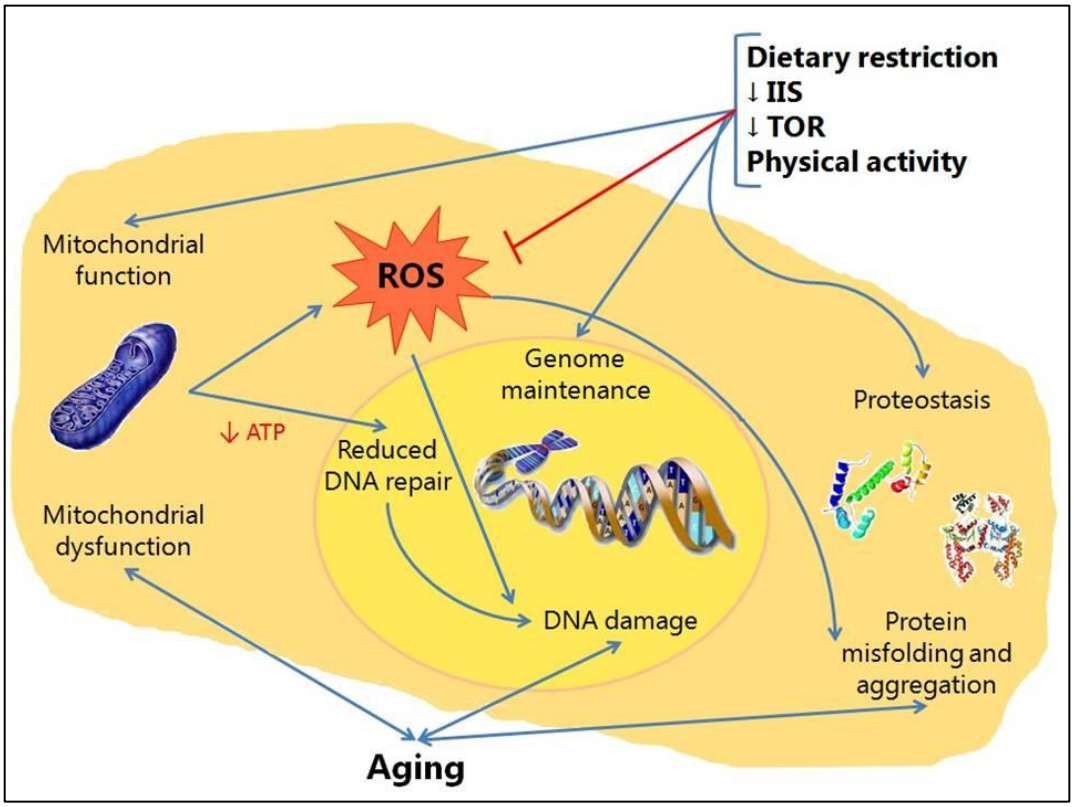

Figure 4 – Age-related stress response. Aging is associated with mitochondrial dysfunction, leading to reduced respiratory metabolism and increased generation of reactive oxygen species (ROS) leading to persistent DNA damage from both increased oxidative damage and increased protein misfolding and aggregation causing age-related cellular damage and physiological decline. On the other hand, dietary restriction extends life span and augments stress resistance by altering cellular metabolism and mobilizing protective stress responses. Most of these gene and protein networks that maintain mitochondrial function, genomic stability, and proteostasis are coordinately regulated by insulin/IGF and TOR signaling. Adapted from Haigis and Yankner [132].

However, in the past few years several studies have challenged this theory and spanned the controversy over this issue (see for example, [133-136]). In fact, experiments involving manipulation of antioxidant defenses have produced inconsistent results [137]; no single antioxidant has yet been shown to consistently increase lifespan in a broad range of species [138]; and no extension of lifespan associated with elevation of defense mechanisms has yet been unequivocally demonstrated to be caused by reduced ROS damage [139].

Decrease in antioxidant enzymes, such as catalase (Cat) or superoxide dismutase (SOD), has been shown to shorten lifespan, suggesting the importance of the detoxification of ROS for the set of lifespan [140, 141]. Additionally, targeted overexpression of these enzymes in some tissues was shown to extend lifespan in various organisms [137, 142, 143], and the administration of antioxidant drugs in organisms with depleted antioxidant defenses have also shown an increase in longevity [144].

1.3.2.4 Other genes and uncharacterized loci

Studies using mutational analysis and QTL mapping have also discovered a series of candidate genes that have not been characterized as members of any known pathway that affects longevity and also several candidate regions with unstudied loci. The single gene alteration approach, either by direct mutagenesis or by ectopic expression, has proven to be a very useful method on the unraveling of the genetic architecture of complex traits. P-element insertion screens have shown that disruption of the G-protein coupled receptor *methuselah* (*mth*) [145] and the Krebs cycle transporter *I'm not dead yet* (*Indy*) extend lifespan [146]. Mutation of the gene *stunted* that encodes the endogenous ligands of Mth, have also been shown to increase lifespan [147], however it has already been shown that Mth receptor exhibits a promiscuous response to other peptides sharing little sequence homology [148].

P type transposable elements have also been used to promote overexpression of certain genes. These elements are engineered to have transcriptional promoters directed out through the end of the element and will often insert upstream of gene coding regions, causing overexpression of the gene and mutant phenotypes [149]. The transcription of these promoters is activated upon feeding the fly with doxycycline, leading to conditional gene overexpression and conditional mutations. This approach has lead to the identification

of several candidate aging genes whose overexpression leads to extended lifespan: *CTP:phosphocholine cytidylyltransferase 1* (*Cct1*), *filamin*, *four wheel drive* (*fwd*), *Sugar baby* (*Sug*), *VhaSFD*, *hebe* and *magu* [150, 151].

The second most used approach is the identification of quantitative trait loci (QTL) affecting naturally occurring variation in lifespan [152]. In *D. melanogaster* at least 22 QTLs affecting variation in lifespan between two inbred lines (Oregon and 2b) have been mapped by linkage to molecular markers [153-156]. Further quantitative complementation tests to mutation at positional candidate genes defined a series of novel candidates affecting longevity that are involved in a variety of cellular functions, such as neuron development (*shuttle craft* (*stc*), *Lim3* and *tailup* (*tup*)); and catecholamine biosynthesis (*Dopa decarboxylase* (*Ddc*), *Catecholamines up* (*Catsup*) and *Dox-A2*) [157-159].

Assessing whole genome transcript abundance with age is another strategy to identify candidate aging genes. Genes showing changes in transcript abundance with age are considered as biomarkers of aging, and if they correspond to QTLs, particularly if they are located inside the regions to which QTL map, they can also be considered as candidate genes [160]. By combining the microarray gene expression data and genetic mapping it was possible to identify 49 candidate genes and four pathways that could potentially be responsible for lifespan regulation and be involved in the aging process [161].

1.4 Assessing natural variation in longevity

In *D. melanogaster* there is significant variation in longevity within natural populations, correlating with latitude and probably driven by differential selection due to climate variation [162]. Longevity can be considered as a typical quantitative trait, with continuous phenotypic variation attributable to the combined segregation of multiple interacting loci with effects that are highly sensitive to environmental challenges [163].

In order to fully understand the mechanisms of longevity, firstly we need to know which loci are responsible for the aging process. This will require not only the identification of the genes involved, but also understanding how their effects are modulated by age and stage-specific environmental variation, and how these interactions are, in turn, influenced by genetic variation at other loci on the genome [156]. Despite the rapidly expanding list of candidate genes for aging, molecular analyses are mostly based in artificially created mutations in the target genes and therefore little is known about whether

the observed genetic variation at these loci occurs in natural populations, or even if this variation would contribute to the phenotypic variance for lifespan observed in nature. For example, not all candidate loci with major effects on longevity may exhibit segregating allelic variation in natural populations. Thus, while the major lifespan effects identified by molecular gerontology may be of biomedical interest, they may be of only limited relevance for our understanding of the evolution of aging in natural populations [163].

The only way to understand the genetic architecture of longevity is to know what subset of loci affecting lifespan also harbors variation in nature, what are the allelic effects at these loci and what molecular polymorphisms define QTL alleles [163]. Moreover, it would be possible to elucidate both how life histories evolve in natural populations and how genotypes translate into phenotypes by doing an evaluation of how polymorphic alleles or individual polymorphisms affect those phenotypes [164].

1.5 Characterization of candidate genes using association studies

The continuous variation observed in natural populations is attributable to the combination of segregating alleles at multiple loci affecting the traits, environmental effects and gene-environment interactions. For any quantitative trait, segregating genetic variation will be caused by a combination of low-frequency alleles with deleterious effects that arose recently by mutation and have not been yet eliminated; selectively neutral alleles that span the range of allele frequencies expected in a population; and alleles at intermediate frequencies that have opposing effects on major components of fitness or are only expressed late in life [152].

Variation for quantitative traits seems to be the raw material on which the forces of selection act to produce phenotypic diversity and adaptation. Evolutionary quantitative genetics has been trying to determine how genetic variation for adaptive quantitative traits is maintained in natural populations. Finding if the loci at which variation occurs within a population are the same responsible for the divergence between populations and species, is an aspect of major interest for the fully comprehension of the evolutionary process, and can provide us a detailed description of the molecular basis of variation in quantitative traits [165].

Association Studies and QTL analyses are largely used for the identification of the genetic basis of the variation observed in complex traits (such as polygenic phenotypes)

[166]. This kind of analyses relies on statistical methods that combine two types of information – phenotypic and genotypic data – in order to try to identify which are the loci responsible for the observed variation in the trait of interest [152].

To perform an association study it is required two or more strains of the organism under study showing genetic variation associated with phenotypic differences regarding the trait of interest and genetic markers that distinguish the two parental lines [165]. In this way, several types of molecular markers can be used, including single nucleotide polymorphisms (SNPs), simple sequence repeats (SSRs, or microsatellites), restriction fragment length polymorphisms (RFLPs), or transposable elements [167-171].

Therefore, an association study starts with the cross between the two parental strains, resulting in a F_1 generation composed by heterozygous individuals. Next, these F_1 individuals are crossed among themselves to create a progeny (F_2 generation) consisting of recombinant individuals containing different fractions of the genome of each parental line. The phenotype and the genotype of each F_2 individual are measured and statistical methods are used to evaluate if there is an association between the molecular marker and the phenotype. Markers that are genetically linked to a locus influencing the trait of interest will segregate more often with a certain value of that trait, whereas unlinked markers will not show any significant association with the phenotype [165].

However, and like most of the tools used in Molecular Biology, an association study is not free of limitations. The main disadvantage is the fact that this kind of studies requires a large sample size, since the statistical power is largely dependent of the sample size. Furthermore, these studies can only map those differences that are captured between the initial parental strains and therefore it will only be possible to analyze genes or loci with alleles that do segregate on the parental individuals [165].

On this way, the main goal of these studies tends to be the identification of candidate loci rather than specific alleles that really explain the phenotypic variation. Thus, the influence of the genes identified by the association study must always be confirmed by functional studies (such as expression analyses, mutagenesis or proteomics) [165].

1.6 *Drosophila americana*

Despite the well-characterized life history variation in natural populations of *D. melanogaster*, it remains unknown if the same set of genes described for this species

can be used to explain lifespan variation in other distantly related *Drosophila* species, so it would be risky to take for granted that molecular variation in this set of genes will explain the observed variation in other species.

In order to address this issue, expression levels, as well as DNA polymorphism, must be assessed for the same set of genes in distantly related *Drosophila* species. Therefore, in this work we studied *Drosophila americana*, a temperate species of the *virilis* group of the *Drosophila* genus that has been diverging from *D. melanogaster* for about 40 My [172]. Nowadays, it has been already shown that *D. americana* can be widely used as model species in comparative studies [171, 173-175]. This species is native to the United States of America and has been independently evolving on this continent for approximately 1 million years [176, 177]. This species is widely distributed, from the Central and Eastern regions of the United States from the South (Texas to the states around the Gulf of Mexico) to the North of the country (from Montana to Maine) (Fig. 5) [178].

Individuals can be easily collected along the margins of marshes, lakes, and rivers, especially those where there is a high density of *Salix* species, when temperatures are below 25º C. It can also be easily maintained under laboratory conditions on a medium containing 10% (mass/volume) yeast, 4% (mass/volume) wheat flour, 8% (mass/volume) sugar and 0.4% (mass/volume) salt diet (1% agar (mass/volume); 0.5% propionic acid) [179].

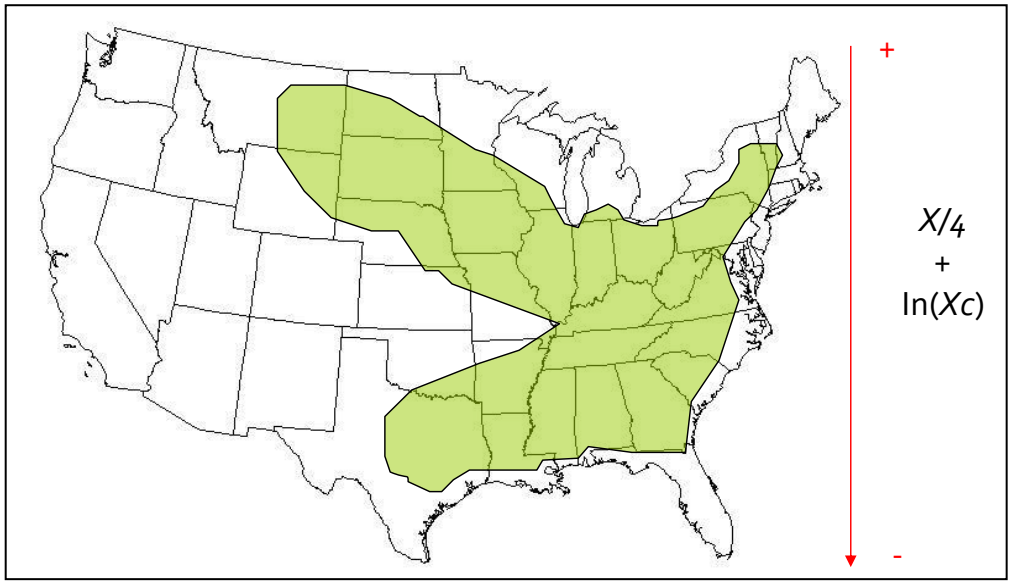

Figure 5 – Areas of distribution of *D. americana* (adapted from Hilton and Hey [180]).

Recent studies indicate that *D. americana* has had a large, stable population and low levels of population structure and there is no evidence of any recent significant reduction in population size [177, 181].

D. americana belongs to the *virilis* group, one of the most well studied groups of the subgenus *Drosophila*. This group consists of 12 species [182] that are further grouped into four subphylad: *D. virilis*, *D. montana*, *D. littoralis* and *D. kanekoi* [183, 184]. Relying primarily on chromosomal differences as centromeric fusions and paracentric inversions, Throckmorton [185] proposed that the *virilis* subphylad included five taxa, namely: *D. virilis*, *D. lummei*, *D. novamexicana*, *D. americana americana*, and *D. a. texana*, the latter three forming the *americana* complex (Fig. 6).

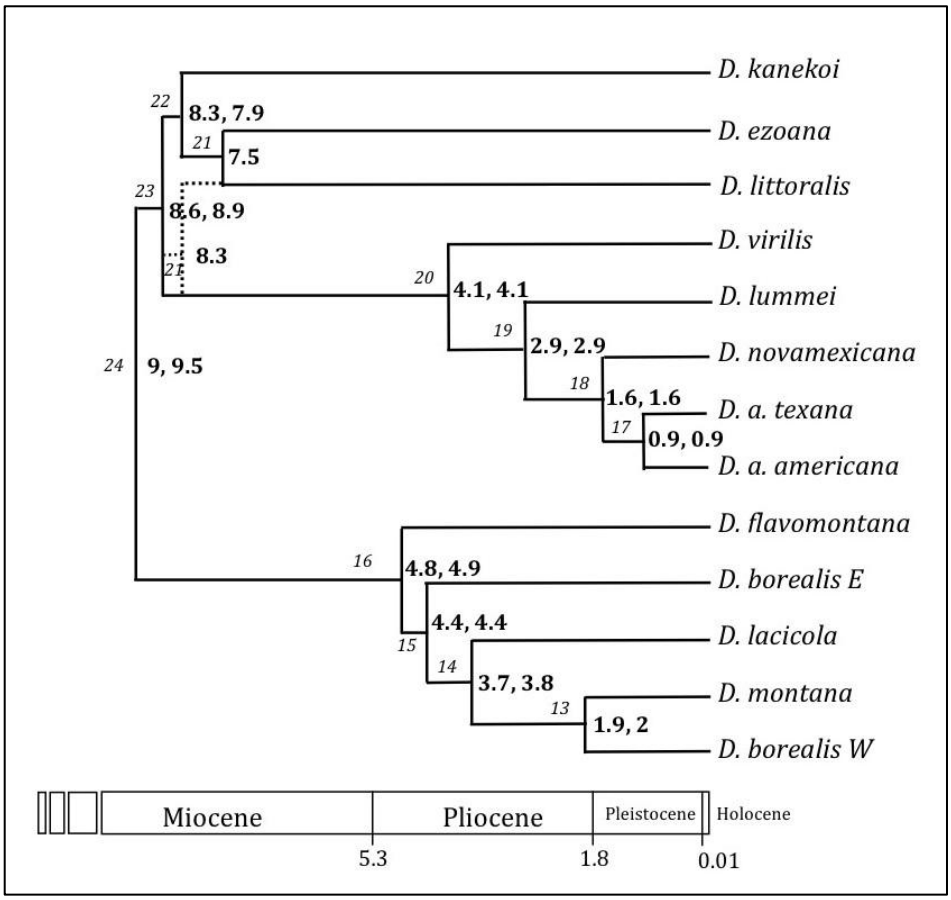

Figure 6 – Phylogenetic tree of the *virilis* group obtained with the BEST topology. Times in MY for the different cladogenesis events are shown inside the nodes in bold; the first is the age obtained for the BEST tree and the second is the age obtained for the Bayesian tree [182].

D. americana possesses a fusion between the *2*nd and the *3*rd chromosomes that is absent in *D. novamexicana*, and it also has a centromeric fusion of chromosomes *X* and *4* that is frequent on the north and almost absent on the southern populations of *D. americana* [186-189]. In fact, based in this latitudinal cline observed for the *X/4* chromosomal fusion, this species was originally thought to be made of two subspecies (*D. a. americana* and *D. a. texana*), that would overlap in a small hybrid zone [180]. However, several molecular studies revealed that *D. a. americana* and *D. a. texana* individuals are indistinguishable at the DNA level for genes on chromosomes *2, 3, 4* and *X* with the exception of genes on the basis of the *X* chromosome [180, 186-188, 190-192]. These results support the idea that there is enough gene flow between individuals with and without this chromosomal arrangement and therefore, *D. americana* is, in fact, a single species [189].

In *D. americana* there are also six inversions with estimated frequency higher than 5% (*X/4* fusion, *Xc*, *2b*, *4a*, *4b*, *5a* and *5b*). Some of these rearrangements are shared between some populations of *D. americana* and *D. novamexicana*. For example, the *Xc* inversion is fixed in *D. novamexicana* and almost fixed in the northern populations of *D. americana*, whereas in south populations it is almost absent [193]. Given this, it is believable that the *X/4* fusion arose in an *Xc* inverted chromosome [186, 188, 189], and that this rearrangement only occurred after the split between this two species (about 1,6 million years ago) [182]. These findings together with the low genome wide levels of population structure observed in *D. americana* [177, 181], have clearly shown that it is not made of two subspecies but it is rather a single taxa showing different chromosomal rearrangement frequencies in different populations [193]. Weak selection probably brought the *X/4* fusion to high frequency and is maintaining the *X/4* fusion *Xc* inversion chromosomal arrangement [186-188]. Moreover, molecular markers are available for the *X/4* fusion, *Xc*, *4ab*, and *5b* chromosomal arrangements, and perfect markers could be developed for *4a* and *5a* since their breakpoint sequences have been determined [194, 195]. Therefore, if desired, genotype-phenotype association studies where such chromosomal rearrangements are not segregating can be confidently set up.

The average recombination rates are high for this species, what is a desirable feature when performing phenotype-genotype studies, either classical F_2 association studies or genome-wide association studies (GWAS).

Another aspect of huge importance that must be taken into account is that the *D. americana* genome has been recently sequenced for two different strains and its whole assembly is now already available. Having two assembled reference genomes greatly facilitates the development of markers (either anonymous or for candidate genes) for F_2 association studies. Moreover, GWAS are now possible in *D. americana* since the reads that are generated in such studies can now be mapped to these reference genomes.

2. OBJECTIVES

The overall objective of this study is to determine whether molecular variation found in candidate genes for lifespan described in *D. melanogaster* would explain phenotypic variability for this trait in the distantly related species *D. americana*. By doing so, this work also intends to verify to what extent the results obtained for the model species, such as *D. melanogaster*, can be generalized to other species. We are also interested in determining if the variation observed in lifespan between different populations is due to gene expression changes or to amino acid substitutions that truly influence the phenotype.

In order to achieve these goals the following steps were addressed:

- Development of an F_2 association study by establishing five F_2 association crosses between five different *D. americana* strains from different geographic regions showing different chromosomal rearrangements and phenotyping of the F_2 individuals for lifespan;

- Development of molecular markers for 21 of the candidate genes described to be involved in longevity determination in *D. melanogaster*;

- Genotyping of the F_2 individuals for the candidate genes (DNA extraction, PCR amplification and restriction enzymes typing) and look for statistical associations between naturally occurring polymorphisms and lifespan;

- Bioinformatics analysis and candidate genes' sequencing looking for evidence of amino acid polymorphisms;

- Gene expression analysis throughout several time points in two *D. americana* strains.

3. METHODOLOGIES

3.1 F_2 Association experiment

Five isofemale strains (H5 from Lake Hurricane, Mississippi; W11, W29 and W46 from Lake Wappapelo, Missouri; and O57 from Fremont, Nebraska) were used to establish five different F_2 association crosses (Table 2). These strains were selected, since they were established with flies originated from distinct regions of the distribution and, according to the markers used they show different chromosomal rearrangements, such as, the *X/4* fusion, *Xc*, *4ab* and *5a*/*5b* inversions and show differences regarding several phenotypic traits. The *X/4* fusion and *Xc* chromosomal rearrangements were typed as described by Vieira *et al.* [186], Vieira *et al.* [188], Reis *et al.* [196]. The *4ab* inversion was typed as described by McAllister [189] and the *5b* inversion as described by Reis *et al.* [171] (Supplementary table 1).

After eclosion about 100 males of each cross were individually collected and maintained at 25º C under 12h light and dark cycle conditions until they were dead in order to determine their lifespan.

Table 2 – F_2 association crosses established between the five *D. americana* strains used in this work.

Cross	Nomenclature
H5 ♂ x W11 ♀	AA
W11 ♂ x W46 ♀	AT
W29 ♂ x O57 ♀	TN
O57 ♂ x H5 ♀	NA
W46 ♂ x W29 ♀	TT

A – strains from the centre of the distribution showing the northern characteristic chromosomal rearrangements (Americana type)

N – strain from the north of the distribution (Nebraska)

T – strains showing the southern characteristic chromosomal rearrangements (Texana type)

3.2 Candidate gene approach

Starting from the amino acid sequence of *D. melanogaster* candidate genes, the tBLASTn tool (http://www.flybase.org) was used to find the gene orthologs in *D. virilis* (a species closely-related to *D. americana*).

The *D. americana* genes' sequences were obtained using BLASTn (http://evolution.ibmc.up.pt) against the complete genome sequences from two *D. americana* strains (H5 and W11) and the *D. virilis* orthologs as queries.

Sequences of candidate genes were then used to find allelic polymorphic sites, inside the gene or in its vicinity, between these two strains. The selected polymorphisms were then screened using Polymerase Chain Reaction (PCR) amplification and specific restriction enzymes in the F_2 individuals described above.

3.2.1 Genomic DNA extraction

Flies were kept at -20º C and placed in liquid nitrogen 30 minutes before the beginning of the extraction protocol. Flies were then macerated in liquid nitrogen and put in a 1.5 mL tube with 180 µL of PBS.

Genomic DNA from single males was then extracted using the QIAamp® DNA Mini Kit (QIAGEN, CA, USA) according to the manufacturer's instructions.

Cell lysis was done adding 20 µL of Proteinase K and 180 µL of lysis buffer AL. The samples were homogenized and incubated for 10 minutes at 70 °C. Then, 200 µL of ethanol 95 % was added and the homogenate was spin filtered through a 2 mL DNeasy spin column by centrifugation for 3 minutes at 8000 rpm to isolate the DNA.

The column was placed in a new 2 mL collecting tube and washed using 500 µL of the wash buffer AW1, followed by centrifugation at 8000 rpm for 1 minute. The isolated DNA attached to the column was then washed a second time using 500 µL of AW2 and centrifuged for 3 minutes at the maximum speed (13200 rpm).

Finally, the DNeasy spin column was placed in a fresh tube of 1.5 mL, and the DNA was eluted using 100 µL of bidestilled water pre-warmed at 50º C. This set was incubated for 3 minutes and then centrifuged for 1 minute at full speed. This latter step was repeated with the addition of another 100 µL of warm bidestilled water.

Extracted genomic DNA was held at -20º C for further utilization.

3.2.2 PCR amplifications and Restriction enzymes typing

PCR amplifications were performed in a T-Professional basic thermocycler (Biometra®). Standard amplification conditions were pre-denaturation at 95º C for 5 min, followed by 35 cycles of denaturation at 94º C for 30 s, primer annealing at specific temperatures (Supplementary table 2) for 45 s, and primer extension at 72º C for 2 min.

PCR reactions were carried out in a 10 µL final volume solution containing 1 µL of genomic DNA, 4.6 µL of bidestilled water, 0.5 µL of each primer (~5 µM), 1 µL of dNTPs (10 mM) (dATP, dCTP, dGTP, dTTP), 1 µL of PCR buffer 10 x with ammonium sulfate, 1.2 µL of magnesium chloride (25 mM) and 0.2 µL of Taq DNA Polimerase (5 U/µL) (Fermentas®). The amplification products were run at 180 volts in 1.5 % (w/v) agarose gels stained with ethidium bromide, using 1 x SGTB (Grisp, Portugal) and observed under UV light.

After successful amplification, products were subjected to digestion with particular restriction enzymes. Most of the times, 2 to 3 µl of PCR product was digested in a 5 µl reaction volume containing 1 U of Restriction Enzyme and 0.5 µl of the 10 x reaction buffer supplied by manufacturer. Incubation was performed during a minimum of 4 h at the specific temperature for each enzyme. Restriction fragments were run at 180 volts in 2 % (w/v) agarose gels stained with ethidium bromide, using 1 x SGTB (Grisp, Portugal) and observed under UV light. Fragments were typed as 0 (when the amplification products were not digested by the enzyme), 1 (when the products were digested by the enzyme) or 1/0 (for heterozygous).

For each gene, information on the primers, restriction enzymes and the PCR amplification conditions used, as well as the SNP that was typed can be found in Supplementary Table 2.

3.2.3 Statistical analysis

Genotype-phenotype associations were tested using non-parametric tests and the software SPSS Statistics 19.0 (SPSS Inc., Chicago, Illinois, USA). Using the same software, linear regression analyses were performed in order to estimate the amount of phenotypic variation explained by variability in candidate genes.

3.3 Sequence analysis

The sequences of the gene coding region (CDS) from the two strains were also aligned among themselves and with the gene CDS of *D. virilis* using Proseq 2.0 [197]. Aligned sequences were translated into protein in order to look for evidence of amino acid variation in these genes between the two *D. americana* strains.

In order to confirm the evidence of amino acid variation in *dilp2* gene, its entire coding region was amplified in the F_0 individuals of all crosses using specific primers (Supplementary Table 3). Amplification and electrophoresis conditions were the same described above.

Amplification products were extracted using the QIAEX® II Gel Extraction Kit (QIAGEN, CA, USA) according to the manufacturer's instructions.

After isolation, agarose gel bands were weighted and 450 µL of QX1 buffer was added per 0.1 g of agarose. Bands were then incubated with 10 µL of QIAEX® II suspension for 10 minutes at 50º C with agitation. After centrifugation at full speed for 1 min another 500 µL of QX1 buffer was added and centrifuged again. Extracted products were washed with 500 µL of PE buffer and eluted in 3 µL of bidistilled water.

Purified amplification products were then cloned using the TOPO-TA Cloning Kit for Sequencing (Invitrogen, Spain). Positive colonies were picked randomly, grown in 5 mL of LB with Ampicilin, and plasmids were extracted using the QIAprep® Spin Miniprep Kit (QIAGEN, CA, USA).

Bacterial material was firstly concentrated by centrifugation at 4000 rpm for 10 minutes and resuspended in 250 µL of P1 buffer. Cells were lysed with 250 µL of P2 buffer for no longer than 5 minutes and lysis reaction was stopped with 350 µL of N3 buffer. Lysate was centrifuged for 10 minutes at full speed and the supernatant was filtered through a QIAprep® Spin column for 1 minute at full speed. The column was then washed with 500 µL of PB buffer and centrifuged for 1 minute. A second wash was made with 750 µL of PE buffer for 3 minutes followed by 1 minute of centrifugation at full speed. Plasmid DNA was eluted with 50 µL of bidistilled water for 5 minutes followed by 1 minute of centrifugation at full speed.

Pools for each individual were made containing equal parts of the extracted plasmids. Sequencing was performed using ABI PRISM Big Dye cycle-sequencing kit 1.1 (Perkin Elmer, CA, USA) and the primers for M13 forward and reverse priming sites of

the pCR2.1 vector (Supplementary Table 3). Sequencing reactions were carried out in a 5 µL final volume solution containing 2.2 µL of PCR product / plasmid DNA, 0.17 µL of bidestilled water, 0.63 µL of primer solution (~5 µM), and 2 µL of Big Dye solution. Cycle sequencing reactions were performed on a T-Professional basic thermocycler (Biometra®) under the following conditions: pre-denaturation at 96º C for 5 min, followed by 25 cycles of denaturation at 95º C for 30 s, primer annealing at 50º C for 15 s, and primer extension at 60º C for 4 min.

Sequencing products were cleaned-up using a salting-out extraction method. Sequencing products were precipitated using 15 µL of a mixture of ethanol 95 % and sodium acetate 3 M (25:1) and left on ice for 30 minutes followed by a centrifugation of 30 minutes at full speed (13200 rpm). Precipitated sequencing products were then washed with 150 µL of ethanol 75 % and centrifuged for 2 minutes at full speed and allowed to dry for 20 minutes. After purification, products were sent by mail to STABVIDA (Lisbon, Portugal).

In order to confirm if there were possible nucleotide missincorporations during the PCR reactions, direct sequencing of the same amplicons was performed using primers dilp2_seqF and dilp2_seqR (Supplementary Table 3). Amplification products were extracted using the QIAquick® Gel Extraction Kit (QIAGEN, CA, USA) according to the manufacturer's instructions.

Amplification products were isolated, excised from agarose gels and purified. After weighing the agarose gel bands, 450 µL of QG buffer were added per 0.1 g of agarose and incubated at 50º C for 10 minutes (until complete dissolution of the agarose slice). PCR products were precipitated with 100 µL of isopropyl alcohol per 0.1 g of agarose and the mixture was transferred to a QIAquick® spin column and centrifuged for 1 minute at full speed in order to bind the DNA to the column. After discarding the supernatant, another 500 µL of QG buffer were added and centrifuged for 1 minute at full speed in order to completely remove agarose residues from the column. Adhered DNA was then washed with 750 µL of PE buffer followed by 1 minute of centrifugation at full speed. After discarding the supernatant, another centrifugation was made in order to completely dry the column.

The column was then placed in a fresh 1.5 mL tube and the DNA was eluted with bidestilled water for 1 hour and 10 minutes followed by centrifugation for 1 minute at full

speed. Purified amplification products were then used for sequencing reaction as described above and all sequencing runs (direct and after cloning) were performed by STABVIDA (Lisbon, Portugal).

3.4 RNA extraction and cDNA synthesis

Expression levels for genes showing a statistically significant association were determined in sets of three individuals with 0, 10, 30 and 60 days-old from the strains H5 and W11.

Living individuals were frozen in liquid nitrogen and conserved at -80º C until extraction. Total RNA was extracted from each set of three individuals using TRIzol Reagent (Invitrogen, Spain) according to the manufacturer's instructions and treated with DNase I (RNase-Free) (Ambion, Portugal).

Frozen flies were homogenized in liquid nitrogen and, after complete homogenization, samples were resuspended in 1 mL of TRIzol Reagent followed by a centrifugation at 12000 x g for 10 minutes at 4º C. The supernatant was then transferred to a fresh tube and incubated for 5 minutes at 20º C. After that, 200 µL of chloroform were added and the samples were shaken vigorously for 15 seconds and incubated at 20º C for 3 minutes, followed by 15 minutes of centrifugation at 12000 x g at 4º C. After this step RNA remains exclusively in the aqueous phase that is transferred to a fresh tube.

Then, 5 µL of RNase-free glycogen were added as a carrier to the aqueous phase and 500 µL of isopropyl alcohol were also added before an incubation at 20º C for 10 minutes followed by a centrifugation at 12000 x g for 10 minutes at 4º C. RNA precipitates as a gel-like pellet on the bottom and side of the tube and, therefore, the supernatant was discarded.

The RNA pellet was washed twice with 1 mL of 75 % ethanol (prepared with RNase-free water) and centrifuged at 7500 x g for 5 minutes at 4º C. Supernatant was discarded, the pellet was allowed to dry for 10 minutes and then it was dissolved with 24.3 µL of RNase-free water at 58º C for 10 minutes.

Eluted RNA was quantified by UV spectrophotometry (Nanodrop® ND-1000, Fisher Scientific, Portugal), diluted to a final concentration of 200 ng/µL and treated with 1.8 µL of DNase I (2 U/µL), 2 µL of DNase I buffer (10 x) and 2.4 µL of magnesium chloride (25 mM) in a 20 µL final reaction volume with 13.8 µL of total RNA. Digestion

reaction was carried out for 2 hours at 37º C and inactivated with 0.2 µL of EDTA (0.5 M) at 75º C for 5 minutes.

cDNA was synthesized by reverse transcription with SuperScript® III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen, Spain) using random primers. 1 µg of total RNA was pre-incubated with 1 µL of random hexamer solution (50 ng/µL) and 1 µL of dNTP mix (10 mM) for 5 minutes at 65º C to allow RNA priming. cDNA synthesis reaction was then carried out in a 20 µL final volume solution containing 10 µL of RNA plus primers solution, 2 µL of RT buffer (10 x), 4 µL of magnesium chloride (25 mM), 2 µL of DTT (0.1 M), 1 µL of RNase OUT (40 U/µL) and 1 µL of SuperScript® III Reverse Transcriptase (200 U/µL). Reactions were performed on a T-Professional basic thermocycler (Biometra®) following the program: 10 minutes at 25º C, followed by 50 minutes at 50º C. Reactions were stopped at 85º C for 5 minutes, followed by chilling on ice. RNA remains were degraded by an incubation of 20 minutes at 37º C with 1 µL of RNase H (2 U/µL).

No-template controls and reactions with RNA that was not reverse transcribed were also performed in order to confirm the absence of genomic DNA contamination. Synthesized cDNA and controls were stored at -20º C for further utilization.

3.5 Gene expression analysis

Highly efficient specific primers (Supplementary Table 4) were used when performing qRT-PCR experiments using the isolated cDNA, being every experiment performed in duplicate. qRT-PCR reactions were carried out in a 20µL final volume solution containing 1 µL of cDNA, 8.5 µL of bidestilled water, 0.25 µL of each primer (10 µM), and 10 µL of the iQ SYBR Green Supermix (Bio-Rad, Portugal). The reactions were performed on a Bio-Rad iCycler with the following program: 3 min at 95º C; 40 cycles of 30 s at 94º C, 30 s at 56º C and 30 s at 72º C followed by a standard melt curve. Specific primers (Supplementary Table 4) were also developed for the endogenous *Ribosomal protein L32* (*RpL32*) which was used as the reference gene. Fold change in expression was calculated using the $2^{-\Delta\Delta CT}$ method [198].

4. RESULTS

4.1 F_2 association experiment

Five different *D. americana* F_2 association crosses have been set up in order to look for evidence of associations between common naturally segregating polymorphisms and lifespan.

In this experiment, considerable variation in lifespan was observed between the individuals and crosses, with the longevity ranging between 12 and 124 days (Fig. 7). The observed average lifespan for each cross is under a normal distribution (One-Sample Kolmogorov-Smirnov Test; for all cases $p > 0.05$). Using a sample of unrelated individuals (one per strain) from different *D. americana* populations the obtained mean lifespan was of 56.3 days (data not shown), this value is not very distant from those obtained with the F_2 individuals used in this study.

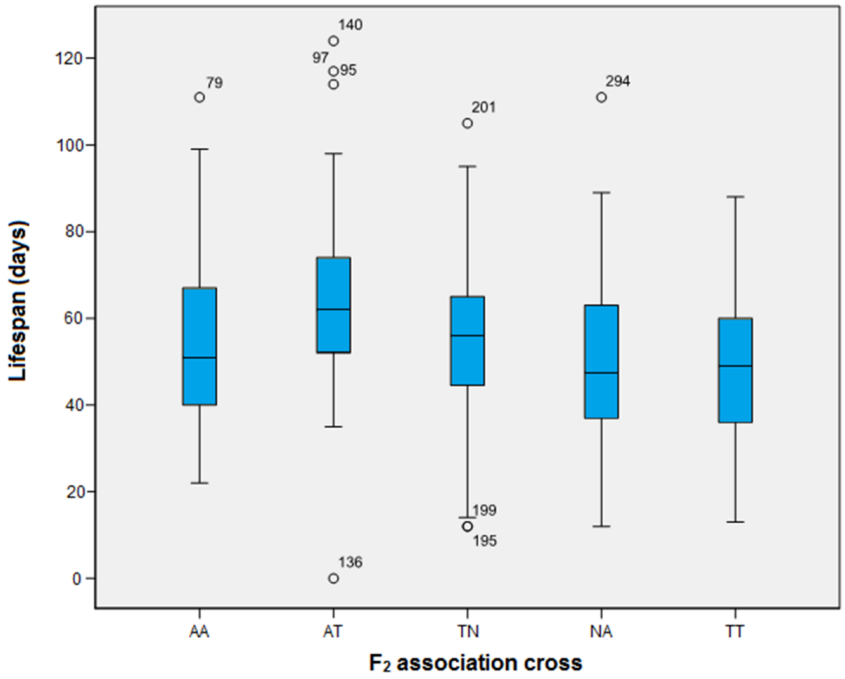

Figure 7 – Lifespan variation observed in the five F_2 populations.
AA (H5♂ X W11♀): Mean = 54.91; Median = 51.00; Std. Dev. = 18.261; N = 89
AT(W11♂ X W46♀): Mean = 64.81; Median = 62.00; Std. Dev. = 20.763; N = 75
TN (W29♂ X O57♀): Mean = 54.59; Median = 56.00; Std. Dev. = 17.863; N = 87
NA (O57♂ X H5♀): Mean = 49.67; Median = 47.50; Std. Dev. = 17.599; N = 94
TT (W46♂ X W29♀): Mean = 46.47; Median = 49.00; Std. Dev. = 17.239; N = 89

However, there are highly significant differences regarding lifespan between the five F_2 populations (non-parametric Kruskal-Wallis test; $p < 0.001$). Seven out of the 10 possible pairwise comparisons involving different crosses are statistically significant (non-parametric Mann-Whitney test; $p < 0.05$; Table 3), although only four are significant after applying the sequential Bonferroni correction for multiple testing (Table 3).

Table 3 – Lifespan differences among F_2 association crosses.

Cross	W11 x W46 (64.81*)	W29 x O57 (54.59*)	O57 x H5 (49.67*)	W46 x W29 (46.47*)
H5 x W11 (54.91*)	***p* < 0.01#**	*p* > 0.05	*p* > 0.05	***p* < 0.05**
W11 x W46 (64.81*)	-	***p* < 0.05**	***p* < 0.001#**	***p* < 0.001#**
W29 x O57 (54.59*)	-	-	***p* < 0.05**	***p* < 0.005#**
O57 x H5 (49.67*)	-	-	-	***p* > 0.05**

* average lifespan in days

significant after applying the sequential Bonferroni correction

In *D. americana* natural populations, there are seven polymorphic chromosomal rearrangements (the *X/4* fusion, and the inversions *Xc*, *2b*, *4a*, *4b*, *5a* and *5b*) that show very different frequencies in the north and the south of the distribution [193]. It should be noted that, according to Hsu [193], 97.5 % of the *X/4* fusion chromosomes harbor the *Xc* inversion, while only 7.5 % of the non-fusion chromosomes show the *Xc* inversion. Moreover, 84.5 % of *4a* inverted chromosomes show the *4b* inversion, while only 3.2 % of *4b* inverted chromosomes do not show the *4a* inversion. There is no physical overlap between inversions *5a* and *5b* but the two inversions are never found on the same chromosome, although a large number of individuals from the centre of the *D. americana* distribution show both inversions in heterozygosity, and at least one of the inversions is always present. Since chromosomal rearrangements can suppress recombination in heterozygotes creating associations between variants from genes located far away from each other, we genotyped the F_0 of all five association crosses as well as the F_2 individuals from the crosses where such chromosomal rearrangements are segregating (Supplementary table 1). There are only two significant associations (Table 4). Nevertheless, these associations are not significant after applying the sequential Bonferroni correction for multiple testing. It should be noted that these two cases are not independent because of the

X/4 fusion. Therefore, at large, it seems that chromosomal arrangements are not being maintained in the populations because of selection for extended lifespan in males, however, this situation points out the possible presence of a variant (or variants) in the *X* and/or in the 4^{th} chromosome that could be influencing lifespan in *D. americana*.

Table 4 – Associations between chromosomal rearrangements and lifespan.

Cross	N*	Average lifespan in days	Association
W11 x W46	43 31	66.1 (hemizygous *X/4 – Xc*); 63.9 (hemizygous standard)	$p > 0.05$
W29 x O57	49 35	58.7 (hemizygous *X/4 – Xc*); 48.7 (hemizygous standard)	$p < 0.05$
W29 x O57	50 35	58. 2 (heterozygous *4ab*) 48.7 (homozygous standard)	$p < 0.05$
W29 x O57	22 18 47	55.5 (homozygous *5a*) 51.1 (homozygous *5b*) 55.5 (heterozygous *5a*/*5b*)	$p > 0.05$
O57 x H5	50 40	49.6 (homozygous *4ab*) 49.2 (homozygous standard)	$p > 0.05$
O57 x H5	16 21 55	54.9 (homozygous *5a*) 51.8 (homozygous *5b*) 46.9 (heterozygous *5a*/*5b*)	$p > 0.05$

* sample size

4.2 Candidate gene approach

In order to look for evidence of associations between common naturally segregating polymorphisms in candidate genes and lifespan, we choose to study 21 out of the 48 genes reported by Paaby and Schmidt [24]. The chromosome localization of these genes in *D. americana* was inferred starting from their position in *D. virilis* and taking into account the several chromosomal rearrangements present between the two species (Fig. 8).

All genes showing evidence for an association between naturally occurring variation in *D. melanogaster* and lifespan were included (*Catsup*, *Ddc*, *Dox-A2*, *Lim3*, *ms(2)35Ci*, *stc* and *tup*; [24]). We also decided to include seven genes harboring evidence for amino acid adaptive evolution (*filamin*, *fwd*, *puc*, *Cat*, *mei-41*, *PTEN* and *EF-1a48D*, [179]) and another seven randomly chosen candidate genes from the set of genes described by the referred authors (*hep*, *ovo*, *InR*, *dFOXO*, *SOD*, *magu* and *dilp2*).

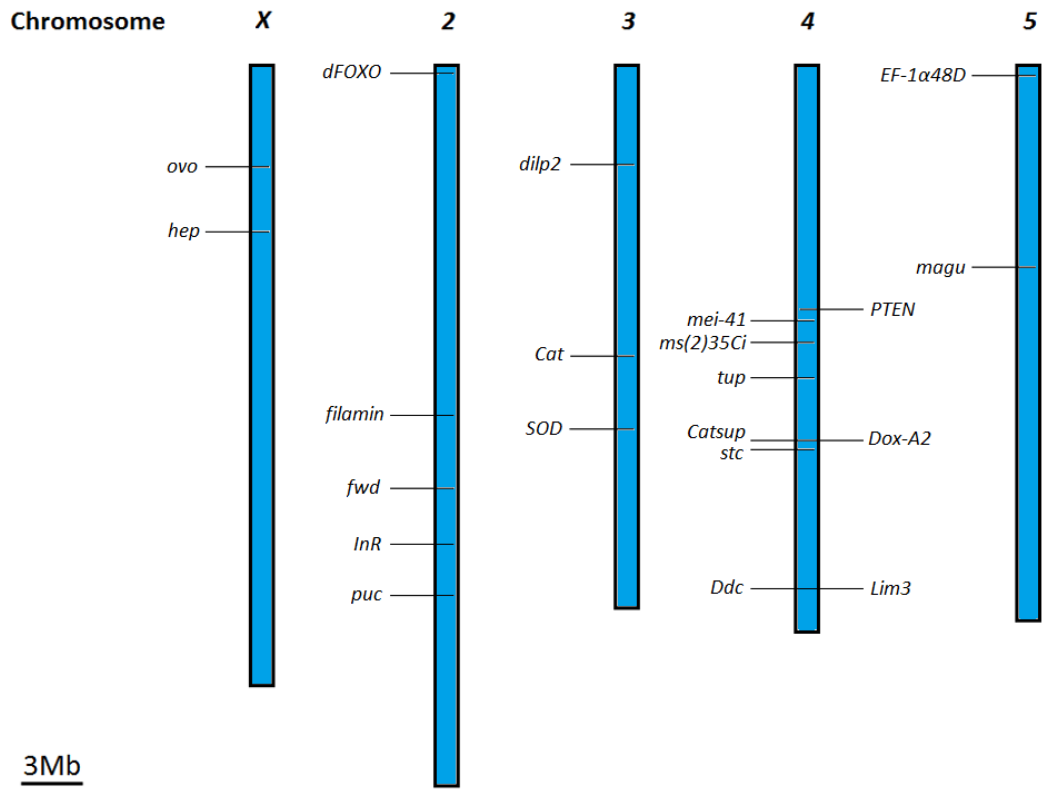

Figure 8 – Schematic representation of chromosomal localization of the studied candidate genes.

Although the markers used have been developed to distinguish polymorphisms occurring between the strains H5 and W11 (AA cross), most of them are also segregating in crosses other than AA cross (on average 36 % of the markers developed are segregating in crosses in which the strains H5 and W11 are not involved, Table 5).

Ten out of the 21 genes studied have shown a statistically significant association in at least one of the five association crosses (Mann-Whitney / Kruskal-Wallis test, $p < 0.05$; Table 5). From these genes, two (*hep* and *Lim3*) remain significant after applying the sequential Bonferroni correction for multiple comparisons. Two genes (*dilp2* and *filamin*) showed an association in two different crosses and, for the case of *filamin* this occurred in two completely independent crosses.

Table 5 – Associations between candidate gene markers and lifespan.

Gene	Chr	F_2 association cross *				
		H5 x W11	W11 x W46	W29 x O57	O57 x H5	W46 x W29
ovo	*X*	$p = 0.116$	$p = 0.378$	n. a.	n. a.	n. a.
hep	*X*	**$p = 0.001$#**	$p = 0.796$	n. a.	$p = 0.543$	$p = 0.490$
dFOXO	*2*	**$p = 0.018$**	$p = 0.473$	$p = 0.828$	$p = 0.728$	n. a.
puc	*2*	$p = 0.524$	$p = 0.229$	n. a.	$p = 0.209$	n. a.
InR	*2*	$p = 0.864$	$p = 0.593$	$p = 0.618$	$p = 0.425$	n. a.
fwd	*2*	$p = 0.862$	$p = 0.580$	n. a.	n. a.	n. a.
filamin	*2*	**$p = 0.024$**	n. a.	**$p = 0.015$**	$p = 0.758$	$p = 0.508$
dilp2	*3*	**$p = 0.003$**	n. a.	$p = 0.985$	**$p = 0.011$**	$p = 0.185$
Cat	*3*	**$p = 0.009$**	n. a.	n. a.	$p = 0.144$	n. a.
SOD	*3*	$p = 0.184$	n. a.	n. a.	**$p = 0.044$**	n. a.
PTEN	*4*	$p = 0.771$	$p = 0.732$	**$p = 0.043$**	n. a.	$p = 0.413$
mei-41	*4*	$p = 0.511$	$p = 0.435$	$p = 0.604$	$p = 0.061$	n. a.
ms(2)35Ci	*4*	$p = 0.124$	$p = 0.984$	$p = 0.644$	$p = 0.159$	n. a.
tup	*4*	$p = 0.081$	n. a.	n. a.	$p = 0.542$	n. a.
Dox-A2	*4*	**$p = 0.034$**	$p = 0.745$	n. a.	n. a.	n. a.
Catsup	*4*	$p = 0{,}067$	$p = 0.369$	n. a.	$p = 0.172$	$p = 0.917$
stc	*4*	*$p = 0.151$*	$p = 0.829$	n. a.	n. a.	$p = 0.405$
Ddc	*4*	**$p = 0.022$**	$p = 0.909$	n. a.	n. a.	n. a.
Lim3	*4*	**$p = 0.002$#**	n. a.	n. a.	$p = 0.431$	n. a.
E1F-1α48D	*5*	$p = 0.629$	$p = 0.742$	n. a.	n. a.	n. a.
magu	*5*	$p = 0.339$	n. a.	$p = 0.325$	$p = 0.153$	$p = 0.926$

* - Candidate genes showing a statistically significant association ($p < 0.05$) are in bold.

n. a. - Crosses without allelic segregation for the selected polymorphism.

significant after applying the sequential Bonferroni correction

We used the Mann-Whitney test to verify if the means associated to each genotype are significantly different, and this led us to assume a genetic model that explains the dominance of an allele over the other one (Table 6).

The difference observed between extreme classes ranges from 6.7 (for *Dox-A2*) to 23.1 days (for *Cat*) and, assuming the given dominance model for the different alleles, it

explains a minimum of 1.3 % (for the gene *hep*) and a maximum of 11.1 % (for *dilp2*) of the total variation regarding lifespan. For the AA cross, in every case, the allele that is associated with short lifespan always comes from the strain H5 (Table 6). Using only the markers for genes showing a significant association, 22.4 % of the lifespan variation observed in this cross is explained (Pearson correlation coefficient = 0.558; R^2 = 0.311; Adjusted R^2 = 0.224; N = 72). From the studied genes *dilp2* seems to be the one that explains a higher percentage of the phenotypic variation observed in these crosses (11.1 % on the AA cross and 8.5 % on the NA cross), and *hep* seems to be the gene with the lower significant effect on the lifespan variation (1.3 %).

Table 6 – Summary of the crosses showing significant associations.

Gene	Chr	F_2 Cross*	0/0	1/0	1/1	DBEC	R2 %
hep	*X*	H5 x W11 (0/0 1/1)	49.0	-	61.8	12.8 (26.1 %)	1.3
dFOXO	*2*	H5 x W11 (1/1 0/0)	62.3	50.9	49.0	13.3 (27.1 %)	10.0**
filamin	*2*	H5 x W11 (1/0 1/1)	36.7	56.6	59.0	22.3 (60.8 %)	9.3**
		W29 x O57 (1/0 1/1)	54.7	62.0	44.6	17.4 (39.0 %)	4.3***
dilp2	*3*	H5 x W11 (0/0 1/1)	39.8	55.3	60.9	21.1 (53.0 %)	11.1**
		O57 x H5 (1/0 0/0)	44.1	54.2	-	10.1 (22.9 %)	8.5
Cat		H5 x W11 (1/1 0/0)	60.4	55.1	37.3	23.1 (61.9 %)	9.9***
SOD	*3*	O57 x H5 (1/1 1/0)	-	45.1	54.2	9.1 (20.2 %)	7.0
PTEN	*4*	W29 x O57 (0/0 1/1)	48.7	58.2	-	9.5 (19.5 %)	6.8
Dox-A2	*4*	H5 x W11 (0/0 1/1)	52.0	58.7	-	6.7 (12.9 %)	3.4
Ddc	*4*	H5 x W11 (0/0 1/1)	51.4	59.6	-	8.2 (16.0 %)	4.9
Lim3	*4*	H5 x W11 (1/0 0/0)	60.6	50.0	-	10.6 (21.2 %)	8.4

Chr - Chromosome

DBEC – difference between extreme classes (difference between the extreme classes in percentage)

* Between brackets is shown the genotype of the F_0 individuals used

** Assuming that 1 is dominant over 0

*** Assuming that 0 is dominant over 1

4.3 Gene sequence variation

Using the data of the sequencing of the genomes of two *D. americana* strains (H5 and W11) it is possible to align the protein sequences, using the annotated *D. virilis* genome as a reference (http://flybase.org). This can be particularly useful in the case of genes showing a significant association in the F_2 individuals of the cross involving these two strains, where it is possible to gain insight about possible causative amino acid substitutions that could explain the observed phenotypic variation. However, from the eight genes that showed a significant association in this cross, only in three (*hep*, *Lim3* and *dilp2*) we have found evidences of amino acid substitutions between these two strains.

Lim3 shows three amino acid polymorphisms between H5 and W11 (Fig. 9): one derived conservative (a Valine by an Alanine at position 102) and one non-conservative (a Proline by a derived Alanine at position 540) amino acid changes in strain H5, and one derived non-conservative amino acid substitution in strain W11 (a Proline by a Serine at position 536).

SEQUENCE NAME	102	417	485	536	540	545
lim3 W11	V	S	T	S	P	A
lim3 H5	A	S	T	P	A	A
GJ17996 Dvir	V	T	N	P	P	S

Figure 9 – Amino acid polymorphic sites observed in Lim3 protein between the strains H5 and W11 and *D. virilis*. The shown numbers represent the position of the replaced amino acid relatively to *D. virilis*.

In Hep sequence we have identified four amino acid polymorphisms (Fig. 10), although many more could be present since this region of the genome is not well represented in either the H5 or W11 strains. Nevertheless, there is one non-conservative change in strain W11 (an Alanine by a derived Threonine at position 123), and two conservative substitutions (an Alanine by a derived Valine at position 703 and by a derived Glycine at position 1028) and one non-conservative substitution (a Glutamine by a derived Leucine at position 864) in strain H5.

SEQUENCE NAME	123	703	864	1028
hep W11	T	A	Q	A
hep H5	A	V	L	G
GJ16904 Dvir	A	A	Q	A

Figure 10 – Amino acid polymorphic sites observed in Hep protein between the strains H5 and W11. The shown numbers represent the position of the replaced amino acid relatively to *D. virilis*.

DILP2 is a well conserved protein across *Drosophila* species [102, 112]. Nevertheless, we observed one amino acid polymorphism (a non-conservative substitution of an Alanine by a derived Threonine at position 20 in strain W11; Fig. 11). However, the further sequencing of the individuals used to establish the F_2 association crosses have shown that this amino acid polymorphism is not segregating in any of the five crosses used (Fig. 12).

SEQUENCE NAME	20	40	60
dilp2 W11	T	T	D
dilp2 H5	A	T	D
GJ12130 Dvir	A	A	E

Figure 11 – Amino acid polymorphic sites observed in DILP2 protein between the strains H5 and W11 and *D. virilis*. The shown numbers represent the position of the replaced amino acid relatively to *D. virilis*.

Despite the fact that there are a few nucleotide changes in some of the F_0 individuals (Fig. 12), all of them represent synonymous substitutions corresponding to the same amino acid and, therefore, there is no evidence of amino acid polymorphisms segregating in the five crosses used in this work.

SEQUENCE NAME	85	135	150	151	162	165	279	297	306	322	333	357	402	417	423	426
AAF0male	C	A	C	A	C	A	C	T	C	A	C	C	C	T	T	C
AAF0female	C	A	C	A	C	A	C	T	C	A	C	C	C	T	T	C
ATF0male	C	A	C	A	C	A	C	T	C	A	C	C	C	T	T	C
ATF0female	C	A	C	A	C	A	C	C	C	A	C	C	C	T	C	C
TNF0male	C	A	C	A	C	A	C	C	C	A	C	C	C	C	T	C
TNF0female	C	A	C	A	C	A	C	Y	C	A	C	C	Y	Y	T	C
NAF0male	C	A	C	A	C	A	C	Y	C	A	C	C	Y	Y	T	C
NAF0female	C	A	C	A	C	A	C	T	C	a	C	C	C	T	T	C
TTF0male	C	A	C	A	C	A	C	C	C	A	C	C	C	T	C	C
TTF0female	C	A	C	A	C	A	C	C	C	A	C	C	C	C	T	C
VIRcds	T	G	T	G	T	G	A	C	T	T	G	T	C	G	T	T

Figure 12– Nucleotide polymorphic sites observed in *dilp2* gene sequences obtained by sequencing of the F_0 individuals of the five crosses used in the association study. The shown numbers represent the position of the substitution relatively to the *D. virilis* CDS. Y = C or T.

4.4 Gene expression analysis

In addition to the presence of the above described amino acid polymorphisms it could be the case that the phenotypic variation observed between populations is due to differences in gene expression levels of candidate genes. The same reasoning could also be done for the genes that do not show evidence of amino acid variation. Therefore, in this study we looked for gene expression levels of the ten candidate genes that showed a significant association in the F_2 association study, in male individuals with 0, 10, 30 and 60 days-old from the strains H5 and W11 (Fig. 13).

A general overview on the data reveals that gene expression levels vary throughout time, and the fold-change between the two strains is also very variable along the several time points studied. However, for the majority of the cases the fold-change values are below 2-fold and, therefore, in these cases it cannot be said that there are significant differences in gene expression between the two strains.

filamin and *PTEN* have a similar gene expression pattern; which is lower at 0 days and then rises at 10 days and goes dropping in the subsequent time points in the strain H5. When looking to the W11 expression levels for these two genes, we find that it gradually lowers throughout the studied time points. *dilp2* and *dFOXO* also have a similar expression pattern with an increase at the 10 days time point and then a decrease in 60 days-old flies. These latter four genes show a minimum expression level at the 60 days time point.

SOD has a growing expression level in the strain W11, whereas, in the strain H5 the level is lower in new-born flies and then it rises at the 10 days time point and remains virtually constant throughout time. *Cat* expression level rises until 30 days and then it goes down between the last two time points in the strain H5, whereas in the strain W11 the expression values start dropping earlier, between 10 and 30 days.

When looking to the remaining genes we cannot observe a clear pattern in the gene expression levels. In the case of *Lim3* it can be seen that the two strains have an opposing behavior (increasing in H5 and decreasing in W11 until 30 days, shifting in the last time point). The other genes have a gene expression level almost stable throughout the studied time points, with a maximum expression level observed at 30 days for *hep*, and the minimum observed for *Ddc* in 10 days-old flies. *Dox-A2* has a similar expression pattern as *Lim3* for the strain W11, but when looking to the strain H5 we can see that relative expression levels remain constant.

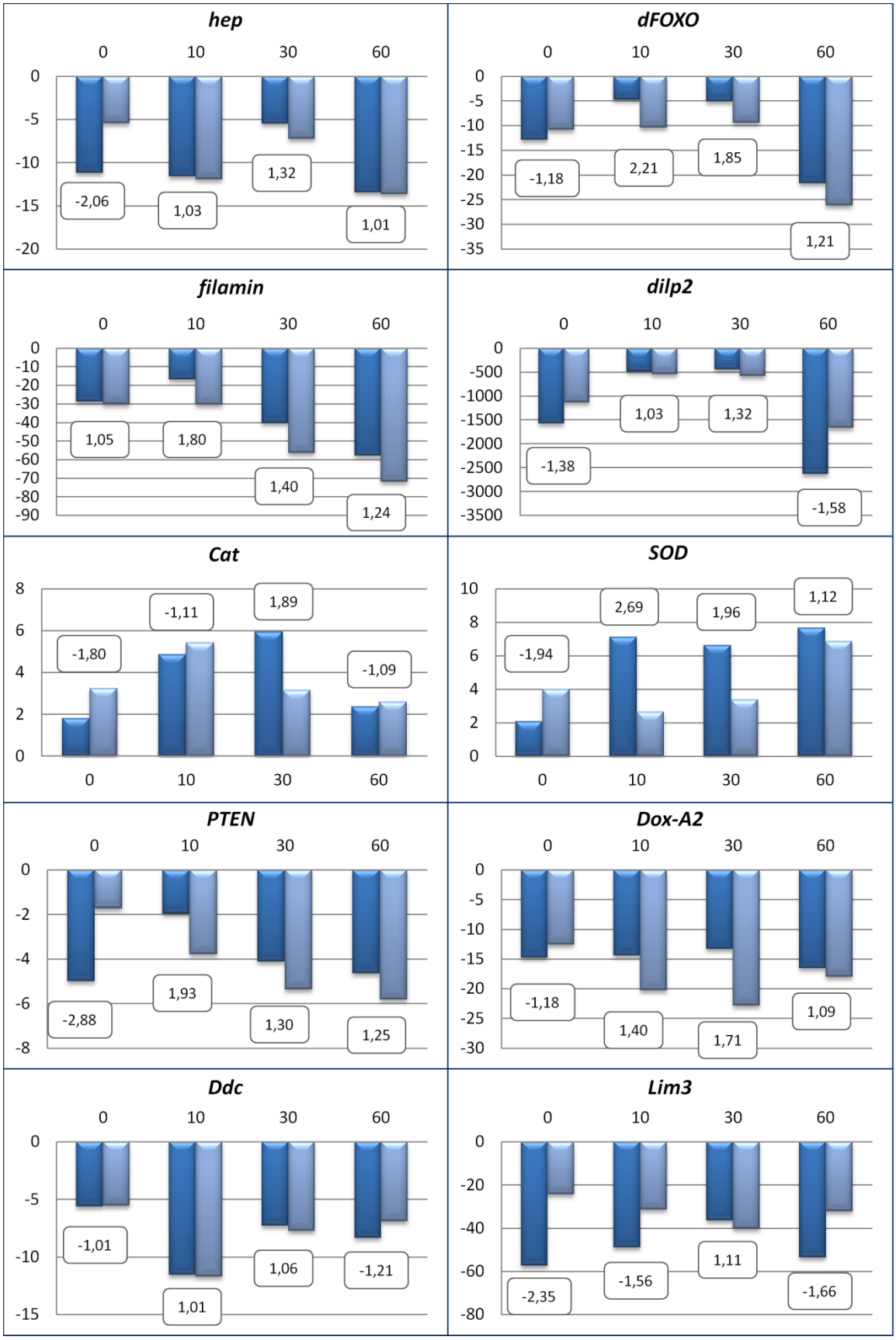

Figure 13 – Gene expression data for the ten candidate genes studied in the strains H5 (dark blue) and W11 (light blue). On the y-axis is represented the expression level relative to the control gene (*RpL32*). On the x-axis is represented the age of the individuals (in days). Inside the boxes is shown the fold-change value between the two *D. americana* strains (H5 relatively to W11).

Analyzing the fold-change between the two studied strains we can see that, for the majority of the genes and time points, the values are not significantly different (below 2-fold) and therefore, we can say that expression levels in those cases are quite the same. This situation occurs for all time points in *filamin*, *dilp2*, *Cat*, *Dox-A2* and *Ddc*, with the latter one showing the lesser variation between the two strains. In the case of *hep*, *PTEN* and *Lim3* we observe that the fold-change value is maximum in new-born flies (these genes are more than 2-fold less expressed in H5 strain relatively to W11) and then it tends to decrease and the expression levels of the two strains match up. In *dFOXO* and *SOD* we observe the highest fold-change between H5 and W11 at the 10 days time point. In both cases gene expression is more than 2-fold higher in H5 than in W11 at this point, and at the 30 days time point the fold-change values are also very close to 2-fold. At 60 days, gene expression is the same on both strains and, in new-born flies the gene expression is lower in the H5 strain, being the fold-change very close to two in the case of *SOD*.

5. DISCUSSION

Limited lifespan is a universal phenomenon, controlled by genetic and environmental factors whose interactions both limit and generate variation in this trait between individuals, populations and species. To understand the genetic architecture of longevity it is necessary to know what loci affect variation in lifespan, what are the allelic effects at these loci and what molecular polymorphisms define these alleles [163].

In *D. melanogaster*, longevity mechanisms have been well studied and several candidate genes have been pointed out to explain lifespan differences within individuals. Nevertheless, it remains unknown if this same set of genes can be used to explain lifespan variation in other distantly related species. Therefore, it is imperative to perform studies and to verify to what extent the results obtained for model species can be generalized to other species.

5.1 F_2 association experiment

The high level of recombination experienced by the *D. americana* genome (Morales-Hojas *et al.*, unpublished results) make it difficult to find associations between randomly selected markers and a multigenic phenotypic trait, as is the case for lifespan. Therefore, F_2 association crosses have to be established between different strains in order to overcome this issue, and their progeny is used to find associations between segregating alleles and the phenotype.

In this work, we performed five different association crosses with moderate-sized samples rather than a single association study with a larger number of individuals since our main goal was to infer the location of common variants with a large effect on the phenotypic trait, because these are the most probable ones that could explain most of the phenotypic variation that is found in natural populations. The strains used to establish the crosses were chosen in order to have an almost complete representation of the distribution of *D. americana*. We used one strain (O57) from the north of the distribution, two strains (H5 and W11) from the centre with the chromosomal rearrangements characteristic from the north of the distribution, and two strains (W29 and W46) with the typical chromosomal rearrangements from the south of the distribution (Supplementary table 1).

For all the five crosses established the F_2 males' lifespan seems to be normally distributed, and the mean lifespan of the individuals is very approximate with the value obtained when a sample of unrelated individuals is used (unpublished results). Another interesting fact is that, in general, lifespan of individuals proceeding from crosses between closely related individuals (AA and TT crosses) live less than the individuals resulting from the cross between parents of different *D. americana* populations (AT, NA and TN crosses). This situation is consistent with the heterosis and inbreeding depression theory that postulates that crosses between genetically dissimilar parents often creates superior offspring, and crosses between closely-related individuals tend to produce inferior ones due to the accumulation of deleterious alleles inside the population, leading to an heterozygote advantage [199]. The only cross that contradicts this hypothesis is the NA cross, whose F_2 individuals show the lowest mean lifespan. This observation could be explained because the strains used to establish this cross are those that on average, live less (data not shown). Therefore, the individuals resulting from this cross possess a series of genetic features that contributes to their lower lifespan.

Since chromosomal rearrangements can suppress recombination in heterozygotes contributing to the maintenance of linkage disequilibrium between alleles, thus creating associations between variants from genes located far away from each other [200], we looked for associations between the markers for the various chromosomal rearrangements present in *D. americana* and lifespan. We only found an association between lifespan and the *X/4* fusion – *Xc* inversion and the *4ab* inversion in one cross (Table 4), and probably this observation is due to the influence of a gene affecting lifespan that is located at the basis of the *X* chromosome or in the proximity of the markers used for the rearrangements genotyping. Therefore, it seems more likely that the presence of these chromosomal rearrangements is not affecting the setting of lifespan and that the latitudinal cline observed for them is not being maintained by selection acting on lifespan.

5.2 Candidate gene approach

A candidate gene approach can be a useful tool to determine if the same set of genes explains phenotypic variation in distantly related species. In *D. melanogaster*, life-history variation in natural populations is already well characterized, however it is unknown if the mechanisms observed in this species are the same responsible for the natural variation

occurred in other species. It has been already shown that *D. americana* can be widely used as model species in comparative studies [171, 173-175]. This species has also high average recombination rates making it a good model for performing phenotype-genotype association studies.

Using the data from the recent sequencing of the *D. americana* genome, markers have been developed to distinguish naturally occurring polymorphisms in the candidate genes, between the two sequenced strains. Initially, the amino acid sequence of the *D. melanogaster* candidate genes was obtained on Flybase and used to find the orthologous genes in *D. virilis*. From the 12 sequenced genomes that are available [100], this species is the one that is more closely-related with *D. americana*, therefore, the utilization of this species' sequences facilitates the achievement of the correct *D. americana* gene sequences in our database.

All *D. americana* candidate gene sequences were annotated and a polymorphism was selected, in a way that allows the discrimination between the two strains. Any kind of polymorphism can be used, but in this work only single nucleotide polymorphisms (SNP) were used that could then be differentiated by PCR and by using specific restriction enzymes that did or did not recognize the polymorphic sequence.

When performing an association study, it is expected that markers that are genetically linked to a locus influencing longevity will segregate more often with a certain value of the trait (higher or lower lifespan), whereas unlinked markers will not show any significant association with the phenotype [165]. Due to genetic linkage it is also likely to find associations in genes located in the same chromosomal region and in genes located in regions of low recombination. However, segregation and recombination work against these associations by reducing linkage disequilibrium each generation [152, 201]. In cases where alleles at two or more loci have correlated effects on fitness, the alleles experience similar selective pressures and under these circumstances it could be advantageous to minimize recombination between these loci [202]. If these genes are not closely linked, a suppressor of recombination between these alleles would be beneficial because it would decrease the occurrence of recombinants between them [203, 204]. Maintenance of positive correlations among alleles is the essence of the co-adapted gene complex and, therefore, if a gene is causatively correlated with a particular trait it will always show an association with that trait.

In this study, it was possible to find associations in ten (*hep*, *dFOXO*, *filamin*, *dilp2*, *Cat*, *SOD*, *PTEN*, *Dox-A2*, *Ddc* and *Lim3*) out of the 21 candidate genes studied (Table 5). However, it must be noted that for genes or crosses where we do not observe a significant association in this sample, we cannot exclude their potential role because an association could have been observed if other association crosses or sample sizes had been used. In addition, it is possible that not all relevant variation present in the candidate genes is segregating in the five F_2 association crosses here performed. We have selected only one polymorphism that is segregating between the strains H5 and W11 and, as we can see, in some cases this particular polymorphism is not segregating in other crosses (Table 5 and Supplementary table 1). Moreover, the sample size used may not be the adequate to detect weak associations, and therefore, in some of the cases we cannot exclude the potential role of the genes in the setting of lifespan.

Filamin is an actin-binding protein that binds to several cell-membrane proteins and intercellular ligands involved in signal transduction, and appears to act as a downstream effector in remodeling of the actin cytoskeleton [205]. It has an important role on the development of actin cytoskeletal structures and has several interactions with potential relevance to aging studies, in particular with Presenilins and with Toll and Tube proteins, affecting immune-response and aging-associated disease [206].

It has been shown that reducing the levels of a subset of the DILPs, by ablation of DILP2, 3 and 5-producing median Neural Stem Cells (mNSCs) in the *pars intercerebralis* of the brain late in the final larval instar, leads to a series of phenotypes including, among others, the extension of median and maximal lifespan [108]. Of all the DILPs produced by the mNSCs, DILP2 is currently thought to be the most important one. It is the most highly expressed, the most potent growth stimulator [109] and has been suggested to play a prominent role in lifespan extension by reduced IIS [93, 130].

FOXOs are a conserved family of transcription factors that are phosphorylated and inactivated in response to IIS, and they regulate various functions including stress responses and metabolism [207]. Therefore, *dFOXO* is negatively regulated by upstream IIS, and its overexpression in fat body under normal nutrition leads to extension of lifespan in female flies [97, 110].

PTEN is a lipid phosphatase with a function in the organization of the actin cytoskeleton [99]. PTEN promotes nuclear localization of endogenous dFOXO (that is

involved in the regulation of the insulin receptor signaling pathway) and thus changes in *PTEN* expression levels or function can interfere with insulin signaling [93, 95-97].

Due to the importance of the insulin/Igf signaling (IIS) pathway in the control of a great variety of physiological mechanisms, including longevity, in a great variety of organisms, from worms to mammals [79-83], it is not surprising that these genes that play a crucial role in the functioning of the IIS pathway may have an effect on lifespan. Genes belonging to pathways that directly interact with the IIS signaling pathway are also likely to play an important role on lifespan. This is the case of *hep*, which is part of a mitogen-activated protein (MAPK) cascade, together with Jun-N-terminal Kinase (JNK), JNK kinase (JNKK) and the JNK Basket (Bsk). In *Drosophila*, JNK confers tolerance to oxidative stress and extends life span by inducing a gene expression program [129]. Interestingly, it has been found that JNK also antagonizes IIS systematically by activating *dFOXO* and downregulating the expression of *dilp2* in insulin-producing cells (IPCs) [130], thereby, *hep* overexpression leads to an extension of lifespan by increased JNK signaling [129].

One of the most well accepted theories of aging is the free radical (or oxidative stress) hypothesis, proposed by Harman [5]. According to this hypothesis aging results from the imbalance between the formation and detoxification of reactive oxygen species (ROS) generated during normal cellular metabolism that leads to the accumulation of oxidative damage to cellular macromolecules, including DNA, proteins and lipids. In this study we observed an association with two genes encoding antioxidant enzymes, Superoxide dismutase (SOD) and Catalase (Cat). These two enzymes act in tandem to remove superoxide anions ($O_2^{\cdot-}$) and hydrogen peroxide (H_2O_2); and it is believed that, together, they provide the main enzymatic antioxidative defenses in insects. There is experimental evidence for their involvement in the determination of *D. melanogaster* adult lifespan by overexpression of both genes [208]. Lifespans of the experimental flies were found to be extended up to 34%, providing direct support for the involvement of reactive oxygen metabolites in the aging process.

Dox-A2 (*Diphenol oxidase A2*) encodes phenol oxidase, a protein involved in metabolism of catecholamines, which are essential for hardening and coloring of the cuticle (it catalyzes transformation of dopamine to dopaquinone, required for pigment formation). It is also known as *Rpn3* and is involved in proteolysis since it is one of the

regulatory particles non-ATPase that composes the 26S proteasome and is capable of binding substrates for degradation. *Ddc* encodes dopa decarboxylase, catalyzing the second reaction in the catecholamine biosynthesis pathway, transformation of dopa to dopamine, the basic catecholamine involved in nerve impulse transfer. *Lim3* is a homeobox gene that encodes a RNA polymerase II transcription factor required for development and function of neurons. QTL analyses and deficiency complementation tests suggested that naturally occurring allelic differences in these genes play a role in longevity control in *D. melanogaster* [209, 210].

However, and as said before, an association study tends to identify candidate loci rather than specific alleles that really explain the phenotypic variation. Therefore, it is likely to find associations in genes that are physically close to each other, not only by its direct influence in the trait but often due to their proximity to the causative locus. This could be the case of the genes *lim3* and *Ddc* that are very close to each other, and when looking to a wider region, the case of *Dox-A2* (Fig. 8). Definitely, there is something located on that region of the 4^{th} chromosome that could be influencing the setting of lifespan in this particular cross (AA). Intriguingly, there are no significant associations observed for the genes that are located very close to *Dox-A2*, as is the case of *stc*, *Catsup* or even *tup*. However, for *tup* and *catsup* the observed *p* value is very close to 0.05, therefore one could speculate that we could find an association in these genes if a larger sample was used.

Another possible explanation for this lack of associations is that, despite the F_0 individuals being homozygous for a certain marker (such as *mei-41* or *ms(2)35Ci*) they could be heterozygous regarding a causative variant (i.e. the marker could not be in linkage disequilibrium with the causative variant, thus it would have a genotype with a bimodal distribution regarding the mean lifespan values and therefore, an association will not be observed). On the other hand, it could be the case that the founding individuals of the cross are heterozygous regarding the maker used (as happens with *tup*, *Catsup* or *stc*) thus sharing an allele that in fact could not be equivalent because despite having the same restriction pattern they could be in linkage with different causative variants not showing a statistical association. On this way, the same causative variant could be in linkage with different alleles of the marker used that is heterozygous in one of the strains resulting in the same outcome.

All these genes referred before are in the *4th* chromosome and, in *D. americana* there is an *X/4* fusion that could also have an impact in these associations. Because in this study only males were used, they will always receive the free ("unfused") *4th* chromosome from the father and another *4th* chromosome (either fused or unfused depending on the strain) from the mother. On these cases, there could be a distortion in the data because the chromosome proceeding from the father is always the same and, on the F_2 individuals, one of the genotypic classes is absent.

While it is conceivable that for some of the genes that we observed a significant association it could be a gene on the vicinity of the candidate gene the one harboring the variation affecting lifespan, for two of the studied genes (*dilp2* and *filamin*) we observed associations in more than one cross, what could be a strong indication that, at least for these genes, the observed association is not an indirect one. In fact, for *filamin*, associations were observed in two completely independent crosses, what could mean that this gene may have an important role on the setting of lifespan.

When looking for the statistics of these two genes on the AA cross (Table 6) it is interesting to see that the average lifespan for each genotypic class and the difference between extreme classes of these cases are very similar. In fact, it is very intriguing why two genes in different chromosomes show the same pattern (*dilp2* is on the *2nd* while *filamin* is on the *3rd* chromosome). However, this fact could be easily explained because in *D. americana* there is a fixed fusion between these two chromosomes and, therefore, these findings could be due to the linkage between the two regions, not excluding, however, the influence of each of these genes individually. In fact, the linkage hypothesis is supported by the finding that there are only 10 (out of 75) recombinant individuals between these two genes. In addition, we see an almost perfect linkage disequilibrium between *dilp2* and the more closely located gene, *Cat* (only 7 recombinant individuals in a total of 86). In contrast with these observations we cannot find linkage between these genes and any other gene located in these two chromosomes, either those more closely located (as *SOD, fwd* or *puc*) or those more distant (like *dFOXO*). This could be due to the fact that there may exist a hotspot of recombination in the regions between *filamin* and *fwd* and between *Cat* and *SOD*, and therefore recombination occurred in these regions breaking the linkage between them. It could also be the case that an event of crossing-over occurred in these fused chromosomes breaking the linkage between the upper (telomere) and the bottom

(centromere) regions. *dFOXO* is more distantly located from *filamin*, near to the telomere of the chromosome, therefore it is very likely that recombination events occurred in this wider region so it is very probable that linkage has been disrupted between these two genes.

Lifespan is a phenotypic trait governed by multiple genes and interactions with small effect [159]. All these factors combined contribute to the setting of this complex trait as a whole. In this work, we used the Mann-Whitney test to verify if the means associated to each genotype are significantly different, and this led us to assume a genetic model that explains the dominance of an allele over the other one and by doing linear regression analysis we can find out how much of the phenotypic variation that is observed could be explained by the allelic variation on a certain gene (Table 6).

Given these models we were able to observe that, in our study *hep* is the gene that explains the lowest percentage of variation in the AA cross (1.3 %). On the other hand, the genes that explain the highest percentage of phenotypic variation in the AA cross are *dilp2* and *dFOXO* (11.1 and 10 %, respectively). This is a quite interesting result since it is known that these two genes interact with each other [93, 97, 98], and the IIS signaling pathway in which they are involved is thought to be one of the most important pathways influencing lifespan [11].

The fact that all these genes with a statistically significant association only explain a low percentage of the phenotypic variation observed in lifespan is in agreement with the assumption that longevity is a complex trait governed by multiple interactions and by multiple genes with a small effect each one contributing with a small percentage for the whole. However, and as said before, with this study we cannot exclude the influence of the genes that did not show a significant association on the setting of lifespan in *D. americana*. On the other hand, we cannot also restrict the influence on lifespan of a certain gene only in the cross where an association was observed. In fact, it is possible that all the studied genes could have (or not) an influence in the setting of the lifespan of this species; however the genes showing an association are those that more likely have a stronger importance for this phenotype.

5.3 Gene sequence analysis

Investigations of natural genetic variation at target genes can be a useful tool on the identification of genic elements that contribute to phenotype evolution, as well as elucidate important dynamics within pathways and characterize genic functionality on a fine scale. Genetic manipulations in the laboratory typically demonstrate functions of whole genes; however the function of smaller regions and even specific nucleotides can be resolved by evaluating natural alleles if there is sufficient variation in the wild, where natural selection can impose subtle pressure over many generations [164]. In *D. melanogaster* there is evidence that nucleotide variation affects longevity phenotypes and natural allelic variation has been characterized in several loci identified as aging genes, so far (see Paaby and Schmidt [24]).

In this study we were able to find associations between lifespan and three (*filamin*, *Cat* and *PTEN*) out of the seven genes with evidence for positive selection at the amino acid level and three (*Ddc*, *Dox-A2* and *Lim3*) out of the seven *D. melanogaster* genes showing associations between naturally occurring variation and lifespan (using deficiency complementation tests) in *D. americana*. However, this set of genes is not more likely to show associations than a random sample of candidate genes (Fisher's exact test; $p > 0.05$).

Using the data from the *D. americana* genome sequencing it is possible to gain insight into whether there is an amino acid mutation on the candidate gene that is causing the observed difference in lifespan. By doing this, we were able to find evidence of amino acid substitutions between the two strains' genomes in three (*hep*, *Lim3* and *dilp2*) out of the seven genes that showed a significant association in the cross in which they are involved, although natural genetic variation at other sites may contribute to the associations we observe. However, gene sequence must always be addressed in the individuals used in the association study because the sequence represented in the genome database does not represent the entire population. Indeed, we have done the sequencing of the entire coding region of *dilp2* and the amino acid substitution that we observed between the sequences of the H5 and W11 genomes was not present in any of the individuals used to establish the five F_2 association crosses. Therefore, this mutation in this gene is not likely to be the one responsible for the differences in lifespan observed in our study.

We only did the sequencing of the gene *dilp2* because it is a very short one and because this gene is thought to be one of the most important genes in the setting of lifespan

because of its function in the IIS signaling. We found it very interesting that this so well conserved gene [102, 112] possessed an amino acid substitution between two *D. americana* strains. However, and despite the fact that it is absent in our association study, it is very likely that this polymorphism does exist in natural populations of *D. americana* and we can speculate that it could affect the protein function and play a role in the setting of lifespan in nature.

In *D. melanogaster*, a handful of candidate aging genes have already been characterized regarding its natural allelic variation. For example, Paaby *et al.* [164] evaluated allelic variation at two members of the IIS signalling: the *Insulin-like Receptor* (*InR*) and its substrate, *chico*, in natural populations of *D. melanogaster*. They found that *InR* shows evidence of positive selection and clines in allele frequency across latitude in two continents (North America and Australia) while *chico* exhibits neutral patterns of evolution. These authors identified a candidate adaptive polymorphism at *InR* that appears to be functionally significant, providing new characterization of genic regions of functionality within *InR*, and probably, it is a component in a set of genes and traits (affecting lifespan) that respond adaptively to climatic variation.

Ddc has been identified as an aging candidate gene by QTL analysis and deficiency complementation tests [211]. Linkage disequilibrium mapping using 173 alleles from a natural population have revealed the existence of three common molecular polymorphisms in this gene, that are being maintained by balancing selection, accounting for 15.5 % of the genetic contribution to variance in lifespan from chromosome *2* [157].

In 2006, Carbone *et al.* [212] sequenced 169 alleles of the gene *Catsup* in a single *D. melanogaster* natural population and identified 33 polymorphisms with little linkage disequilibrium. From these, one is associated with longevity and results in a leucine-valine replacement in the third N-terminal transmembrane helix of the protein, being a probable casual variant affecting this complex trait.

However, in *D. americana* this is the first study that has been carried out regarding this issue. In the future, the entire gene region of these candidate genes (especially those showing a statistically significant association) should be sequenced in several strains in order to get a greater comprehension about the nature of the allelic variation existing in these genes and try to identify the possible causative polymorphisms affecting lifespan.

5.4 Gene expression analysis

Despite the evidence that, at least for some of the studied genes, there could be amino acid variation causing the phenotypic differences observed, changes in gene expression level could also be responsible for the observed differences in lifespan. In *D. melanogaster* several studies have identified a series of candidate genes that extend lifespan by decreased or increased gene expression activity (reviewed by Paaby and Schmidt [24]). However, the evaluation of the effects of gene expression changes on lifespan has been achieved through mutational analysis and, therefore, does not necessarily represent what happens in nature.

On this way, in the present work we decided to look for the gene expression levels of the genes for which we observed a significant association. We used individuals with different ages from two *D. americana* strains in order to verify if there are differences in gene expression throughout life that could explain the differences observed in longevity between the strains.

We decided to study the strains H5 and W11, not only because these are the two strains for which we have the complete genome sequenced making it easier to develop highly specific primers, but also because we have experimental evidence that these two strains show great differences regarding the mean lifespan of their individuals. Individuals from the W11 strain usually develop slower and live longer than individuals from the strain H5 (data not shown). Supporting this evidence is the fact that, in our association study, when we looked to mean lifespan of the individuals of each genotype, we saw that individuals that live less are usually those carrying the alleles present on the parent coming from the strain H5 (Table 6).

We also decided to look for gene expression levels in individuals of different ages in order to verify if the possible existing differences are constant throughout life or if there are differences in a certain time point that could be the responsible for the differences in lifespan. We chose to study new-born individuals to look if there are differences in gene expression early in life that could be influencing the entire adult phase and determine lifespan. We also decided to study flies with 10 days-old, a time point where flies are fully mature young adults, and at 30 days where we can consider flies as “middle-aged”. The last time point was chosen to be 60 days because several flies do not reach this age,

especially those from strains with shorten lifespan (as is the case of H5), and therefore these can be considered old flies.

In our study we observed that gene expression levels vary throughout time, and the fold-change between the two strains is also very variable along the several time points studied. However, for some genes (like *filamin*, *dilp2*, *Cat*, *Dox-A2* and *Ddc*) fold-change values are always below 2-fold and, in other cases like *hep*, *Dox-A2*, *Ddc* and *Lim3*, despite some variations, expression levels are very similar throughout life. These results are very interesting because, in *D. melanogaster*, has been already shown that nearly 20 % of the genome changes with age [161] and in our study we see that almost every gene has variations in expression during lifetime. This is not unexpected since all of the studied genes are candidate aging genes and the most interesting of them are expected to exhibit differences in expression both between strains and between young and old flies [161].

In *D. melanogaster*, it has been shown that for genes *Cat*, *filamin*, *hep*, *dFOXO*, *SOD* and *PTEN* high levels of expression are associated with an extension of lifespan [93, 97, 108, 129, 150, 208, 213]. In the present work, we observed that for three of these genes (*hep*, *SOD* and *PTEN*), in *D. americana*, alleles associated with long lifespan (those coming from strain W11) show higher expression levels at 0 days than those alleles associated with short lifespan (those coming from strain H5). However, at 10, 30 and 60 days of life, alleles associated with long lifespan show similar or lower expression levels than alleles associated with a short lifespan. These results may suggest that what happens early in adult life may influence lifespan more strongly than what happens late in life. In *D. melanogaster*, it has already been suggested that much of the variability to be expressed in lifespan may be present in latent form very early in life [214].

When looking to expression patterns throughout time of each gene we can see that for most of the cases what happens in *D. americana* is very similar to what has been described for *D. melanogaster* (see the High-Throughput Expression Data available at http://flybase.org/).

In the case of *hep*, expression levels are almost constant along life with a slight increase until 30 days (no further data is shown at FlyBase). In our data we see that *hep* expression levels are variable along life, but at 30 days they are higher than in the other time points in H5 and just slightly lower than at 0 days in W11 strain. Elevated expression levels of *hep* have been associated with an increase in stress resistance and longer lifespan

in *D. melanogaster* [129]. In our study, and as said before, at 0 days individuals from W11 strain have 2-fold higher expression levels of this gene than those from H5 strain what could have an important effect on the stress resistance and the setting of these individuals' lifespan, because in the remaining time point expression levels are equal between strains.

dFOXO (as other genes involved in the IIS pathway) has its maximum expression in *D. melanogaster* in the larval stages (due to the importance of this pathway on growth, [79]); in the adulthood what is seen is that expression levels are low in the beginning, then they rise and go down again late in life. In this work we saw a very similar pattern in both *D. americana* strains, with an increase between 0 and 10 days and a decrease between 30 and 60 days. This is very similar to what is observed in *dilp2* gene expression levels, another gene involved in the IIS signaling. On this case the data from FlyBase shows an increase in expression levels until 30 days, the same that we observe in our study. This increase in the expression levels of these genes early in life could be associated with the period of sexual maturation (which we know that in *D. americana* occurs during the first five days; unpublished results), and that has been shown to require high levels of IIS signaling in several species including *D. melanogaster* [215, 216]. In the late stages of life a decrease in metabolism could be responsible for the age-associated physiological decline that is observed [9].

In *D. melanogaster* has been shown that high levels of *dFOXO* are responsible for an increase in stress resistance and in lipid metabolism and a decrease in IIS signaling resulting in longer lifespan, whereas for *dilp2* the observations are the opposite [93, 97, 108]. In the present work, we did not observe any significant differences in *dilp2* gene expression between the two strains, while for *dFOXO* we saw, at 10 days, that H5 individuals have a 2.2-fold higher gene level and at 30 days this fold change is 1.8. This is an unexpected result since alleles coming from the H5 strain are those associated with a shorter lifespan. This observation could be due to a cause other than lifespan. It is known that *dFOXO* has multiple interactions with other pathways and regulates several other biological functions such as response to stress, nutrient levels and starvation, regulation of insulin receptor signaling pathway, regulation of growth, response to bacterium, or rhythmic process [89, 91-93, 98, 217-219] that directly or indirectly influence lifespan.

SOD is a gene involved in stress resistance that has been found to extend lifespan when gene activity is increased [220]. In *D. melanogaster*, high-throughput expression

analysis revealed an increasing level of expression of this gene until 30 days, a pattern similar to what we see in *D. americana* W11 strain, despite a slight decrease between 0 and 10 days. For H5 strain we see more variable expression levels; however they remain almost constant in 10, 30 and 60 days time-points. According to what is observed in *D. melanogaster* it was expected that long-lived strains show higher levels of *SOD* expression, however in our study this is the case only when we look to new born flies. In 10 and 30 days-old flies we see that those from H5 strain have higher *SOD* expression than those from W11 strain. This increment in *SOD* levels in the strain H5 very early in life are very intriguing and unexpected, and we can speculate that this strain could have, for an unknown reason, an elevated production of ROS or a deregulated stress response mechanism that could be harmful for cells and be responsible for the lower lifespan of these individuals, however further studies need to be done in order to confirm this issue.

As already seen, *Cat* is also involved in the oxidative stress response along with *SOD* [208]. According to the *D. melanogaster* high-throughput expression data present in the FlyBase database, *Cat* expression levels during life are not very variable, with an increase in the first five days of life. This is compatible to what we observe in our work, where an increase in *Cat* expression was observed until the 10^{th} day in the strain W11, followed by a not very pronounced decrease in expression levels till the 60^{th} day. In the H5 strain, we observed that the expression values also increased between 10 and 30 days, decreasing only after this time point. The behavior of the expression of this gene between the two studied strains is very similar to what is observed in *SOD* (except at the 10 days time point). Thereby, and due to the known interaction between these two genes [208], we can do the same reasoning as we did for the influence of *SOD* in lifespan determination on these *D. americana* strains. However, this hypothesis is just one speculation and, in fact, in our study, gene expression differences between the two strains are never higher than 2-fold and therefore they could not be very significant.

filamin and *PTEN* both have a role in the cytoskeleton organization and play important roles in immune response and cell proliferation [99, 206, 221] and, according to Flybase, have a decreasing but almost constant gene expression levels across life. These data are consistent to what we observe in our study, especially for W11, and for *filamin* this decrease although gradual is quite marked. Landis *et al.* [150] and Hwangbo *et al.* [93] have shown that an increase in expression of these genes leads to an increment in

D. melanogaster lifespan. In our study we see that *filamin* expression levels are very similar between the two strains in all time points and for *PTEN* we see a very low level of gene expression in new born flies of H5 strain when comparing to W11 (fold-change of -2.88). This observation reinforces what has been said before that what happens early in adult life may be influencing lifespan more strongly than what happens late in life. These genes have also a role in the immune system and, in fact, there is growing evidence that the molecular pathways underlying immune responses and lifespan are interlinked and it has been hypothesized that the immune system plays a major role in aging and lifespan determination [161, 222, 223].

Dox-A2, *Ddc* and *Lim3* play roles in very different physiological processes and have been identified as aging candidate genes by QTL analysis [157, 159]. These genes show no significant variations in gene expression throughout life and, in fact, an integrated microarray gene expression and QTL mapping analysis performed by Lai *et al.* [161] have not shown differences in expression of these genes between two different *D. melanogaster* lines and between young and old flies. Our results are generally in line with these observations. The most intriguing observation is that in new-born flies *Lim3* is more than 2-fold less expressed in individuals from the H5 strain than in those from the W11 strain. Accordingly, Rybina and Pasyukova [210] have shown that, in *D. melanogaster*, a naturally occurring polymorphism in the regulatory region of *Lim3* is responsible for a 25 % change in lifespan and that markers associated with long lifespan and intermediate *Lim3* expression are present at high frequencies in the population. We can speculate that, the very low levels of *Lim3* expression in H5 strain at 0 days could be responsible for its lower longevity when comparing to other strains, such as W11, that could possibly show intermediate levels of *Lim3* expression and, therefore, live longer.

Our inability to detect differences in gene expression in some of the studied candidate genes could be due to the fact that some of these genes are more importantly expressed at a different developmental period, differences in transcript abundance that are too small to be detected given the sample size and the method used, or because the genes are not regulated at the level of transcription [161]. Nevertheless, incorporation of knowledge about variation in transcript abundance, allelic variation and the data proceeding from QTL or association studies will greatly increase our comprehension about how candidate genes in QTL regions or identified by mutational analysis are affecting the

determination of lifespan in natural populations. However, and due to the high complexity and pleiotropic nature of this trait, its full characterization would only be achieved on a genome-wide level.

6. CONCLUDING REMARKS

The complexity of the aging process makes it difficult to understand the mechanisms responsible for the setting of lifespan. Most of the present knowledge has been obtained throughout mutational analysis using model species such as *D. melanogaster*. Therefore, little is known about whether the observed genetic variation at these loci occurs in natural populations, or even if this type of variation would contribute to the phenotypic variance for lifespan observed in nature. It also remains unknown if the same set of genes described for model species can explain lifespan variation in other distantly related species. To understand the genetic architecture of longevity it is necessary to know what loci affect variation in lifespan, and which of these loci harbor naturally occurring allelic variation and what polymorphisms actually affect the trait. The analysis of the causal relationships between gene structure, transcription level and lifespan will provide insight into conserved regulatory pathways controlling lifespan.

A candidate gene approach can be a useful tool to determine if the same set of genes explains phenotypic variation in distantly related species. In this study, ten out of 21 candidate genes showed a statistically significant association, what could mean that the molecular basis of life span variability is at least partially conserved between these distantly related species.

Association mapping can be useful on the identification of genes affecting a quantitative trait, as is the case of lifespan; however it provides no information as if the polymorphisms distinguishing the parental lines are the functional ones or whether they are rare or common in the population from which the parental lines are derived. The availability of two *D. americana* genomes will greatly facilitate the identification of the actual causative mutations in these genes. In this work we observed putative amino acid changes in two of the genes with significant associations that could be the responsible for the observed phenotypic differences between strains.

Nevertheless, expression level changes in these genes could also be responsible for the observed differences in lifespan. Genes that exhibit changes in transcript abundance with age are biomarkers of aging, and candidate genes for investigation by mutational analysis. Furthermore, genes showing differences between strains are also excellent candidate genes to explain lifespan differences, particularly if they are located inside

QTLs. In our study we found differences between "young" and "old" flies in five (*dFOXO*, *filamin*, *dilp2*, *SOD* and *PTEN*) of the genes showing significant associations with some of them showing differences between the two strains in some of the time points analyzed. This could mean that differences in different life stages could be important for the differences in lifespan, and what happens early in adult life may influence lifespan more strongly than what happens late in life as is suggested by the differences observed in *hep*, *SOD*, *PTEN* and *Lim3* expression levels in new-born flies and because of the absence of differences in all genes expression levels in 60 days-old flies.

However, DNA sequence variation or gene expression levels do not affect quantitative traits directly, but does so throughout networks of intermediate molecular phenotypes. Understanding the relationship between DNA sequence variation, transcriptional, protein and metabolite networks and phenotypes at the organism level is the main challenge for the future and will add the missing biological context to genotype–phenotype associations.

Therefore, in the future it would be interesting to perform association studies using larger samples or a larger number of generations to look if more recombination events would cause a weakening or a strengthening on the observed associations. The complete sequencing and gene expression level characterization of all the candidate genes in more strains would also be useful to understand the effects of particular amino acid polymorphisms and gene expression changes on the differences that are observed between different strains and individuals.

QTL, proteomics and mutational analyses will greatly increase our knowledge about the mechanisms underlying lifespan in *D. americana*. However, and due to the high complexity and pleiotropic nature of this trait its full characterization will only be achieved on a genome-wide level.

7. REFERENCES

1. Speakman JR (2005) Body size, energy metabolism and lifespan. J Exp Biol 208: 1717-1730.

2. Rubner M (1908) Das problem det lebensdaur und seiner beziehunger zum wachstum und ernarnhung. Munich: Oldenberg.

3. Pearl R (1922) The biology of death. Philadelphia: J. B. Lippincott.

4. Pearl R (1928) The rate of living. London: University of London Press.

5. Harman D (1956) Aging: a theory based on free radical and radiation chemistry. J Gerontol 11: 298-300.

6. Sohal RS, Weindruch R (1996) Oxidative stress, caloric restriction, and aging. Science 273: 59-63.

7. Beckman KB, Ames BN (1998) The free radical theory of aging matures. Physiol Rev 78: 547-581.

8. Helfand SL, Rogina B (2003) Genetics of aging in the fruit fly, *Drosophila melanogaster*. Annu Rev Genet 37: 329-348.

9. Helfand SL, Rogina B (2003) Molecular genetics of aging in the fly: Is this the end of the beginning? Bioessays 25: 134-141.

10. Helfand SL, Inouye SK (2002) Rejuvenating views of the ageing process. Nat Rev Genet 3: 149-153.

11. Kenyon CJ (2010) The genetics of ageing. Nature 467: 622-622.

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

13. Hughes KA (2010) Mutation and the evolution of ageing: from biometrics to system genetics. Philos Trans R Soc Lond B Biol Sci 365: 1273-1279.

14. Katic M, Kahn CR (2005) The role of insulin and IGF-1 signaling in longevity. Cell Mol Life Sci 62: 320-343.

15. Kirkwood TBL, Shanley DP (2005) Food restriction, evolution and ageing. Mech Ageing Dev 126: 1011-1016.

16. Medawar PB (1946) Old age and natural death. Modern Quart 1: 30-56.

17. Medawar PB (1952) An unsolved problem of biology. London, UK: H. K. Lewis.

18. Williams GC (1957) Pleiotropy, natural selection, and the evolution of senescence. Evolution 11: 398-411.

19. Hamilton WD (1966) The moulding of senescence by natural selection. J Theor Biol 12: 12-45.

20. Eyre-Walker A (2006) The genomic rate of adaptive evolution. Trends Ecol Evol 21: 569-575.

21. Vijg J, Campisi J (2008) Puzzles, promises and a cure for ageing. Nature 454: 1065-1071.

22. Kirkwood TBL (2002) Evolution of ageing. Mech Ageing Dev 123: 737-745.

23. Sezgin E (2004) Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. Genetics 168: 923-931.

24. Paaby AB, Schmidt PS (2009) Dissecting the genetics of longevity in *Drosophila melanogaster*. Fly 3: 29-38.

25. Ito K, Hotta Y (1992) Proliferation pattern of postembryonic neuroblasts in the brain of *Drosophila melanogaster*. Dev Biol 149: 134-148.

26. Arking R (1991) Biology of aging: Observations and principles. Englewood Cliffs, NJ: Prentice-Hall.

27. Fontana L, Partridge L, Longo VD (2010) Extending healthy life span - From yeast to humans. Science 328: 321-326.

28. De Loof A (2011) Longevity and aging in insects: Is reproduction costly; cheap; beneficial or irrelevant? A critical evaluation of the "trade-off" concept. J Insect Physiol 57: 1-11.

29. Lithgow GJ, White TM, Melov S, Johnson TE (1995) Thermotolerance and extended life-span conferred by single-gene mutations and induced by thermal stress. Proc Natl Acad Sci U S A 92: 7540-7544.

30. Tatar M, Khazaeli AA, Curtsinger JW (1997) Chaperoning extended life. Nature 390: 30-30.

31. Hercus M, Loeschcke V, Rattan S (2003) Lifespan extension of *Drosophila melanogaster* through hormesis by repeated mild heat stress. Biogerontology 4: 149-156.

32. Wang H-D, Kazemi-Esfarjani P, Benzer S (2004) Multiple-stress analysis for isolation of *Drosophila* longevity genes. Proc Natl Acad Sci U S A 101: 12610-12615.

33. Maynard Smith J (1958) The effects of temperature and of egg-laying on the longevity of *Drosophila subobscura*. J Exp Biol 35: 832-842.

34. Morrow G, Tanguay RM (2003) Heat shock proteins and aging in *Drosophila melanogaster*. Semin Cell Dev Biol 14: 291-299.

35. Tower J (2009) Hsps and aging. Trends Endocrinol Metab 20: 216-222.

36. Defays R, Gómez FH, Sambucetti P, Scannapieco AC, Loeschcke V, et al. (2011) Quantitative trait loci for longevity in heat-stressed *Drosophila melanogaster*. Exp Gerontol 46: 819-826.

37. Lamb MJ (1964) The effects of radiation on the longevity of female *Drosophila subobscura*. J Insect Physiol 10: 487-497.

38. Lamb MJ (1978) Ageing. In: Ashburner M, Wright T, editors. The Genetics and Biology of *Drosophila*. London: Academic.

39. Miquel J, Lundgren PR, Bensch KG, Atlan H (1976) Effects of temperature on the life span, vitality and fine structure of *Drosophila melanogaster*. Mech Ageing Dev 5: 347-370.

40. Denlinger DL (2002) Regulation of diapause. Annu Rev Entomol 47: 93-122.

41. Tatar M, Yin CM (2001) Slow aging during insect reproductive diapause: why butterflies, grasshoppers and flies are like worms. Exp Gerontol 36: 723-738.

42. Sohal RS, Buchan PB (1981) Relationship between physical activity and life span in the adult housefly, *Musca domestica*. Exp Gerontol 16: 157-162.

43. Iliadi KG, Iliadi NN, Boulianne GL (2009) Regulation of *Drosophila* life-span: Effect of genetic background, sex, mating and social status. Exp Gerontol 44: 546-553.

44. Vinogradov AE (1998) Male reproductive strategy and decreased longevity. Acta Biotheor 46: 157-160.

45. Carranza J, Pérez-Barbería FJ (2007) Sexual selection and senescence: Male size-dimorphic ungulates evolved relatively smaller molars than females. Am Nat 170: 370-380.

46. Bonduriansky R, Maklakov A, Zajitschek F, Brooks R (2008) Sexual selection, sexual conflict and the evolution of ageing and life span. Funct Ecol 22: 443-453.

47. Tatar M (2004) The neuroendocrine regulation of *Drosophila* aging. Exp Gerontol 39: 1745-1750.

48. Flatt T, Tu M-P, Tatar M (2005) Hormonal pleiotropy and the juvenile hormone regulation of *Drosophila* development and life history. Bioessays 27: 999-1010.

49. Keller L, Jemielity S (2006) Social insects as a model to study the molecular basis of ageing. Exp Gerontol 41: 553-556.

50. Rolff J (2002) Bateman's principle and immunity. Proc R Soc Lond B Biol Sci 269: 867-872.

51. Kirkwood TBL (1977) Evolution of ageing. Nature 270: 301-304.

52. Holliday R (1989) Food, reproduction and longevity: Is the extended lifespan of calorie-restricted animals an evolutionary adaptation? Bioessays 10: 125-127.

53. Rose MR, Bradley TJ (1998) Evolutionary physiology of the cost of reproduction. Oikos 83: 443-451.

54. Flatt T, Schmidt PS (2009) Integrating evolutionary and molecular genetics of aging. Biochim Biophys Acta 1790: 951-962.

55. Kotiaho JS, Simmons LW (2003) Longevity cost of reproduction for males but no longevity cost of mating or courtship for females in the male-dimorphic dung beetle *Onthophagus binodis*. J Insect Physiol 49: 817-822.

56. Reguera P, Pomiankowski A, Fowler K, Chapman T (2004) Low cost of reproduction in female stalk-eyed flies, *Cyrtodiopsis dalmanni*. J Insect Physiol 50: 103-108.

57. Davies S, Kattel R, Bhatia B, Petherwick A, Chapman T (2005) The effect of diet, sex and mating status on longevity in Mediterranean fruit flies (*Ceratitis capitata*), Diptera: Tephritidae. Exp Gerontol 40: 784-792.

58. Barnes AI, Boone JM, Jacobson J, Partridge L, Chapman T (2006) No extension of lifespan by ablation of germ line in *Drosophila*. Proc R Soc Lond B Biol Sci 273: 939-947.

59. Trevitt S, Partridge L (1991) A cost of receiving sperm in the female fruitfly *Drosophila melanogaster*. J Insect Physiol 37: 471-475.

60. Kuijper B, Stewart AD, Rice WR (2006) The cost of mating rises nonlinearly with copulation frequency in a laboratory population of *Drosophila melanogaster*. J Evol Biol 19: 1795-1802.

61. den Hollander M, Gwynne DT (2009) Female fitness consequences of male harassment and copulation in seed beetles, *Callosobruchus maculatus*. Anim Behav 78: 1061-1070.

62. Wigby S, Chapman T (2005) Sex peptide causes mating costs in female *Drosophila melanogaster*. Curr Biol 15: 316-321.

63. Kim Y-J, Bartalska K, Audsley N, Yamanaka N, Yapici N, et al. (2010) MIPs are ancestral ligands for the sex peptide receptor. Proc Natl Acad Sci U S A 107: 6520-6525.

64. Poels J, Van Loy T, Vandersmissen HP, Van Hiel B, Van Soest S, et al. (2010) Myoinhibiting peptides are the ancestral ligands of the promiscuous *Drosophila* sex peptide receptor. Cell Mol Life Sci 67: 3511-3522.

65. Piper MDW, Skorupa D, Partridge L (2005) Diet, metabolism and lifespan in *Drosophila*. Exp Gerontol 40: 857-862.

66. Chapman T, Partridge L (1996) Female fitness in *Drosophila melanogaster*: An interaction between the effect of nutrition and of encounter rate with males. Proc R Soc Lond B Biol Sci 263: 755-759.

67. Chippindale AK, Leroi AM, Kim SB, Rose MR (1993) Phenotypic plasticity and selection in *Drosophila* life-history evolution. I. Nutrition and the cost of reproduction. J Evol Biol 6: 171-193.

68. Finch CE, Ruvkun G (2001) The genetics of aging. Annu Rev Genom Hum Genet 2: 435-462.

69. Koubova J (2003) How does calorie restriction work? Genes Dev 17: 313-321.

70. Longo VD, Finch CE (2003) Evolutionary medicine: From dwarf model systems to healthy centenarians? Science 299: 1342-1346.

71. Oishi K, Shiota M, Sakamoto K, Kasamatsu M, Ishida N (2004) Feeding is not a more potent Zeitgeber than the light-dark cycle in *Drosophila*. Neuroreport 15: 739-743.

72. Mair W, Piper MDW, Partridge L (2005) Calories do not explain extension of life span by dietary restriction in *Drosophila*. PLoS Biol 3: e223.

73. Wong R, Piper MDW, Blanc E, Partridge L (2008) Pitfalls of measuring feeding rate in the fruit fly *Drosophila melanogaster*. Nat Methods 5: 214-215.

74. Marden JH, Rogina B, Montooth KL, Helfand SL (2003) Conditional tradeoffs between aging and organismal performance of Indy long-lived mutant flies. Proc Natl Acad Sci U S A 100: 3369-3373.

75. Simmons FH, Bradley TJ (1997) An analysis of resource allocation in response to dietary yeast in *Drosophila melanogaster*. J Insect Physiol 43: 779-788.

76. Rauser CL, Mueller LD, Rose MR (2004) Dietary restriction in *Drosophila*. Science 303: 1610-1612.

77. Finch CE (1990) Longevity, senescence, and the genome. Chicago: The University of Chicago Press.

78. Skorokhod A, Gamulin V, Gundacker D, Kavsan V, Muller IM, et al. (1999) Origin of insulin receptor-like tyrosine kinases in marine sponges. Biol Bull 197: 198-206.

79. Butler AA, Roith DL (2001) Control of growth by the somatropic axis: Growth hormone and the insulin-like growth factors have related and independent roles. Annu Rev Physiol 63: 141-164.

80. Saltiel AR, Kahn CR (2001) Insulin signalling and the regulation of glucose and lipid metabolism. Nature 414: 799-806.

81. Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloen A, et al. (2003) IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. Nature 421: 182-187.

82. Kimura KD, Tissenbaum HA, Liu Y, Ruvkun G (1997) *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. Science 277: 942-946.

83. Piper MDW, Selman C, McElwee JJ, Partridge L (2008) Separating cause from effect: how does insulin/IGF signalling control lifespan in worms, flies and mice? J Intern Med 263: 179-191.

84. Ruan Y, Chen C, Cao Y, Garofalo RS (1995) The *Drosophila* insulin receptor contains a novel carboxyl-terminal extension likely to play an important role in signal transduction. J Biol Chem 270: 4236-4243.

85. Bartke A (2001) Mutations prolong life in flies; implications for aging in mammals. Trends Endocrinol Metab 12: 233-234.

86. Kohn AD, Summers SA, Birnbaum MJ, Roth RA (1996) Expression of a constitutively active Akt Ser/Thr kinase in 3T3-L1 adipocytes stimulates glucose uptake and glucose transporter 4 translocation. J Biol Chem 271: 31372-31378.

87. Calera MR, Martinez C, Liu H, Jack AKE, Birnbaum MJ, et al. (1998) Insulin increases the association of Akt-2 with Glut4-containing vesicles. J Biol Chem 273: 7201-7204.

88. van der Heide Lars P, Hoekman Marco FM, Smidt Marten P (2004) The ins and outs of FoxO shuttling: mechanisms of FoxO translocation and transcriptional regulation. Biochem J 380: 297.

89. Gilley J, Coffer PJ, Ham J (2003) FOXO transcription factors directly activate *bim* gene expression and promote apoptosis in sympathetic neurons. J Cell Biol 162: 613-622.

90. Martinez-Gac L, Alvarez B, Garcia Z, Marques M, Arrizabalaga M, et al. (2004) Phosphoinositide 3-kinase and Forkhead, a switch for cell division. Biochem Soc Trans 32: 360-361.

91. Tran H, Brunet A, Grenier JM, Datta SR, Fornace AJ, et al. (2002) DNA repair pathway stimulated by the Forkhead transcription factor FOXO3a through the Gadd45 protein. Science 296: 530-534.

92. Brunet A, Sweeney LB, Sturgill JF, Chua KF, Greer PL, et al. (2004) Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. Sci Aging Knowl Environ 2004: or2-.

93. Hwangbo DS, 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.

94. McElwee JJ, Schuster E, Blanc E, Piper MD, Thomas JH, et al. (2007) Evolutionary conservation of regulated longevity assurance mechanisms. Genome Biol 8: R132.

95. Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, et al. (2001) Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. Science 292: 104-106.

96. Tatar M, Kopelman A, Epstein D, Tu M-P, Yin C-M, et al. (2001) A mutant *Drosophila* insulin receptor homolog that extends life-span and impairs neuroendocrine function. Science 292: 107-110.

97. Giannakou ME, Goss M, Jünger MA, Hafen E, Leevers SJ, et al. (2004) Long-lived *Drosophila* with overexpressed *dFOXO* in adult fat body. Science 305: 361-361.

98. Giannakou ME, Goss M, Partridge L (2008) Role of dFOXO in lifespan extension by dietary restriction in *Drosophila melanogaster*: not required, but its activity modulates the response. Aging Cell 7: 187-198.

99. von Stein W, Ramrath A, Grimm A, Müller-Borg M, Wodarz A (2005) Direct association of Bazooka/PAR-3 with the lipid phosphatase PTEN reveals a link between the PAR/aPKC complex and phosphoinositide signaling. Development 132: 1675-1686.

100. *Drosophila* 12 Genomes C, Clark AG, Eisen MB, Smith DR, Bergman CM, et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203-218.

101. Brogiolo W, Stocker H, Ikeya T, Rintelen F, Fernandez R, et al. (2001) An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. Curr Biol 11: 213-221.

102. Guirao-Rico S, Aguadé M (2011) Molecular evolution of the ligands of the insulin-signaling pathway: *dilp* genes in the genus *Drosophila*. Mol Biol Evol 28: 1557-1560.

103. Li C, Kim K (2008) Neuropeptides. WormBook: 1-36.

104. Wu Q, Brown MR (2006) Signaling and function of insulin-like peptides in insects. Annu Rev Entomol 51: 1-24.

105. Ikeya T, Galic M, Belawat P, Nairz K, Hafen E (2002) Nutrient-dependent expression of insulin-like peptides from neuroendocrine cells in the CNS contributes to growth regulation in *Drosophila*. Curr Biol 12: 1293-1300.

106. Okamoto N, Yamanaka N, Satake H, Saegusa H, Kataoka H, et al. (2009) An ecdysteroid-inducible insulin-like growth factor-like peptide regulates adult development of the silkmoth *Bombyx mori*. FEBS J 276: 1221-1232.

107. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591.

108. Broughton SJ, Piper MD, Ikeya T, Bass TM, Jacobson J, et al. (2005) Longer lifespan, altered metabolism, and stress resistance in *Drosophila* from ablation of cells making insulin-like ligands. Proc Natl Acad Sci U S A 102: 3105-3110.

109. Rulifson EJ (2002) Ablation of insulin-producing neurons in flies: Growth and diabetic phenotypes. Science 296: 1118-1120.

110. Giannakou ME, Partridge L (2007) Role of insulin-like signalling in *Drosophila* lifespan. Trends Biochem Sci 32: 180-188.

111. Hsu H-J, Drummond-Barbosa D (2009) Insulin levels control female germline stem cell maintenance via the niche in *Drosophila*. Proc Natl Acad Sci U S A.

112. Grönke S, Clarke D-F, Broughton S, Andrews TD, Partridge L (2010) Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. PLoS Genet 6: e1000857.

113. Broughton SJ, Alic N, Slack C, Bass T, Ikeya T, et al. (2008) Reduction of DILP2 in *Drosophila* triages a metabolic phenotype from lifespan revealing redundancy and compensation among DILPs. PLoS ONE 3: e3721.

114. Hafen E (2004) Interplay between growth factor and nutrient signaling: Lessons from *Drosophila* TOR. In: Thomas G, Sabatini DM, Hall MN, editors. TOR: Springer Berlin Heidelberg. pp. 153-167.

115. Davis RJ (2000) Signal transduction by the JNK group of MAP kinases. Cell 103: 239-252.

116. Wullschleger S, Loewith R, Hall MN (2006) TOR signaling in growth and metabolism. Cell 124: 471-484.

117. Kapahi P, Zid BM, Harper T, Koslover D, Sapin V, et al. (2004) Regulation of lifespan in *Drosophila* by modulation of genes in the TOR signaling pathway. Curr Biol 14: 885-890.

118. Montagne J, Stewart MJ, Stocker H, Hafen E, Kozma SC, et al. (1999) *Drosophila* S6 kinase: A regulator of cell size. Science 285: 2126-2129.

119. Um SH, D'Alessio D, Thomas G (2006) Nutrient overload, insulin resistance, and ribosomal protein S6 kinase 1, S6K1. Cell Metabolism 3: 393-402.

120. Baird SD, Turcotte M, Korneluk RG, Holcik M (2006) Searching for IRES. RNA 12: 1755-1785.

121. Johnson GL, Nakamura K (2007) The c-jun kinase/stress-activated pathway: Regulation, function and role in human disease. Biochim Biophys Acta 1773: 1341-1348.

122. Weston CR, Davis RJ (2007) The JNK signal transduction pathway. Curr Opin Cell Biol 19: 142-149.

123. McEwen DG, Peifer M (2005) Puckered, a *Drosophila* MAPK phosphatase, ensures cell viability by antagonizing JNK-induced apoptosis. Development 132: 3935-3946.

124. Boutros M, Agaisse H, Perrimon N (2002) Sequential activation of signaling pathways during innate immune responses in *Drosophila*. Dev Cell 3: 711-722.

125. Chen W (2002) Stimulus-specific requirements for MAP3 kinases in activating the JNK pathway. J Biol Chem 277: 49105-49110.

126. Geuking P, Narasimamurthy R, Lemaitre B, Basler K, Leulier F (2009) A non-redundant role for *Drosophila* Mkk4 and Hemipterous/Mkk7 in TAK1-mediated activation of JNK. Plos One 4: e7709.

127. Igaki T (2009) Correcting developmental errors by apoptosis: lessons from *Drosophila* JNK signaling. Apoptosis 14: 1021-1028.

128. Sabio G, Davis RJ (2010) cJun NH2-terminal kinase 1 (JNK1): roles in metabolic regulation of insulin resistance. Trends Biochem Sci 35: 490-496.

129. Wang MC, Bohmann D, Jasper H (2003) JNK signaling confers tolerance to oxidative stress and extends lifespan in *Drosophila*. Dev Cell 5: 811-816.

130. Wang MC, Bohmann D, Jasper H (2005) JNK extends life span and limits growth by antagonizing cellular and organism-wide responses to insulin signaling. Cell 121: 115-125.

131. Bokov A, Chaudhuri A, Richardson A (2004) The role of oxidative damage and stress in aging. Mech Ageing Dev 125: 811-826.

132. Haigis MC, Yankner BA (2010) The aging stress response. Mol Cell 40: 333-344.

133. Pérez VI, Bokov A, Remmen HV, Mele J, Ran Q, et al. (2009) Is the oxidative stress theory of aging dead? Biochim Biophys Acta 1790: 1005-1014.

134. Lapointe J, Hekimi S (2009) When a theory of aging ages badly. Cell Mol Life Sci 67: 1-8.

135. Back P, Matthijssens F, Vlaeminck C, Braeckman BP, Vanfleteren JR (2010) Effects of *sod* gene overexpression and deletion mutation on the expression profiles of reporter genes of major detoxification pathways in *Caenorhabditis elegans*. Exp Gerontol 45: 603-610.

136. Partridge L (2009) Some highlights of research on aging with invertebrates, 2009. Aging Cell 8: 509-513.

137. Phillips JP, Parkes TL, Hilliker AJ (2000) Targeted neuronal gene expression and longevity in *Drosophila*. Exp Gerontol 35: 1157-1164.

138. Bayne A-CV, Sohal RS (2002) Effects of superoxide dismutase/catalase mimetics on life span and oxidative stress resistance in the housefly, *Musca domestica*. Free Radical Biol Med 32: 1229-1234.

139. Dugan LL, Quick KL (2005) Reactive oxygen species and aging: Evolving questions. Sci Aging Knowl Environ 2005: pe20-.

140. Kirby K (2002) RNA interference-mediated silencing of *Sod2* in *Drosophila* leads to early adult-onset mortality and elevated endogenous oxidative stress. Proc Natl Acad Sci U S A 99: 16162-16167.

141. Phillips JP, Campbell SD, Michaud D, Charbonneau M, Hilliker AJ (1989) Null mutation of copper/zinc superoxide dismutase in *Drosophila* confers hypersensitivity to paraquat and reduced longevity. Proc Natl Acad Sci U S A 86: 2761-2765.

142. Sun J, Molitor J, Tower J (2004) Effects of simultaneous over-expression of Cu/ZnSOD and MnSOD on *Drosophila melanogaster* life span. Mech Ageing Dev 125: 341-349.

143. Schriner SE (2005) Extension of murine life span by overexpression of Catalase targeted to mitochondria. Science 308: 1909-1911.

144. Magwere T, West M, Riyahi K, Murphy MP, Smith RAJ, et al. (2006) The effects of exogenous antioxidants on lifespan and oxidative stress resistance in *Drosophila melanogaster*. Mech Ageing Dev 127: 356-370.

145. Lin Y (1998) Extended life-span and stress resistance in the *Drosophila* mutant *methuselah*. Science 282: 943-946.

146. Rogina B (2000) Extended life-span conferred by cotransporter gene mutations in *Drosophila*. Science 290: 2137-2140.

147. Cvejic S, Zhu Z, Felice SJ, Berman Y, Huang X-Y (2004) The endogenous ligand Stunted of the GPCR Methuselah extends lifespan in *Drosophila*. Nat Cell Biol 6: 540-546.

148. Ja WW, Carvalho GB, Madrigal M, Roberts RW, Benzer S (2009) The *Drosophila* G protein-coupled receptor, Methuselah, exhibits a promiscuous response to peptides. Protein Sci 18: 2203-2208.

149. Bieschke ET, Wheeler JC, Tower J (1998) Doxycycline-induced transgene expression during *Drosophila* development and aging. Mol Gen Genet 258: 571-579.

150. Landis GN, Bhole D, Tower J (2003) A search for doxycycline-dependent mutations that increase *Drosophila melanogaster* life span identifies the *VhaSFD*, *Sugar baby*, *filamin*, *fwd* and *Cctl* genes. Genome Biol 4: R8.

151. Li Y, Tower J (2008) Adult-specific over-expression of the *Drosophila* genes *magu* and *hebe* increases life span and modulates late-age female fecundity. Mol Genet Genomics 281: 147-162.

152. Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. London: Prentice Hall.

153. Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng Z-B, Mackay TFC (1997) Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. Proc Natl Acad Sci U S A 94: 9734-9739.

154. Vieira C, Pasyukova EG, Zeng Z-B, Hackett JB, Lyman RF, et al. (2000) Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. Genetics 154: 213-227.

155. Leips J, Mackay TFC (2000) Quantitative trait loci for life span in *Drosophila melanogaster*: Interactions with genetic background and larval density. Genetics 155: 1773-1788.

156. Leips J, Mackay TFC (2002) The complex genetic architecture of *Drosophila* life span. Exp Aging Res 28: 361-390.

157. De Luca M, Roshina NV, Geiger-Thornsberry GL, Lyman RF, Pasyukova EG, et al. (2003) Dopa decarboxylase (*Ddc*) affects variation in *Drosophila* longevity. Nat Genet 34: 429-433.

158. Pasyukova EG, Roshina NV, Mackay TFC (2004) *Shuttle craft*: a candidate quantitative trait gene for *Drosophila* lifespan. Aging Cell 3: 297-307.

159. Mackay TF, Roshina NV, Leips JW, Pasyukova EG (2006) Complex genetic architecture of *Drosophila* longevity. In: Masoro EJ, Austad SN, editors. Handbook of the Biology of Aging. Amesterdam: Elsevier. pp. 181-216.

160. Wayne ML, McIntyre LM (2002) Combining mapping and arraying: An approach to candidate gene identification. Proc Natl Acad Sci U S A 99: 14903-14906.

161. Lai C-Q, Parnell LD, Lyman RF, Ordovas JM, Mackay TFC (2007) Candidate genes affecting *Drosophila* life span identified by integrating microarray gene expression analysis and QTL mapping. Mech Ageing Dev 128: 237-249.

162. Schmidt PS, Paaby AB, Heschel MS (2005) Genetic variance for diapause expression and associated life histories in *Drosophila melanogaster*. Evolution 59: 2616-2625.

163. Geiger-Thornsberry GL, Mackay TFC (2004) Quantitative trait loci affecting natural variation in *Drosophila* longevity. Mech Ageing Dev 125: 179-189.

164. Paaby AB, Blacket MJ, Hoffmann AA, Schmidt PS (2010) Identification of a candidate adaptive polymorphism for *Drosophila* life history by parallel independent clines on two continents. Mol Ecol 19: 760-774.

165. Mackay TFC (2001) Quantitative trait loci in *Drosophila*. Nat Rev Genet 2: 11-20.

166. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9: 356-369.

167. Casa AM (2000) Inaugural Article: The MITE family Heartbreaker (Hbr): Molecular markers in maize. Proc Natl Acad Sci U S A 97: 10083-10089.

168. Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol 34: 275.

169. Gupta P, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomics 4: 139-162.

170. Henry RJ (2006) Plant conservation genetics. New York: Haworth Press.

171. Reis M, Vieira CP, Morales-Hojas R, Aguiar B, Rocha H, et al. (2011) A comparative study of the short term cold resistance response in distantly related *Drosophila* species: The role of *regucalcin* and *Frost*. PLoS ONE 6: e25520.

172. Russo CA, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol 12: 391-404.

173. Reis M, Sousa-Guimarães S, Vieira CP, Sunkel CE, Vieira J (2011) *Drosophila* genes that affect meiosis duration are among the meiosis related genes that are more often found duplicated. PLoS ONE 6: e17512.

174. Wittkopp PJ, Williams BL, Selegue JE, Carroll SB (2003) *Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes. Proc Natl Acad Sci U S A 100: 1808-1813.

175. Wittkopp PJ, Smith-Winberry G, Arnold LL, Thompson EM, Cooley AM, et al. (2011) Local adaptation for body color in *Drosophila americana*. Heredity 106: 592-602.

176. Caletka BC, McAllister BF (2004) A genealogical view of chromosomal evolution and species delimitation in the *Drosophila virilis* species subgroup. Mol Phylogen Evol 33: 664-670.

177. Morales-Hojas R, Vieira CP, Vieira J (2008) Inferring the evolutionary history of *Drosophila americana* and *Drosophila novamexicana* using a multilocus approach and the influence of chromosomal rearrangements in single gene analyses. Mol Ecol 17: 2910-2926.

178. Patterson JT, Stone WS (1952) Evolution in the genus *Drosophila*. New York: Macmillan.

179. Fonseca NA, Morales-Hojas R, Reis M, Rocha H, Vieira CP, et al. (2012) *D. americana* as a model species for comparative studies on the molecular basis of phenotypic variation. Genome Biol Evol submitted.

180. Hilton H, Hey J (1996) DNA sequence variation at the *Period* locus reveals the history of species and speciation events in the *Drosophila virilis* group. Genetics 144: 1015-1025.

181. Schäfer MA, Orsini L, McAllister BF, Schlotterer C (2006) Patterns of microsatellite variation through a transition zone of a chromosomal cline in *Drosophila americana*. Heredity 97: 291-295.

182. Morales-Hojas R, Reis M, Vieira CP, Vieira J (2011) Resolving the phylogenetic relationships and evolutionary history of the *Drosophila virilis* group using multilocus data. Mol Phylogen Evol 60: 249-258.

183. Spicer GS (1992) Reevaluation of the phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae). Ann Entomol Soc Am 85: 11-25.

184. Spicer GS, Bell CD (2002) Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA genes. Ann Entomol Soc Am 95: 156-161.

185. Throckmorton LH (1982) The *virilis* species group. In: Ashburner M, Novistky E, editors. The genetics and biology of *Drosophila*. London: Academic. pp. 227-297.

186. Vieira J, McAllister BF, Charlesworth B (2001) Evidence for selection at the *fused1* locus of *Drosophila americana*. Genetics 158: 279-290.

187. Vieira CP, Coelho PA, Vieira J (2003) Inferences on the evolutionary history of the *Drosophila americana* polymorphic *X/4* fusion from patterns of polymorphism at the *X*-linked *paralytic* and *elav* genes. Genetics 164: 1459-1469.

188. Vieira CP, Almeida A, Dias JD, Vieira J (2006) On the location of the gene(s) harbouring the advantageous variant that maintains the *X/4* fusion of *Drosophila americana*. Genet Res 87: 163.

189. McAllister BF (2002) Chromosomal and allelic variation in *Drosophila americana*: selective maintenance of a chromosomal cline. Genome 45: 13-21.

190. Hilton H, Hey J (1997) A multilocus view of speciation in the *Drosophila virilis* species group reveals complex histories and taxonomic conflicts. Genet Res 70: 185-194.

191. McAllister BF, Charlesworth B (1999) Reduced sequence variability on the *Neo-Y* chromosome of *Drosophila americana americana*. Genetics 153: 221-233.

192. McAllister BF, McVean GAT (2000) Neutral evolution of the sex-determining gene *transformer* in *Drosophila*. Genetics 154: 1711-1720.

193. Hsu TC (1952) Chromosomal variation and evolution in the *virilis* group of *Drosophila*. University of Texas Publications 5204: 35-72.

194. Evans AL, Mena PA, McAllister BF (2007) Positive selection near an inversion breakpoint on the Neo-*X* chromosome of *Drosophila americana*. Genetics 177: 1303-1319.

195. Fonseca NA, Vieira CP, Schlötterer C, Vieira J (2012) The DAIBAM MITE element is involved in the origin of one fixed and two polymorphic *Drosophila virilis* phylad inversions. Fly 6: 71-74.

196. Reis M, Vieira CP, Morales-Hojas R, Vieira J (2008) An old *bilbo*-like non-LTR retroelement insertion provides insight into the relationship of species of the *virilis* group. Gene 425: 48-55.

197. Filatov DA (2002) proseq: A software for preparation and evolutionary analysis of DNA sequence data sets. Mol Ecol Notes 2: 621-624.

198. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25: 402-408.

199. Darwin C (1868) The variation of animals and plants under domestication. London: J. Murray.

200. Dobzhansky T (1970) Genetics of the evolutionary process. New York,: Columbia University Press. ix, 505 p.

201. Futuyma DJ (1998) Evolutionary biology. Sunderland, MA: Sinauer Associates. xviii, 763, 749 p.

202. Sinervo B, Svensson E (2002) Correlational selection and the evolution of genomic architecture. Heredity 89: 329-338.

203. Feldman MW, Otto SP, Christiansen FB (1996) Population genetic perspectives on the evolution of recombination. Annu Rev Genet 30: 261-295.

204. Pepper JW (2003) The evolution of evolvability in genetic linkage patterns. BioSyst 69: 115-126.

205. Li MG, Serr M, Edwards K, Ludmann S, Yamamoto D, et al. (1999) Filamin is required for ring canal assembly and actin organization during *Drosophila* oogenesis. J Cell Biol 146: 1061-1074.

206. Guo Y, Zhang SX, Sokol N, Cooley L, Boulianne GL (2000) Physical and genetic interaction of filamin with presenilin in *Drosophila*. J Cell Sci 113: 3499-3508.

207. Baumeister R (2006) Endocrine signaling in *Caenorhabditis elegans* controls stress response and longevity. J Endocrinol 190: 191-202.

208. Orr W, Sohal R (1994) Extension of life-span by overexpression of superoxide dismutase and catalase in *Drosophila melanogaster*. Science 263: 1128-1130.

209. Roshina NV, Pasyukova EG (2007) Genes regulating the development and functioning of the nervous system determine life span in *Drosophila melanogaster*. Russ J Genet 43: 275-280.

210. Rybina OY, Pasyukova EG (2010) A naturally occurring polymorphism at *Drosophila melanogaster Lim3* locus, a homolog of human *LHX3/4*, affects *Lim3* transcription and fly lifespan. PLoS ONE 5: e12621.

211. Pasyukova EG, Vieira C, Mackay TFC (2000) Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. Genetics 156: 1129-1146.

212. Carbone MA, Jordan KW, Lyman RF, Harbison ST, Leips J, et al. (2006) Phenotypic variation and natural selection at *Catsup*, a pleiotropic quantitative trait gene in *Drosophila*. Curr Biol 16: 912-919.

213. Tatar M (2005) Comment on "Long-lived *Drosophila* with overexpressed *dFOXO* in adult fat body". Science 307: 675.

214. Flanagan JR (1980) Detecting early-life components in the determination of the age of death. Mech Ageing Dev 13: 41-62.

215. Sim C, Denlinger DL (2008) Insulin signaling and FOXO regulate the overwintering diapause of the mosquito *Culex pipiens*. Proc Natl Acad Sci U S A 105: 6777-6781.

216. Tennessen JM, Thummel CS (2011) Coordinating growth and maturation - Insights from *Drosophila*. Curr Biol 21: R750-R757.

217. Puig O, Marr MT, Ruhf ML, Tjian R (2003) Control of cell number by *Drosophila* FOXO: downstream and feedback regulation of the insulin receptor pathway. Genes Dev 17: 2006-2020.

218. Zheng X, Yang Z, Yue Z, Alvarez JD, Sehgal A (2007) FOXO and insulin signaling regulate sensitivity of the circadian clock to oxidative stress. Proceedings of the National Academy of Sciences 104: 15899-15904.

219. Alic N, Andrews TD, Giannakou ME, Papatheodorou I, Slack C, et al. (2011) Genome-wide dFOXO targets and topology of the transcriptomic response to stress and insulin signalling. Mol Syst Biol 7.

220. Spencer CC, Howell CE, Wright AR, Promislow DEL (2003) Testing an 'aging gene' in long-lived *Drosophila* strains: increased longevity depends on sex and genetic background. Aging Cell 2: 123-130.

221. Howell L, Sampson CJ, Xavier MJ, Bolukbasi E, Heck MMS, et al. (2012) A directed miniscreen for genes involved in the *Drosophila* anti-parasitoid immune response. Immunogenetics 64: 155-161.

222. Garsin DA, Villanueva JM, Begun J, Kim DH, Sifri CD, et al. (2003) Long-lived *C. elegans daf-2* mutants are resistant to bacterial pathogens. Science 300: 1921.

223. Doroszuk A, Jonker M, Pul N, Breit T, Zwaan B (2012) Transcriptome analysis of a long-lived natural *Drosophila* variant: a prominent role of stress- and reproduction-genes in lifespan extension. BMC Genomics 13: 167.

SUPPLEMENTARY MATERIAL

Supplementary table 1 – Genotypes of the F_0 individuals used to establish the five F_2 association crosses.

	H5 x W11		W11 x W46		W29 x O57		O57 x H5		W46 x W29	
	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀
Chromosomal markers*										
fused	0	0	0	1	1	0	0	0	1	1
bib	0	0	0	0	0	1	1/0	0	0	0
GJ17741	1	1	1	1	1	0	0	1	1	1
Candidate genes										
ovo	1	0	0	1	1	1	1	1	1	1
hep	0	1	1	0	1	1	1	0	0	1
dFOXO	1	0	0	1	1	0	0	1	1	1
puc	1/0	1/0	1/0	0	0	0	0	1/0	0	0
InR	1	1/0	1/0	1	1	1/0	1/0	1	1	1
fwd	0	1/0	1/0	0	0	0	0	0	0	0
filamin	1/0	1	1	1	1/0	1	1	0	1	0
dilp2	0	1	1	1	0	1/0	1/0	0	1	0
Cat	1	0	0	0	0	0	0	1/0	0	0
SOD	1/0	1	1	1	1	1	1	1/0	1	1
PTEN	1/0	0	0	1	0	1	1/0	1	1	0
mei-41	0	1	1/0	1/0	0	1/0	1/0	1/0	0	0
ms(2)35Ci	0	1	1	0	0	1/0	1	0	0	0
tup	1/0	0	0	0	0	0	0	1/0	0	0
Dox-A2	0	1	1/0	0	0	0	0	0	0	0
Catsup	1/0	0	0	1	0	0	1/0	1	1	0
stc	1/0	1/0	1/0	1/0	0	0	1/0	0	1/0	0
Ddc	0	1	1/0	0	0	0	0	0	0	0
Lim3	1/0	0	0	0	0	0	0	1	0	0
E1F-1α48D	0	1	1	0	0	0	0	0	0	0
magu	0	1	1	1	0	0	1	0	1	0

* *fused*: 0 – presence of the *X/4* fusion; 1 – absence of the *X/4* fusion

bib: 0 – absence of the *4ab* inversion; 1 – presence of the *4ab* inversion

GJ17741: 0 – presence of the *5b* inversion; 1 – presence of the *5a* inversion

Supplementary table 2 – Primers, PCR amplification conditions, size of the amplification products and restriction enzymes used in the F_2 association studies. Variable nucleotides are shown in bold.

Gene	Primers	Annealing Temperature	Product size	Restriction Enzyme	
				Name	Recognition Site
Cat	F 5' GGGCAACAGGTGGATAGG 3' R 5' GCTGGACGCTGAAAATAC 3'	55ºC	451bp	*Cla*I	A**T**CGAT
Catsup	F 5' TTATCAGTGCCGCTCCAT 3' R 5' GAAGTCGCCAATCTCGTG 3'	55ºC	632bp	*Pst*I	C**T**GCAG
Ddc	F 5' GTTTGTTCTGCGTCTGTA 3' R 5' GCTGTTCACTTCTTGTCC 3'	54ºC	369bp	*Rsa*I	GT**A**C
dFOXO	F 5' GCAAGCCCGTCTATCTGT 3' R 5' GTTTCCCTTCCCATTTAC 3'	50ºC	908bp	*Dde*I	CTG**A**G
dilp2	F 5' GGCGTTTCCTTCTCCTCA 3' R 5' TCGTTTGTTTGCTCCTTT 3'	54ºC	978bp	*Fok*I	GG**A**TG
Dox-A2	F 5' TGGTTCTGATTGGAGTTA 3' R 5' TTGATTTTTCTTTTCGTC 3'	48ºC	512bp	*Acc*I	GTATA**C**
EF-1a48D	F 5' AGCGATGGTGTGACTGAG 3' R 5' AGGCAAAAACGAAACTGG 3'	51ºC	645bp	*Msl*I	CAYNNNNRT**G**
filamin	F 5' GCGAACCAGTCTCTCAGT 3' R 5' TAGTCAAGGAATCAGCAT 3'	48ºC	599bp	*Dde*I	CTNA**G** (CTCA**G**)
fwd	F 5' GGTATTTCATTTTCTTTG 3' R 5' CGGTATCATTTCTCGTAG 3'	51ºC	257bp	*Bsr*DI	GCAA**T**G
hep	F 5' GGAAACGGACAAGAAACT 3' R 5' CAAAGCAGCCCAAACACT 3'	53ºC	290bp	*Bsr*DI	GCAA**T**G
InR	F 5' TGCTGACTGTTTGATTTT 3' R 5' AGGGCTATTCCACTATGA 3'	52ºC	449bp	*Ple*I	GACT**C**
Lim3	F 5' AGCAGCAGCAGCAACAAT 3' R 5' CAGCCACCCACAACGAGT 3'	62ºC	421bp	*Hha*I	GCG**C**
magu	F 5' GGGCAATAAACTAACCAT 3' R 5' ATCCTCCGACTTCCACAT 3'	53ºC	669bp	*Ava*I	CYCGR**G**
mei-41	F 5' ACCGTAAAGTCGTCAGTT 3' R 5' AAATGCTCAGTTCTCCAC 3'	53ºC	918bp	*Eco*RI	G**A**ATTC
ms(2)35Ci	F 5' CCTGCTTATGCTCTGATT 3' R 5' CCTGTCTGCCTGTTCCTC 3'	50ºC	442bp	*Ple*I	CT**C**AG
ovo	F 5' TGTTCTCGCCATCTTTCC 3' R 5' TCTTTGCCCGTTTTTGAC 3'	52ºC	280bp	*Bgl*I	GCCNNNNNG**G**C
PTEN	F 5' TCCTGCGTTGACATCTAA 3' R 5' GGTGCTTTTTCCCATTTT 3'	50ºC	727bp	*Ssp*I	A**AT**ATT
puc	F 5´ ACGAGAGCAAAAATGTCA 3´ R 5´ CAACAACAAACGGTCCAG 3´	51ºC	586bp	*Hha*I	GCG**C**
SOD	F 5' CACTTCAATCCCTACCAG 3' R 5' GTGCGTCCAATAATGCTA 3'	52ºC	160bp	*Bgl*I	G**C**CACCTGGGC
stc	F 5' ACTAACAAAACAAAGCAC 3' R 5' AAAACAACATACCAAAGC 3'	49ºC	467bp	*Acl*I	**A**ACGTT
tup	F 5' CGTGCGTGTTGCGGTCTA 3' R 5' TGCGGGCGTGTGCGTATT 3'	58ºC	764bp	*Nde*I	CAT**ATG**

Supplementary table 3 – Primers used for the gene sequencing analysis

Gene	Primer Sequence
dilp2	F: 5’ ATCAGTTCCAAAGCATAG 3’ R: 5’ CACAAAGCAAATACCATA 3’
M13	F (-20): 5’ GTAAAACGACGGCCAG 3 R: 5’ CAGGAAACAGCTATGAC 3’

Supplementary table 4 – Primers used for the gene expression analysis

Gene	Primer Sequence
hep	F: 5’ TCCGAAGAAACCAAAGTA 3’ R: 5’ CACCTCAAAGTCCGTATT 3’
dFOXO	F: 5’ GCCGTCTGAGTCCCATAC 3’ R: 5’ CTGTGCCTGCGTCATTGT 3’
filamin	F: 5’ TTGCTGCCCACCTTGCTC 3’ R: 5’ TGCCCCTGCCGAATCTGA 3’
dilp2	F: 5’ ACATTGAGGAGAAGGAAA 3’ R: 5’ CAGCGAGTTGAGGACATT 3’
Cat	F: 5’ GTCTGGTCGCACAAGGAA 3’ F: 5’ CGGACTGAACGCAATCTG 3’
SOD	F: 5’ CTGCCCCGTGAAGGTTAC 3’ R: 5’ CAGCCATTCGTGTTGTCG 3’
PTEN	F: 5’ TAACCCTCCTACCATTGA 3’ R: 5’ GCACGGCTACGACATTTA 3’
Dox-A2	F: 5’ CGTTTGATGCTTGACTCG 3’ R: 5’ ATCGGTGCTCTCCTTGCT 3’
Ddc	F: 5’ TCTACCTGAAGCACGACA 3’ R: 5’ CAGACGCAGAACAAACCA 3’
Lim3	F: 5’ AGGCGAAGGGTCTCTACT 3’ R: 5’ TGCGGTTTTCAGTGTCTC 3’
RpL32	F: 5’ ACAACAGAGTGCGTCGTC 3’ R: 5’ ATCTCCTTGCGTTTCTTC 3’