Natural Genetic Variation Underlying

UVR Sensitivity in *Drosophila melanogaster*

Thesis submitted for degree of

**Doctor of Philosophy**

at the University of Leicester

By

Kamaldeep Chana

Genetics department

University of Leicester

September 2014
Abstract

Solar ultraviolet radiation has been a major environmental factor throughout the evolution of life. Nature has evolved a plethora of defence mechanisms against this biologically harmful agent, and the genes underlying these mechanisms (e.g. DNA repair) have been the direct target of natural selection.

The main aim of this study was to assess the level of genetic diversity in *Drosophila*, and phenotypic variation in Ultraviolet Radiation (UVR) sensitivity amongst a naturally-derived panel of fly strains (DGRP), and to evaluate the extent of genetic variation underlying this trait. The DGRP strains were screened for UVR-sensitivity using simulated solar radiation, and substantial phenotypic variation was identified. A genome-wide association analysis detected 114 SNPs across all three major chromosomes, at a FDR of 0.0001. Some of these SNPs lie within known UV response genes (*Xpac, hay*), and some novel UVR-associated genes were identified (*TwdIB, raw*). Subsequently, a significant association, SNP 3R:12383617 in gene CG42342 was validated using complementation tests. In addition, enriched network analysis detected both novel UVR-associated pathways (e.g. GPCRs), and pathways wholly implicated in UV response.

In parallel, a candidate gene focussed approach was taken to identify natural phenotypic variation in an allelic series of *p53* strains (27 congeneric strains) derived from European populations. These strains exhibited both phenotypic and genotypic variation. Association analysis revealed single intronic SNP (3R:18875989), which was associated with UVR-induced oxidative stress.

Another set of experiments addressed the role of BTK, a kinase triggered by UVR, in ageing. Two BTK inhibitors (X and Y) were tested as potential novel anti-ageing drugs. The drugs were administrated at different doses in the food, and both extended lifespan, in a p53-dependent manner. The treated flies also showed a weight gain and improved motor function. Patents are currently pending for these two drugs in anti-ageing therapy.
Acknowledgements

I am extremely grateful to my supervisors Dr Eran Tauber and Dr Marcus Cooke, who have helped and guided me through this work, imparting knowledge and understanding. Having the opportunity to work across two labs did not only enhance this work, but has made my project truly enjoyable. I would also like to express my appreciation to Marcus for his continued help after leaving Leicester for sunnier shores.

My undergraduate supervisor Dr Raffael Schaffrath, ignited my passion for research, and entrusted a challenging project to a young researcher. This work both helped me decide what path to take in life and provided a step on the right direction.

Thank you, to all my friends and colleagues, antagonisers and drinking buddies in lab 124 for all the good times over the last four years. A big shout-out for Dr Emma Picot and Nathaniel Davies, for helping (and teaching me) with bioinformatics, and great scientific discussions. I thank Dr Mirko Pegoraro and Giorgio Fedele for all their help in lab.

I would like to acknowledge and thank our collaborators Mohammad Althubiti and his supervisor Dr Salvador Macip, Dept. of Biochemistry, University of Leicester. It has been thoroughly enjoyable to work in partnership with your lab and to increase the scope of my work.

I am truly grateful to my family for all their support, encouragement and love which have carried me through my seven years at Leicester. A special thank you to Balwinder and all my family in Leicester, for always being there for me!
Chapter 1: Introduction ............................................................................................................. 1
  1.1 Solar UVR as an environmental factor ............................................................................. 1
  1.2 Dynamics of solar UVR and the ozone layer ..................................................................... 4
  1.3 Role of UVR-sensitivity in Evolution ................................................................................. 9
  1.4 Targets of UVR damage ..................................................................................................... 12
    1.4.1 Resistance to solar UVR by Screening ........................................................................... 15
    1.4.2 Cellular repair mechanisms ......................................................................................... 16
    1.4.3 Behavioural mechanisms to avoid solar UVR damage ............................................... 20
  1.5 UVR-sensitivity as a complex phenotypic trait ................................................................. 20
    1.5.1 Quantitative Genetics of UVR resistance ................................................................. 21
  1.6 UVR-sensitivity and p53 pathways .................................................................................... 27
  1.7 p53 and ageing ................................................................................................................. 29
  1.8 Aims and objectives of this study ....................................................................................... 32

2. General Methods .................................................................................................................... 34
  2.1 Fly stocks and maintenance ............................................................................................... 34
  2.2 Mass production of maize meal food for vials, bottles and petri dishes ....................... 34
  2.3 Preparing and dispensing food with drug ........................................................................... 34

Chapter 3: Natural genetic polymorphism underlying variability in UVR-sensitivity in D. melanogaster ................................................................................................................... 36
  3.1 Introduction ......................................................................................................................... 36
  3.2 Methods ............................................................................................................................ 39
    3.2.1 Fly stocks ...................................................................................................................... 39
    3.2.2 Assay development for testing UVR-sensitivity ......................................................... 39
      Finding the LD_{50} of Canton-S in the embryo viability assay ............................................ 41
    3.2.3 Phenotypic screen of the DGRP for the GWAS ....................................................... 44
    3.2.4 Validation of candidate SNPs identified in the GWAS ............................................. 46
    3.2.5 Analysis of enriched pathways network and gene ontology ..................................... 49
  3.3 Results ............................................................................................................................. 50
    3.3.1 GWAS of UVR-sensitivity ......................................................................................... 50
    3.3.2 Validation of significant SNPs using complementation tests .................................... 55
    3.3.3 Enriched pathways network analysis and gene ontology ......................................... 56
  3.4 Discussion ........................................................................................................................ 59
Chapter 4: Natural genetic polymorphism in *Drosophila p53* underlying variation in UVR-sensitivity .......................................................... 64

4.1 Introduction ................................................................................................................. 64
4.2 Methods ......................................................................................................................... 66
  4.2.1 Fly stocks .................................................................................................................. 66
  4.2.2 Creating congenic strains carrying natural Dmp53 variants. ...................... 67
  4.2.3 Sequencing the p53 alleles ....................................................................................... 71
    4.2.3.1 DNA extraction from multiple flies ................................................................. 71
    4.2.3.2 DNA amplification by Polymerase Chain Reaction (PCR) ......................... 71
    4.2.3.3 Visualisation of DNA on an agarose gel ......................................................... 74
    4.2.3.4 Pooling DNA amplifications and sequencing .............................................. 74
  4.2.4 Assembling and aligning the sequences ............................................................... 74
  4.2.5 Phenotyping the congenic strains .......................................................................... 75
    4.2.5.1 Adult viability Assay (Solar simulator) ............................................................ 75
    4.2.5.2 UVB assays ....................................................................................................... 76
      Oxidative stress assay (UVB) .................................................................................. 78
      Real-time PCR assay (UVB) ................................................................................... 79
        RNA extraction ....................................................................................................... 79
        First-strand cDNA synthesis ............................................................................... 80
        Real-time PCR reaction ....................................................................................... 81
  4.2.6 Association analysis .............................................................................................. 82
    Testing for the signature of selection ........................................................................ 83
  4.2.8 Visualisation and analysis of 3D protein structure of Dmp53 ......................... 83
4.3 Results ......................................................................................................................... 84
  4.3.1 Adult viability Assay (Solar simulator) ................................................................. 84
  4.3.2 Oxidative stress assay (UVB) .............................................................................. 86
  4.3.3 Real-time PCR assay (UVB) ............................................................................... 88
  4.3.4 Association analysis ............................................................................................. 90
  4.3.5 Haplotype association ......................................................................................... 90
  4.3.6 Summary of SNPs associated with UVR-sensitivity ........................................ 92
  4.3.6 Analysis of Dmp53 sequence polymorphism .................................................... 93
4.4 Discussion .................................................................................................................. 99

Chapter 5: Increased longevity in *Drosophila* is induced by inhibitor of Bruton’s tyrosine kinase (BTK) and requires p53 ......................................................... 104

5.1 Introduction ............................................................................................................. 104
5.2 Methods ................................................................................................................... 108
5.2.1 Longevity assay ................................................................. 108
5.2.2 Climbing assay ................................................................. 108
5.2.3 Weight of flies after drug administration ................................. 109
5.3 Results .................................................................................. 110
5.4 Discussion ............................................................................. 117

Chapter 6: General discussion ......................................................... 121

Chapter 7: Appendices .................................................................. 127
7.1 Appendix for Chapter 4 ............................................................. 127
7.2 Appendix for Chapter 5 ............................................................. 137
7.3 Appendix for Chapter 6 ............................................................. 144

Chapter 8: Bibliography ................................................................. 146
# Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig.1.1</td>
<td>Conceptual graph showing different factors affecting the environment throughout evolution</td>
</tr>
<tr>
<td>Fig.1.2</td>
<td>A typical profile of ozone density and solar UV irradiance in the northern hemisphere</td>
</tr>
<tr>
<td>Fig.1.3</td>
<td>Global UV index map</td>
</tr>
<tr>
<td>Fig.1.4</td>
<td>Predictions of future UVR</td>
</tr>
<tr>
<td>Fig.1.5</td>
<td>Direct UVB damage resulting in thymine dimers</td>
</tr>
<tr>
<td>Fig.1.6</td>
<td>The production of 6,4-photoproducts from thymine</td>
</tr>
<tr>
<td>Fig.1.7</td>
<td>DNA-damage response (DDR) pathways</td>
</tr>
<tr>
<td>Fig.1.8</td>
<td>Allelic variation underlying UVR resistance in <em>CPD photolyase</em> in rice sp. and homologues</td>
</tr>
<tr>
<td>Fig.1.9</td>
<td>Map showing relationship between <em>p53</em> polymorphism and Latitude</td>
</tr>
<tr>
<td>Fig.1.10</td>
<td><em>p53Arg72</em> and <em>MDM2</em> frequencies correlated to temperature and Latitude...</td>
</tr>
<tr>
<td>Fig.1.11</td>
<td>The ‘Evolution Canyon’ in Israel</td>
</tr>
<tr>
<td>Fig.1.12</td>
<td>Hallmarks of ageing</td>
</tr>
<tr>
<td>Fig.3.1</td>
<td>Sample of SNPs in the DGRP strains</td>
</tr>
<tr>
<td>Fig.3.2</td>
<td>Dose response for UVR in <em>Canton-S</em> strain</td>
</tr>
<tr>
<td>Fig.3.3</td>
<td>Effect of stimulus duration (intensity 25mW/cm²)</td>
</tr>
<tr>
<td>Fig.3.4</td>
<td>Reproducibility of two strains in the embryo viability assay at a dose of 18KJ/m²</td>
</tr>
<tr>
<td>Fig.3.5</td>
<td>Variation in survivorship of embryos among DGRP strains (normalised scores)</td>
</tr>
<tr>
<td>Fig.3.6</td>
<td>Distribution of W values in the GWAS</td>
</tr>
<tr>
<td>Fig.3.7</td>
<td>Boxplots of significant SNPs associated with UVR-sensitivity</td>
</tr>
<tr>
<td>Fig.3.8</td>
<td>Complementation test of significant SNP 3R:12383617 in CG42342</td>
</tr>
<tr>
<td>Fig.3.9</td>
<td>Network of genes (N=400) for UVR-sensitivity</td>
</tr>
<tr>
<td>Fig.3.10</td>
<td>Network of genes (N=200) for UVR-sensitivity</td>
</tr>
<tr>
<td>Fig.4.1</td>
<td>Isoforms of <em>Dmp53</em></td>
</tr>
<tr>
<td>Fig.4.2</td>
<td>Geographical origin of isofemale strains</td>
</tr>
<tr>
<td>Fig.4.3</td>
<td>Amplification and sequencing <em>Dmp53</em></td>
</tr>
<tr>
<td>Fig.4.4</td>
<td>Adult viability after UVB irradiation</td>
</tr>
<tr>
<td>Fig.4.5</td>
<td>Validation of oxidative stress assay</td>
</tr>
<tr>
<td>Fig.4.6</td>
<td>Adult viability following solar irradiation</td>
</tr>
<tr>
<td>Fig.4.7</td>
<td>Oxidative stress in the congenic strains</td>
</tr>
<tr>
<td>Fig.4.8</td>
<td>Relative expression of <em>Dmp53</em> in congenic strains</td>
</tr>
<tr>
<td>Fig.4.9</td>
<td>UVR induced oxidative stress associated with SNP 4123 in <em>p53</em></td>
</tr>
</tbody>
</table>
Fig. 4.16 Phylogenetic tree of Dmp53 gene of the congenic strains

Fig. 4.19. Schematic of Dmp53 gene with UVR-sensitivity associated SNPs

Fig. 4.20 Tajima’s D along Dmp53

Fig. 4.21. Linkage disequilibrium across Dmp53

Fig. 4.22 3D model of Dmp53

Fig. 5.1 Survival curves for flies with drug X treatment

Fig. 5.2 Survival curves for flies with drug Y treatment

Fig. 5.3 Temporal analysis of drug intervention

Fig. 5.4 Survival curves for drug intervention in Dmp53 null mutant flies

Fig. 5.5 Motor function during drug X intervention

Fig. 5.6 Wet weight of flies after treatment with drug X

Fig. 7.1.1 Crosses for complementation tests

Fig. 7.1.2 Complementation tests of significant SNPs

Fig. 7.1.3 Variation in survivorship of embryos for DGRP strain under control and experimental conditions

Fig. 7.1.4 Association of Wolbachia pipientis infection with UVR-sensitivity

Fig. 7.1.5 Network of genes for UVR-sensitivity

Fig. 7.2.1 Larval viability in response to solar light

Fig. 7.2.2 UVI in Europe on 1st August 2014

Fig. 7.2.3. Lack of association of adult viability and p53 expression level with SNP 4213

Fig. 7.3.1 Survival curve for preliminary experiment
Table of tables

Table.2.1 Ingredients for maize food................................................................. 34
Table.3.1 SNPs identified in the GWAS with corresponding Deficiency strains .......... 48
Table.4.1 List of isofemale strains ........................................................................ 67
Table.4.2 Primers for real-time PCR ..................................................................... 81
Table.4.3 Haplotype sequence inferred by ML....................................................... 92
Table.4.4 Significant SNPs in Dmp53 associated with UVR response ..................... 93
Table.7.1.1 P-values for t-tests comparing independent embryo viability experiments ......................................................................................................................... 127
Table.7.1.2 DGRP strains used in the embryo viability assay for the GWAS .......... 128
Table.7.1.3 Crosses for quantitative complementation tests ................................. 129
Table.7.1.4 Significant associations detected in the GWAS ................................. 132
Table.7.1.5 Wolbachia pipientis infection status in DGRP strains ............................ 133
Table.7.1.6 PANTHER GO terms ......................................................................... 135
Table.7.1.7 Pathways identified using PANTHER.................................................. 136
Table.7.2.1 Primers for PCR and sequencing......................................................... 137
Table.7.2.2 SNP annotations (BDGP 5) for congenic strains ................................. 138
Table.7.2.3 SNPs in Dmp53 from European (congenic) strains .............................. 139
Table.7.2.4 Genes deleted/disrupted in Df(3R)BSC803 .......................................... 141
Table.7.3.1 Mann-Whitney tests with adjusted p-values ........................................ 143
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADI</td>
<td>Acceptable daily intake</td>
</tr>
<tr>
<td>°C</td>
<td>Degrees Celsius</td>
</tr>
<tr>
<td>BDGP</td>
<td>Berkeley Drosophila Genome Project</td>
</tr>
<tr>
<td>Df</td>
<td>Deficiency strains</td>
</tr>
<tr>
<td>DGRP</td>
<td>Drosophila Genetic Reference Panel</td>
</tr>
<tr>
<td>DPGP</td>
<td>Drosophila Population Genomes Project</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and drug administration</td>
</tr>
<tr>
<td>GAL4</td>
<td>Galactose-induced gene 4 (Yeast)</td>
</tr>
<tr>
<td>GPCR</td>
<td>G-protein coupled receptors</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study</td>
</tr>
<tr>
<td>HPC</td>
<td>High-throughput computing</td>
</tr>
<tr>
<td>IQR</td>
<td>Interquartile range</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>MAF</td>
<td>Minor allele frequency</td>
</tr>
<tr>
<td>Mg</td>
<td>Milligram</td>
</tr>
<tr>
<td>min</td>
<td>Minute</td>
</tr>
<tr>
<td>MI</td>
<td>Millilitre</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>MOA</td>
<td>Mechanism of action</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NEB</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>Ng</td>
<td>Nanogram</td>
</tr>
<tr>
<td>OMI</td>
<td>Ozone Monitoring Instrument</td>
</tr>
<tr>
<td>Pc3p21</td>
<td>Human cancer prostate cell line with p21 tet off system</td>
</tr>
<tr>
<td>POMC</td>
<td>Proopiomelanocortin</td>
</tr>
<tr>
<td>PTEN</td>
<td>Phosphatase and tensin</td>
</tr>
<tr>
<td>RNAi</td>
<td>Ribonucleic Acid Interference</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species or Radiation and Oxidative stress</td>
</tr>
<tr>
<td>sec</td>
<td>Second</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>SIFT</td>
<td>Sorting Intolerant From Tolerant</td>
</tr>
<tr>
<td>SPECTRE</td>
<td>Special Computational Teaching and Research Environment</td>
</tr>
<tr>
<td>TASSEL</td>
<td>Trait Analysis by aSSociation, Evolution and Linkage</td>
</tr>
<tr>
<td>TOMS</td>
<td>Total Ozone Mapping Spectrometer</td>
</tr>
<tr>
<td>UAS</td>
<td>Upstream activating sequence</td>
</tr>
<tr>
<td>UV/UVR</td>
<td>Ultraviolet radiation/irradiation</td>
</tr>
<tr>
<td>UVA/B/C</td>
<td>Ultraviolet radiation/irradiation A/B.C</td>
</tr>
<tr>
<td>UVI</td>
<td>Ultraviolet radiation index</td>
</tr>
<tr>
<td>WMO</td>
<td>World Meteorological Organization</td>
</tr>
<tr>
<td>X</td>
<td>Drug 'X'</td>
</tr>
<tr>
<td>XLA</td>
<td>X-linked agammaglobulinemia</td>
</tr>
<tr>
<td>Y</td>
<td>Drug 'Y'</td>
</tr>
<tr>
<td>μg</td>
<td>Microgram</td>
</tr>
<tr>
<td>μl</td>
<td>Microliter</td>
</tr>
<tr>
<td>μM</td>
<td>Micromolar</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

Natural genetic variation refers to the genotypic differences between individuals within (and among) populations. These differences in the genotype may be a major contributor to the phenotypic differences between individuals (natural phenotypic variation). This can be distinguished from the genetic diversity of a species which the total number of genetic characteristics. Both genetic variation and genetic diversity contribute to the biodiversity within ecosystems. Healthy functioning ecosystems often have relatively high levels of biodiversity (Cardinale et al., 2012).

Solar ultraviolet radiation (UVR\(^1\)) is a major environmental factor that influences ecosystems, penetrating the earth’s atmosphere and reaching the surface. UVR is a narrow band of the electromagnet spectrum between ~215nm to ~400nm. Different amounts of UVR can penetrate the atmosphere according to wavelength, generally, 95% of UVA (315-400nm), 5% UVB (285-315nm) and traces of UVC (285-215nm) reach the surface. The distribution of UVR is dependent on the orbit of the earth and the composition of the atmosphere. Levels of UVR vary greatly and are dependent on many factors including longitude-latitude, aerosols, pollution, cloud cover and the ozone\(^2\) (McKenzie et al., 2011). In most cases UVR is hazardous to biological organisms. In this chapter, I will discuss the biological relevance of UVR throughout history and the role it has played in the evolution of life.

1.1 Solar UVR as an environmental factor

Life on earth began around 3.5 billion years ago during the Archean period, and the earliest photosynthetic cyanobacteria date between 2.45 to 2.1 billion years ago (Tomitani et al., 2006). The great oxygenation event started around 2.3 billion years ago and coincided with the development of more complex life forms. Before this time, there was no or very little oxygen in the atmosphere (which makes up the ozone layer)

---

1 UVR- Ultraviolet Radiation. I may use the terms UVR, UV and solar UVR interchangeably for Ultraviolet Radiation from sunlight

2 Ozone- An allotrope of oxygen found in the stratosphere
(Anbar et al., 2007). In the absence of an ozone layer, the levels of surface-UVR would have been much higher and the shorter wavelengths of UVC would have penetrated the atmosphere. Some studies suggest carbon dioxide would have screened extreme shortwave UVC (<190nm) (Ogawa, 1971) and hydrocarbons were also able to screen some UV (Sagan & Chyba, 1997), though less efficiently than the ozone.

Early Archean atmosphere was anoxic and there was no ozone layer but there may have been high enough levels of elemental sulphur to screen UVR (Fig.1.1). Volcanically produced sulphur dioxide (SO$_2$) and hydrogen sulphide (H$_2$S) converted to sulphur vapour (S$_8$) by solar radiation may have created an effective solar UVR screen. SO$_2$ can absorb wavelengths above 260nm, H$_2$S below 260nm and S$_8$ can absorb most of the UV band (Kasting et al., 1989).

It has been suggested, that in the late Archean period, the levels of sulphur decreased and levels of methane increased (Fig.1.1). The increase in methane coincides with the arrival of cyanobacteria and an increase in atmospheric oxygen. This high level of methane, generated by bacteria, could have offered as much UVR screening as atmospheric sulphur vapour (Pavlov et al., 2001). As well as methane, the UVR absorbance properties of oceanic water with high ferrous iron concentrations would likely have attenuated the penetration of UVR. This affect would have given added protection to the pre-terrestrial organisms of the time (Cockell, 2002). For periods between high levels of methane and the formation of the ozone layer, other organic compounds such as aldehydes may have offered some atmospheric UVR screening (Fig.1.1).
Fig. 1.1 Conceptual graph showing different factors affecting the environment throughout evolution
The relative levels of solar luminosity, UV-screening, atmospheric oxygen alongside major evolutionary events are depicted. The rapid increase in species diversity known as the Precambrian explosion (red line) occurred between the Proterozoic and the Phanerozoic eons (Hessen, 2008).

The luminosity of the sun has increased by approximately 25% since the Archean period (Gough, 1981). However a combination of factors has reduced the levels of UVR. Factors that are biologically relevant and can reduce solar irradiance\(^3\) include the presence of the ozone, other UVR screening atmospheric gases and aquatic UVR screening substances (these will be discussed later). Overall, the evidence suggests UVR has been a major environmental factor throughout evolution from the earliest common ancestor to present day.

\(^3\) Solar Irradiance- power of radiation on the earth’s surface (power unit/unit of area)
1.2 Dynamics of solar UVR and the ozone layer

The definition of climate change is the variation in climate conditions over time; however the term is often used to describe global climate changes caused by human activity. There are both natural and anthropogenic causes for changes to a climate. Some of the causes of natural climate change include: variability in solar radiation, changes in ocean depth, changes in the earth’s orbit and volcanic eruptions. Recent climate change, whether natural or anthropogenic, has had an impact on solar UVR irradiance (McKenzie et al., 2011), and placed new environmental pressures on ecosystems.

The structure of ecosystems is shaped by internal and external factors. The different species that inhabit an ecosystem for example are considered internal factors, as they can influence ecosystem processes\(^ 4\) and in-turn are influenced by them. Conversely, climate is considered an external factor which can influence ecosystems, but cannot be influenced by them. Climates can have major effects on ecosystems (Chapin III et al., 2002) and the geographic variation in climate is the strongest determinant for variation in ecosystems (Chapin III et al., 2002). Species within an ecosystem are often unable to survive in another one due to the change in climate. Although in most cases, a community of organisms is considered only to operate as an internal factor, the collective effect of human activity is able to influence the climate (Chapin III et al., 2002), primarily through ozone depletion.

The ozone layer is the principal screen against solar UVR irradiance and is able to screen nearly all UVC and most UVB (Fig. 1.2). Since the 1980s, some regions of low ozone have led to higher levels of UVB which have caused plant growth to decline (Ballare et al., 2011). These levels have affected aquatic food webs, reducing viability and growth (Llabrés et al., 2013). The decline in amphibian populations due to UVB (Stuart et al., 2004) is also a concern to conservationists (Wake & Vredenburg, 2008), as many view amphibians as valuable biological indicators to environmental change. Conversely, there is evidence that solar UVB promotes plant defence systems (Ballare et al., 2012) and act as a disinfectant of aquatic pathogens. UVC is also known as

\(^4\) Ecosystem processes- The transfer of energy and abiotic material to/from an ecosystem
germicidal UVR and is potentially the most damaging to cells (albeit with less penetration than UVB) (Uehara et al., 2014), but fortunately most of the radiation is screened by the ozone (Fig. 1.2). The targets of UVR damage and its effect on fitness are discussed in chapter 1.4.

![A typical profile of ozone density and solar UV irradiance in the northern hemisphere](https://www.espo.nasa.gov/solveII/implement.html)

**Fig. 1.2** A typical profile of ozone density and solar UV irradiance in the northern hemisphere

A typical profile of ozone density (Dobson Units/kilometre) in the stratosphere (the layer that lies between the stratopause and tropopause) overlaid with typical penetration of UVA, UVB and UVC in the atmosphere over the northern hemisphere (Adapted from Fig. 1, https://www.espo.nasa.gov/solveII/implement.html).

Global distribution of UVB is heavily influenced by the ozone layer, and varies over geographical regions. UVA on the other hand, is not absorbed by the ozone and the global distribution pattern of this radiation is generally dictated by the orbit of the earth and the atmosphere, without much influence from changing ozone. For this reason, solar UVA levels are highest near the equator and decrease towards the poles. The distribution of global UVR (as UVI) is shown in Fig. 1.3. This pattern is mainly due to the angle of incidence of solar radiation. UV light, like all other wavelengths comprising the electromagnetic spectrum, is subject to the projection effect. The projection effect is the relationship between the angle of incidence of light, and the area of the exposure of that light. For example, if the angle of light irradiating an area 1m by 1m is
changed by 60 degrees from perpendicular, the area is increased to 1m by 2m. Therefore, changing the angle reduces the number of photons that hit a given area and so the intensity of the light is reduced. This effect means earth’s poles get less light per area and this has a great effect on UVR intensity. Another reason for reduced light intensity at the poles is that more atmosphere has to be penetrated by the light. This greater amount of atmosphere will affect UVB far more as it is more easily absorbed by the ozone.

Fig.1.3 Global UV index map
Global UV index$^5$ (UVI) on the 28th October 2004. The UVR gradually decreases towards the poles. The map shows high variability in UVI globally and within continents such as Europe (appx. 4-8). There is area of high UVR near the South Pole caused by a hole in the ozone layer. (Adapted from [http://maps.grida.no/go/graphic/the-global-solar-uv-index](http://maps.grida.no/go/graphic/the-global-solar-uv-index)).

$^5$ UV Index- International standard linear scale for UV intensity
Ozone depletion first became a concern in the 1970s when Molina & Rowland proposed that chlorofluorocarbons that entered the atmosphere would lead to ozone destruction (Molina & Rowland, 1974). Further evidence for ozone depletion was published in 1985 by Farman and colleagues who reported an ‘ozone hole’ over Antarctica (Farman et al., 1985). At the time, there was such concern over ozone depleting substances (ODSs) that a UN treaty to stop their use was signed. The treaty called the Montreal protocol was the first treaty ever to be ratified by all member states. This consensus in reducing the use of ODSs has had a major impact on helping the ozone to recover (Slaper et al., 1996). The Montreal protocol also reduced the use of greenhouse gases as many of them are ODSs (Velders et al., 2007).

One of the main drivers for unity on the treaty was the adverse effects of high solar UVR to human health. Solar UV irradiation can cause immunosuppression, eye diseases such as cataracts (Norval et al., 2011) and is a major risk factor for skin cancers (de Gruijl, 1999; Gilchrest et al., 1999). The most common skin cancer in humans is non-melanoma skin cancer (NMSC) which in 2006 affected 1.17% of the US population (Rogers et al., 2010). Although there are many studies linking high exposure of UV and skin cancers, it is likely the increases are due to climate change and modern lifestyles (de Gruijl, 1999).

The long lifespan of some ODSs means the recovery will be slow. By 2040-2060 the levels of Antarctic ozone are predicted to recover to 1980 levels (Previdi & Polvani, 2014). Fig.1.4A shows the projected recovery in Ozone levels, which in its current course is likely to recover by the end of the century. This will lead to particles and aerosols becoming the dominate drivers for global UV exposure levels. Over this time, ecosystems will be subjected shifts in intensity and spatial variation of UV exposure. An assumption is often made that an increase in ozone will lead to decreased solar UV irradiance; however this is disputed and the evidence suggests a mixture of increases and decreases (Hegglin & Shepherd, 2009).
Fig.1.4 Predictions of future UVR
A. Conceptual graph showing the recovery of the ozone over the last 40 decades and the projected recovery (WMO (World Meteorological Organization), 2007). B. Predictions of future global UVR using multi-model average (McKenzie et al., 2011).
The most interesting aspect of UVR distributions, from a biological perspective, is the global variation in solar UV irradiance. This can be seen in Europe where there are large differences in solar UV irradiance between the north and south (Fig.1.3). Global climate change is having a major impact on these distributions and will continue to do so in the future. McKenzie et al., (2011) predict that the levels of erythemal UV are likely to decrease at the poles by 2100. The northern high latitudes are predicted to decrease by ~9%, and the southern high latitude by ~20% (during the early summers) (Fig.1.4B). The mid-latitudes and topics are predicted to increase ~4%. Solar ultraviolet radiation under current predictions will change according to weather patterns and, the interaction between the climate and solar radiation is a dynamic ever changing phenomenon.

Over this time, these variations in solar UV irradiance might place different selection pressures on populations, which may contribute to genetic variation between populations and alter the spatial distribution of alleles at those loci (associated with UVR-sensitivity). In fruit flies, several clines have been reported in variants of genes across that associated with traits such as temperature-resistance (McColl & McKechnie, 1999) and biological rhythms (Costa et al., 1992), but currently no cline has been reported for UVR-associated genes.

**1.3 Role of UVR-sensitivity in Evolution**

UVR has played an important role throughout evolution and is an interesting environmental factor playing a dual role both as a natural selective force, as well as a mutagen (and consequently a source of genetic variation). These roles are discussed below.

DNA molecules have been shown to have high resilience when exposed to UVR. Certain wavelengths of UVR such as UVB are able to penetrate the skin easily and can directly interact (be absorbed) by DNA (Pfeifer et al., 2005). It has been shown that DNA has the ability to transform energy from UVR (UV-photons) into harmless heat.

---

6 Cline- A cline (or ecocline) is a gradient of genetic or physiological change which occurs along a geographical line.
(vibration energy) very quickly- a process called ‘ultrafast’ internal conversion (Pecourt et al., 2001). This characteristic of DNA serves to dissipate nearly all energy from UV-photons into harmless heat energy and is highly efficient, thus acts as photoprotection (Pecourt et al., 2001). When DNA is damaged, a large enough dose has to be delivered as only a small fraction of photons cause damage. Many commercial sunscreens contain ingredients that act in a similar way (Kimbrough, 1997).

The natural photoprotective ability of DNA was probably one of the first evolutionary steps to deal with UVR. The biochemistry of DNA would have been under high selection pressure during early life and the UVR absorbing properties may have acted as a natural sunscreen (Pecourt et al., 2001). During this time the more harmful UVB and UVC bands were able to reach the surface of the earth (Ch 1.2, p.4-5) creating a substantial mutation load. Although this would have caused many deleterious mutations, it is likely the mutagenic properties of UV helped to increase the complexity of DNA sequences and biochemical processes (Rothschild, 1999).

Over billions of years nature has evolved a myriad of physiological, cellular and behavioural adaptions. Many of these adaptions can have high energy costs and this reflects just how harmful UVR is. Conversely, many organisms have adapted to high UV exposure to the extent it has become essential for their fitness. For example, some reptiles use UV for Vitamin D synthesis and are able to cope with high UV exposure by efficiently repairing DNA damage and shredding sunburnt skin (Chang & Zheng, 2003). On the other hand, a reduced level of UV exposure can lead to developmental abnormalities (Oonincx et al., 2010) and proliferation of skin pathogens. In addition, Vitamin D production is evolutionarily conserved (Holick, 2003) and essential for terrestrial vertebrates to develop strong bones and increase in size (Bouillon & Suda, 2014).

Development of complex terrestrial organisms has been helped by the lowering levels of UVR on land, the high mutation load would have stunted evolution. The lower levels of mutation promoted variance and perhaps contributed to the acceleration of evolution in complex terrestrial organisms (Cockell & Raven, 2007). This is in contrast
to organisms in the oceans where UV exposure is significantly lower and less impact on evolution from the Archean period (Cockell & Raven, 2007).

After the Precambrian explosion the levels of oxygen increased rapidly and this coupled with the increase in oxygen combined with high solar UVR levels would have increased DNA and membrane damage. The increase in mutation rates may have favoured species with high non-coding to coding DNA ratios as a high ratio could have buffered some of the mutations caused by UV. More complex organism have high ratios of non-coding to total DNA (Taft et al., 2007) and this theory suggests UV-induced DNA damage could have driven selection of more complex organisms with bigger genomes.

There is some evidence that sexual processes (in bacteria) evolved because of damage to DNA by UVR (Mongold, 1992). The theory suggests sexual organisms have an advantage in recovering from DNA damage because they carry multiple copies of each chromosome; during sexual reproduction, disadvantageous mutations may be recombined, resulting in some mutation-free progeny, or neutralised by complementation. On the other hand, asexual organisms cannot recover irreversible mutations through processes such as recombination, resulting in a mechanism called Muller’s Ratchet (Muller, 1964). Additionally, natural selection has been shown to be less effective at removing disadvantageous mutations compared to sexual reproduction (Kondrashov, 1988). In an evolutionary context, UV would be one of the prime suspects for the evolution of sex through DNA damage repair (Hoelzer & Michod, 1991).

Another similar theory suggests UVR may have promoted evolution of polyploidy. In the absence of sexual preproduction, some species in high UVR regions tend to have chromosomal duplications (Aguilera et al., 2007). Polyploidy allows these organisms to counter Muller’s ratchet, increase genetic variation and allow for differential expression of genes. In wheat and cotton plants, polyploidy changes levels and transcriptions and interactions between genes creating new phenotypes (Appels, 2009; Bottley & Koebner, 2008). Polyploidy also plays a role in protecting species

---

7 Polyploidy- Chromosomal duplications
against extinction e.g. the unisexual fish Amazon molly (*Poecilia formosa*) (Loewe & Lamatsch, 2008).

**1.4 Targets of UVR damage**

There are many mechanisms underlying UVR-sensitivity that can be categorised into three groups: screening or blocking UVR (e.g. producing UV absorbing melanin), biochemical or cellular mechanisms (e.g. DNA repair) and changes in behaviour (e.g. nocturnal animals). The damage caused by UV irradiation has many molecular targets. The most significant and well-studied is DNA damage, which can be sub-divided into direct and indirect DNA damage. These are described below.

*Direct DNA damage*

Direct DNA damage can be caused by UVA, UVB and UVC, but in nature is predominately caused by UVB (Park *et al.*, 2002). This is because the wavelength is ideally suited to penetrate cells and is readily absorbed by DNA. There have been many *in vivo* experiments demonstrating induction of mutations by UVB (Pecourt *et al.*, 2001; Ravanat *et al.*, 2001; Pfeifer *et al.*, 2005). The genotoxicity of UVB has been well documented as causing the classic Cyt-Thy\(^8\) transition mutations. The mutations occur when high energy UVB photons collide with DNA and cause two consecutive cytosines nucleobases to form dimers via strong covalent bonds (Fig.1.5). The presence of two consecutive cytosines is considered a UVB damage hotspot. Other names for these dimers are pyrimidine dimers or cyclobutane pyrimidine dimers (CPD). Enzymes involved in the replication of DNA are usually able to read Thy<>Thy, Thy<>Cyt, and Cyt<>Thy dimers, but Cyt<>Cyt dimers are often read as Thy<>Thy. This error in replication produces mutations that are biased towards Adenine (as A is complementary to T) (Ravanat *et al.*, 2001).

---

\(^8\) Cyt, Thy, Ade, Gua- Abbreviations for Cytosine, Thymine, Adenine and guanine
UVB can also produce the more harmful DNA lesions called 6,4-photoproducts also known as 6,4 pyrimidine-pyrimidones (Fig.1.6). The frequency of these photoproducts is lower than CPDs and requires different mechanisms for their repair (Ch 1.4.2, p.16). Fig.1.6 shows how unstable oxetane is formed via cycloaddition involving the C5-C6 double bond located in the 5’ end and carboxyl C4 at the 3’ end. This intermediate spontaneously forms pyrimidine (6-4) pyrimidone, the 6,4-photoproduct. The pyrimidine is able to undergo conversion to Dewar valence isomer with UVB, the photoproducts are dependent on whether homo-heterodimers of Cyt/Thy are involved. Dinucleotide with 5’ cytosines may also deaminate to thymines. Both the isomer and the 6,4-photoproduction deaminate a hundred times slower than CPDs and so are more damaging (Ravanat et al., 2001).
Fig. 1.6 The production of 6,4-photoproducts from thymine

The formation of 6,4-photoproducts with intermediate step of oxetane production. Further exposure to UVB results in Dewar valence isomer adapted from (Ravanat et al., 2001).

Purine bases are also damaged by UVB radiation, causing dimerisation of adenine or addition to vicinal\(^9\) thymine (Koning et al., 1990). These Ade<>Thy adducts are mutagenic (Zhao & Taylor, 1996) but occur at a much lower frequency than adenine dimers (Sharma & Davies, 1989), with CPDs and 6,4-photoproducts making up the predominate lesions associated with UV-induced DNA damage.

UVR can directly damage many molecules in cells including mitochondrial DNA, amino acids such as tryptophan, flavins, quinones, NADPH and NADH. Absorption of UVR by these chromophores\(^10\) is the main route for indirect DNA damage.

Indirect DNA damage

Indirect DNA damage occurs when UVR triggers the production of harmful DNA damaging agents in the cells. Chromophores absorb high energy photons of UVR but are unable to convert the energy to harmless heat, or respond to it in any other way. DNA and some pigments such as melanin can convert the energy through ‘ultrafast’ internal conversion (Ch 1.3, p.10), however some chromophores are left in an excited high energy state. These molecules are not able to maintain this state, and therefore favour base modification. Unfortunately, the favoured pathway to return to a low energy state generates harmful agents such as reactive oxygen species (ROS) (Ravanat

---

\(^9\) Vicinal- Two functional groups next to each other are bonded by two adjacent carbons

\(^10\) Chromophore- A (or part of) molecule that is a particular colour i.e. absorbs certain wavelengths
15

et al., 2001). Excited chromophores can interact with DNA directly or can react with other species (Bishop, 1994), the most common being triplet oxygen ($^3\text{O}_2$). Chromophores reacting with this ‘normal’ oxygen, produce singlet oxygen ($^1\text{O}_2$). The singlet oxygen is then able to react with DNA to return to $^3\text{O}_2$. The most frequent reaction with singlet oxygen occurs with guanine to produce 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxodGuo) (Ravanat et al., 2001). Other ROS species including free radicals can be very dangerous to DNA as they can move through tissue and attack different organs. ROS can also damage pools of nucleotides in the nucleus prior to DNA replication (Sekiguchi & Tsuzuki, 2002).

There is reportedly no immediate pain or reaction from DNA damage caused by ROS, unlike direct damage which can be measured by sunburn (Erythemal UV$^{11}$). UVA is the major cause of indirect DNA damage which cannot be measured well with the erythemal UV exposure scale. For this reason, early sun-block lotions did not have UVA protection as it was not considered harmful (Armeni et al., 2004).

In 2011, Lease and Papageorgio created an algorithm to identify the signatures of UV-induced DNA damage in the human genome. The authors carried out an in silico study of human gene sequences to identify which genes are vulnerable to UVB-induced DNA damage. They also accounted for likely functional changes such as differential splicing and amino acid changes. Some proto-oncogenes implicated in skin cancers such as tumour suppressor p53 (Tp53) were found to have above average likelihood for UVB-induced mutations (Lease & Papageorgio, 2011).

1.4.1 Resistance to solar UVR by Screening

Ecosystems that are subjected to high UVR often have many species with high levels of pigmentation. In humans, melanocytes produce the pigment melanin through melanogenesis. This pigment is a very effective UV screen absorbing >99.9% of photons (Meredith & Riesz, 2004). The production of melanin is induced by UVB exposure and this leads to darkening of the skin (Park et al., 2009). Though melanin is

---

$^{11}$ Erythemal UV- Measure of potential biological damage caused by sunlight based on the erythemal action spectrum
found in plants (catecholic melanin), many plants produce flavonoids and phenolic components which are also effective UV absorbing compounds (Bieza & Lois, 2001). The variation of pigmentation in aquatic organisms suggests that the production of these photoprotective compounds is associated with a cost; organisms that inhabit deeper water often have less pigmentation compared shallower waters where UV can penetrate. This is reinforced by a selective pressure against pigmentation in order to avoid predators (Hansson, 2000).

1.4.2 Cellular repair mechanisms

DNA repair mechanisms are only employed when photoprotection fails. The three main repair pathways for UV-induced DNA damage are direct reversal, nucleotide excision repair (NER) and base excision repair (BER); these pathways are evolutionary conserved (Guzder et al., 1995; Huang et al., 1992; Kim et al., 2012; Shuck et al., 2008).

DNA that has been damaged by UV, yielding CPDs and 6,4-photoproducts, can be repaired by direct reversal of the adduct. Enzymes such as photolyases can use UV/blue light to drive the conversion of dimers back to two individual bases (photoreactivation). This process is dependent on UV/blue light and photolyase enzyme, which is interesting as the same wavelengths that caused the damage, are able to reverse it (Hidema et al., 2007). Furthermore, it has be found that these adducts have an absorption peak that coincides with UVC, meaning UVC irradiation can reverse/split CPDs via direct absorption (Ravanat et al., 2001). Photolyases are an evolutionary conserved family of proteins whose genes are found throughout diverse range of taxonomic groups.

Cells use the NER system for the excision of CPDs and 6,4-photoproducts before DNA replication. The NER can be sub-divided into global genomic repair (GGR) and transcription coupled repair (TCR), which differ in the ability to recognise DNA distortions. TCR repairs lesions that have stalled transcription and GG-NER repairs constantly throughout the untranscribed genome (Lans et al., 2010). The purpose of NER is to recognise the DNA lesion and make an excision of nucleotides leaving a
portion of single stranded DNA. The opposite strand can then be used as a template to polymerise new nucleotides with no change to the sequence.

Xeroderma pigmentosum (XP) and Cockayne syndrome (CS) are two human diseases from which many of the proteins involved in NER were named after. There are many proteins involved in NER including recognition proteins DDB and XPC (Sugasawa et al., 2005), unwinding proteins (XPB and XPD), stabilising proteins, endonucleases of 5’ (ERCC1) and 3’ (XPG) ends, proteins to protect the opposite strand (RPA), and polymerase and ligation proteins (Rechkunova & Lavrik, 2010). XP sufferers have inefficient NER and are sensitive to light especially UVR which can cause major sunburn in small doses (Halpern et al., 2008).

The BER system is another mechanism that is able to excise bases damaged by ROS. The damage caused by ROS involves greater modification to bases and this distorts the double helical structure. Often guanine is converted to 8-Oxoguanine (8-oxyG) (Dianov et al., 1998), and a DNA glycosylase is able to detect and remove the base (Bigot et al., 2009). DNA glycosylases can hydrolyse the N-glycosidic bond leaving an apurinic/apyrimidinic\(^\text{12}\) (AP) site. The AP site is then recognised by an AP endonuclease and excised, followed by polymerisation and ligation. Other bases that can be detected by BER are alkylated bases, deaminated bases and Uracil in DNA.

Collectively, these entire evolutionarily conserved DNA repair pathways are known as DNA damage response (DDR) pathways (Fig.1.7). Initiation of any of the DD Rs eventually leads to activation of the tumour suppressor protein p53, an important component of genome stability (See Chapter 1.6 for detailed overview). Briefly, once p53 is stabilised through phosphorylation, it moves into the nucleus and activates transcription of target genes (Lakin & Jackson, 1999). There are several cellular outcomes (effector mechanisms) for the cell including cell cycle arrest, senescence and apoptosis.

\(^{12}\) Apurinic/Apyrimidinic site - A site with neither a purine or pyrimidine, just the sugar-phosphate backbone
Fig. 1.7 DNA-damage response (DDR) pathways
When DNA is damaged, specific factors are able to detect the damage, transduce, and increase the response signal. If the DNA damage is not repaired, the signal amplifies and eventually leads to activation of p53. P53 dictates cellular outcomes by promoting expression of multiple sets of genes. Adapted from (Blanpain et al., 2011).

When cells are no longer needed or become a threat to an organism they can undergo programmed death (apoptosis). In a healthy human, millions of the cells in the intestine and bone marrow undergo apoptosis every hour, and apoptosis plays a crucial role in normal development (Meier et al., 2000). A study by Yin et al., (2011) showed that an anti-apoptotic gene UVRAG (UV radiation resistance-associated gene) is important for resistance to UVR (Yin et al., 2011). Apoptosis has also been associated to UVR-sensitivity phenotypes in models such as, human cell-lines (HeLa cells) and D. melanogaster (Kamarajan & Chao, 2000; Sun et al., 2010). However, there is conflicting
evidence when comparing cell-lines to complex organisms. In cell-lines, compromising apoptosis may increase resistance to UV exposure (Sun et al., 2010) but in complex organisms it often leads UV hypersensitivity/cancer (Bruins et al., 2004).

Cellular senescence is another DDR pathway outcome (Fig.1.7). A cell has undergone senescence when it stops proliferating but continues normal metabolic processes. Cells in a healthy human can pass through 40-60 cell divisions before they senesce (known as the Hayflick limit). The telomeres shorten until a critical point and the cell will no longer divide (replicative senescence). In some organisms, enzymatic processes are able to recover telomere length, giving them negligible/reduced senescence and increased longevity (Klapper et al., 1998).

Cell cycle arrest can occur at any of three cell cycle checkpoints: Gap 1 (G₁) phase, Gap 2 (G₂) phase and Mitosis (M) phase. At G₁ the cell can detect DNA damage and stop the cycle (Fig.1.7). Cells may senesce permanently at G₁, cell cycle arrest, or this can be a transient process allowing time for DNA repair (Pellegata et al., 1996).

Senescence in cells can be induced by DNA damage and ROS following UV irradiation (Figue.1.7). This cellular response is mediated through p53 protein which switches on cell cycle arrest by activating p21 and E2F expression (Strozyk & Kulms, 2013). p21, p16, E2F retinoblastoma (Rb) proteins form a complex signaling network which maintains cell cycle arrest (Rufini et al., 2013). A number of studies have shown that an increase in senescent cells is related to tissue and that long term exposure to solar UV irradiation increases the ageing process (photoageing) (Wlaschek et al., 2001; Wang et al., 2009; Herbig et al., 2006; Jeyapalan et al., 2007).

Several studies have shown that repeated exposures to UVR can also increase the induction of antioxidant enzymes (Poswig et al., 1999). A cell is under oxidative stress when an accumulation of ROS cannot be balanced by antioxidants (Sies, 1997). Many cellular mechanisms have evolved to deal with consequences of high ROS (e.g. DNA damage). As a preventative measure, antioxidants are produced to balance the ROS levels and buffer increases. Some antioxidants are enzymatic such as catalase and peroxidases, and others are small compounds such as vitamin C, vitamin E and uric acid (Shindo et al., 1993).
1.4.3 Behavioural mechanisms to avoid solar UVR damage

The UVR avoidance mechanism is a behavioural trait and a strategy employed by many organisms, especially marine species (Wold & Norrbin, 2004). For example, aquatic organisms are able to sense sunlight and migrate to deeper water to avoid harmful UVR (photavoidance) (Hansson et al., 2007). It has been suggested that the first photolyases evolved in marine organisms millions of years ago. These proteins allowed marine organisms to detect and repair UVR damage. These early photolyases evolved into a family of UV/blue light sensing proteins called the Cryptochrome/Photolyase Family (CPF) (Chaves et al., 2011). Another study by Xiang et al., (2010) has found that D. melanogaster larva use retinal structures around the eye and sensors that cover their entire bodies for negative phototaxis. They discovered Class IV dendritic-arborization-neurons-ablated larva show decrease UV avoidance. These photoreceptors on their bodies are a part of the neuronal system and associate with a UVR avoidance phenotype (Xiang et al., 2010).

1.5 UVR-sensitivity as a complex phenotypic trait

UVR has the potential to harm biological organisms in many different ways. Cells employ a plethora of different defence mechanisms to counter and avoid this damage (Ch 1.4, p.12). For this reason, UVR-sensitivity involves many genes and as a trait may not follow Mendelian inheritance patterns, each gene associated with this complex trait may have a small/varying influence on the phenotype. The variation in genes associated with this trait is expected to underlie the phenotypic variation exhibited by the population (alternatively, it is possible for phenotypic variation of a trait to be dominated by variation in a single gene). Typically, complex traits (such as UVR-sensitivity) show continuous range of phenotypes (rather than distinct groups of phenotype or level). Individuals in the population may fall into a complete range of UVR-sensitivity phenotypes from very sensitive to resistant and the trait can be regarded as a continuous (quantitative) trait.

Measuring UVR associated phenotypes such as UV-induced DNA damage and melanisation shows that there is also considerable phenotypic variation among
species. Although environmental factors such as diet can influence the variation in UVR-sensitivity, there is evidence for a strong genetic component underlying this trait in many animal models (Matsumura et al., 2004; Murakami & Johnson, 1996; Gomez et al., 2013). This genetic component is best demonstrated by human diseases such as Xeroderma pigmentosum (XP), Cockayne syndrome (CS) and UV-sensitive syndrome (Davis et al., 1994; Day, 1975; Itoh et al., 1994; Schmickel et al., 1977). UV-sensitive syndrome is characterised by extreme sunburn after even small amounts of UV exposure, and mutations in the ERCC6, ERCC8 or UVSSA genes can cause this disease phenotype. The ERRCC6, ERCC8 and UVSSA genes are important for the TCR mechanism of UV-induced DNA damage (Ch 1.4.2, p.16).

1.5.1 Quantitative Genetics of UVR resistance

One method of identifying natural genetic variation underlying a complex trait is quantitative trait loci (QTL) mapping. This approach identifies loci in the genome that mostly account for the variation in phenotype, and therefore are critical for understanding the evolution of the trait. These loci are likely the targets of natural selection and significant for the evolution of the trait (Falconer & Mackay, 1996).

QTL mapping has been successful in identifying important loci/genes in agriculture and human diseases. In crop studies, QTLs have been found for pest control, yield and UV resistance (Mutschler et al., 1996; Kim et al., 2012; Ueda et al., 2005). Studies into human diseases have identified QTLs for cancers (Breast cancer, BRAC1/2 genes) and neurodegenerative diseases (Alzheimer’s disease, APP gene).

In 2010, a study by Jarosz and Lindquist found UVR-resistance QTLs in yeast. They discovered that yeast strains exhibited substantial variability in their resistance to UVR, and were able to map a major QTL to a recombinase gene, MEC1. Furthermore, they recognized that UVR-sensitivity phenotype associated to MEC1 alleles is dependent on the abundance of stress induced protein chaperone Hsp90. Hsp90 is a potentiator of many stress responses in yeast including several drugs (hydroxyurea and chloronitrobenzene) (Jarosz & Lindquist, 2010).
In 2005, Ueda et al., used backcross inbred lines (BILs) to identify a major QTL that conferred UVB resistance in rice (rice \textit{qUVR-10}). The BILs were generated by multiple backcrossing between two strains with different UVR-sensitivity phenotypes (Sato \textit{et al.}, 2003). Two natural strains of \textit{rice}, \textit{japonica} variety Nipponbare which is UV-resistant, and \textit{indica} variety Kasalath which is UV-sensitive were used to create the BILs. Nipponbare strain was backcrossed with Kasalath for five generations whilst selecting for UVR-sensitive plants. The resulting BILs were genotyped and the markers compared to the Nipponbare strain, to identify regions on the genome that may account for the UV-sensitive trait (QTL- \textit{qUVR-10}). Ueda et al., irradiated rice grains were with UVB (normal light conditions plus UVB), and leaf lesions of adult plants were measured as a phenotype. They confirmed the \textit{qUVR-10} region varies in two natural strains of rice, Nipponbare (UV-resistant) and Kasalath (UV-sensitive), and further analysis of this locus revealed the phenotypic variation mapped to a \textit{CPD photolyase} gene. Moreover, comparing the \textit{CPD photolyase} genes of the two strains revealed a base substitution that results in a non-synonymous mutation. This single amino acid change explained most of the variation in UVR-sensitivity (Fig.1.8). Additionally, overexpression in transgenic Nipponbare plants correlated to decreased UVR-sensitivity (Ueda \textit{et al.}, 2005).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{allelic_variation.pdf}
\caption{Allelic variation underlying UVR resistance in \textit{CPD photolyase} in rice sp. and homologues}
\end{figure}

The rice strains Nipponbare and Kasalath are different at amino acid position 296 (Q/H). Orthologous from different species including \textit{D. melanogaster} (Dm) have the ‘UV-resistant’ amino acid at this position.

Shmookler Reis \textit{et al.}, (2006) used recombinant backcrossed lines (RILs) to access loci associated with stress in \textit{C. elegans}. Briefly, RILs are similar to BILs, except
multiple generations of sibling inbreeding is used to create isogenic strains carrying a mix of parental genotypes. The RILs are then genotyped using parental makers. Shmookler et al., used RILs to assess survival after UVR and heat shock. The phenotypic data together with the genotypic data was used to perform QTL analysis, detecting four heat-shock resistance and four UVR-sensitivity QTLs across all five chromosomes (Shmookler Reis et al., 2006).

A study by Gomez et al., (2013) showed all the major chromosomes (X, Chr2, Chr3) in *D. melanogaster* are implicated in UVC resistance. QTL analysis was carried out using (RILS) and several significant QTLs were detected including a major locus which spanned the pericentromeric region of Chromosome two (Q2), a region previous linked to thermotolerance (Loeschcke et al., 2011; Morgan & Mackay, 2006). One QTL was detected in chromosome X (QT1) and three QTLs in chromosome three (QT3, MT1, MT2). The regions detected were rather large e.g. QT1- ~20cM and QT2- ~10cM. The authors concluded there is substantial genetic variation underlying UVC resistance amongst the RILs. They also noted individuals that are continuously irradiated rather than intermittently, show a more complex response (Gomez et al., 2013). It is intriguing that genetic variation underlying UVC resistance is segregating, given that this wavelength band is rather biologically irrelevant, as only trace amounts are reaching the ecosystem.

Association analysis is an alternative quantitative genetic approach, which is particularly popular in human studies, given that pedigree information (or crosses) is not required. Association analyses are able to independently test markers across genomes in unrelated individuals. An accumulation of genome data has led to powerful genome-wide studies to be performed using high density of markers (SNPs), which can be casual or in LD with causal loci (Anholt & Mackay, 2004). In 2009, Shi et al., compared a cohort of 4000+ humans in 67 populations representing 41 ethnic nationalities from Africa to eastern Asia. The study focused on a common polymorphism in the p53$^{13}$ gene (resulting in hp53 pro72Arg amino acid change), and

---

$^{13}$ The role of p53 in UVR response is described in detail in chapter 1.6
identified a latitudinal cline\textsuperscript{14} in the allele frequencies in populations from Africa to Eastern Asia (Fig.10), suggesting that this polymorphism is driven by natural selection. However, a latter study suggests these data do not necessarily show strong signatures of selection as population history was not accounted for (Sucheston \textit{et al.}, 2011).

\textbf{Fig.1.9 Map showing relationship between \textit{p53} polymorphism and Latitude}

The map shows the distribution of the \textit{p53} polymorphism (frequencies of the Arg72 allele) in different wild populations. The darker areas show the higher frequencies and the overall frequencies show a small but significant latitudinal cline (from Shi \textit{et al.}, 2009).

The isoform carrying the \textit{hp53Arg72} allele encodes a protein that has higher apoptosis activity (Murphy, 2006) and with increased metabolic rates, and hence increases fitness in colder climates. Indeed, the frequency of \textit{p53Arg72} was higher at low temperatures and a cline is observed from low to high temperatures (Fig.1.10, A).

In the same study (Shi \textit{et al.}, 2009), a latitudinal cline in another gene, \textit{Mouse double minute 2 homolog} (\textit{Mdm2}) was also identified. Mdm2 is known to negatively regulate of \textit{p53} by suppressing its transcriptional ability (Kussie \textit{et al.}, 1996). An intronic polymorphism in the \textit{Mdm2} oncogene, \textit{MDM2 SNP309} (rs2279744), effects the levels of transcription (Grochola \textit{et al.}, 2010); therefore modulating \textit{p53} activity. Shi \textit{et al.}, found that allele frequencies of \textit{Mdm2 SNP309} correlated with UV intensity.

\textsuperscript{14} Cline- A cline (or ecocline) is a gradient of genetic or physiological change which occurs along a geographical line.
from high UVR in Africa to low UVR in eastern Asia (Fig. 1.10, B). The activity of hp53 is controlled by hp53 and mdm2 genes which is critical for embryonic development and homeostasis in mice models (Lozano, 2007). Given the interaction between MDM2 and p53, the two independent clines seem to be working together to balance p53 levels. They proposed that individuals with increased p53 activity have a selective advantage in colder areas due to increases in metabolism and embryo implantation, but a surplus of p53 activity leads to embryonic lethality. Conversely, an increase in MDM2 activity is able to modulate p53 activity (Shi et al., 2009). This study is important in highlighting that UVR-resistant alleles that are advantageous to a population, may not undergo fixation, even though they are expected to be subject to purifying selection, possibly due to their role in other processes (e.g. pleiotropic constraint). These are discussed in chapter 6.

**Fig. 1.10** *p53Arg72* and *MDM2* frequencies correlated to temperature and Latitude
A. A negative correlation between increasing temperature and *p53Arg72* frequency. B. A negative correlation between increasing natural UV irradiation and *MDM2 SNP309G* (rs2279744) frequency.
Natural variation in nucleotide excision repair was also reported in *Drosophila* (Lupu *et al.*, 2004). Whilst studying *Drosophila* populations from the ‘Evolution Canyon’ in Israel, they discovered a population which has variation in their nucleotide excision repair machinery. Two distinct fly populations are located on each of the opposite slopes of the canyon, one is considered ‘thermosensitive’ and the other ‘thermotolerant’, due to their microclimate (Fig1.11). Importantly, the two populations can meet and mate, and so can be considered as one larger population which is subdivided. The population on the south facing slope (SFS) are subjected to substantially higher levels of solar UV irradiation and temperature and fittingly the thermotolerant population on that slope have a higher efficiency in repairing a particular type of CPD damage. In experiments with flies from wild populations, DNA damage was induced via a chemical mutagen which has a similar but not identical signature to UVR-induced DNA damage. The authors of the study reason that the repair machinery used by the flies is also used in UVR-induced DNA damage repair. They conclude that variation found in the nucleotide excision repair machinery suggests natural variation in UVR-sensitivity in flies from the canyon; with the ‘African-like’ SFS flies being more resistant then the ‘European-like’ NFS flies (Lupu *et al.*, 2004).

**Fig.1.11 The ‘Evolution Canyon’ in Israel**

A picture of ‘Evolution Canyon’ in Israel depicting the angle of sunlight. The South facing slope-SFS side is more exposed then the North facing slope-NFS, as a result, two microclimates exist on opposite sides of the canyon.
Overall these examples indicate that genes associated with UVR resistance are expected to be under strong purifying selection (and consequently retain little variation) to their involvement in core biological processes e.g. DNA repair. However, there is evidence to the contrary (Gomez et al., 2013; Lupu et al., 2004; Shi et al., 2009; Shmookler Reis et al., 2006; Ueda et al., 2005). The extent of the natural genetic variation in wild populations has not been comprehensively studied.

1.6 UVR-sensitivity and p53 pathways

The p53 gene, also known as tumour suppressor p53 (Tp53), encodes an evolutionary conserved protein that has important roles in regulating many cellular functions including cell cycle arrest, apoptosis and DNA repair (Fig.1.7). Studies have shown p53 also regulates broader processes such as metabolism, reproduction and development (Gottlieb & Vousden, 2010; Levine et al., 2011; Kon et al., 2011). As so many critical functions in a cell are mediated through p53, it has been described as the “guardian of the genome”. Under stressful conditions p53 transcriptional activity is switched on by, its stabilisation, and phosphorylation (and other post-translational modifications) of its transcription domain (Prives & Hall, 1999). As shown in Fig.1.7, p53 is an effector protein controlling cellular outcomes, and is found downstream of sensors, transducers, and mediators. For this reason, p53 has been studied in the context of many pathways and acquired the title as a “master regulator”. p53 shows conservation in many species including Drosophila (Marcel et al., 2011).

A number of studies in different models have established a link between p53 and resistance to UV exposure (Bruins et al., 2004; McKay et al., 2000; Palomera-Sanchez et al., 2010). Many of the pathways that respond to UV-induced damage require activation of p53. The most well-known transcriptional responses to UV irradiation are the activation of nuclear factor nB (NFnB), activating protein 1 (AP1) and p53 transcription factors. NFnB switches on immunosuppression and inflammation pathways (Legrand-Poels et al., 1998) and AP1 seems to act predominately on apoptosis pathways (Shaulian & Karin, 2002). A number of DNA microarray studies have probed the transcriptomes of UV-irradiated skin cells (Dazard et al., 2003; Valéry
et al., 2001; McKay et al., 2004). The majority of the differential expression occurs in genes involved in DDR pathways, and these pathways have transcriptionally highly divergent responses (Gentile et al., 2003).

In healthy humans the level of p53 in the epidermis is very low. When skin is exposed to solar radiation, the levels of active p53 rapidly increase in both the dermis and epidermis. p53 promotes cell cycle arrest and apoptosis, which causes sunburn. Individuals that have reduced p53 response are prone to skin malignancies (Hall et al., 1993). Interestingly, cultured skin cells are similar to cancerous cells as they are immortal, and have low levels of p53 which suppresses DDR (senescence and apoptosis, Fig.1.7) increasing resistance to UV irradiation (Chaturvedi et al., 1999).

Another cellular response involving p53 is oxidative stress (Sablina et al., 2005). High levels of oxidative stress can cause DNA damage and DDR pathways to be activated. Conversely, at low levels, p53 activates the transcription of antioxidant genes such as glutathione peroxidase (Budanov et al., 2004). Furthermore, p53 helps balance ROS levels by regulating ROS-producing metabolic pathways such glycolysis (Bensaad et al., 2006).

A study by Cui et al., (2007) has shown that p53 is required for a tanning (pigmentation) response in mice. p53 plays the role of a UV-sensor and activates the transcription of the POMC gene, switching on the UV-pigment response pathway leading to melanogenesis. Their work clarifies the link between p53 and the production of the photoprotective agent melanin (Cui et al., 2007).

The p53 protein is involved in all three categories of defensive mechanisms against UV outlined in chapter 1.4. The gene is evolutionary conserved across eukaryotes (Yang et al., 2002) but functional mutation exists in natural populations (Shi et al., 2009). Therefore, genetic variation in the p53 gene is highly likely to result in phenotypic variation in UVR-sensitivity.
1.7 p53 and ageing

Ageing is a key factor in most human pathologies such as neurodegenerative disorders, cancers and cardiovascular disorders. In a comprehensive review, Lopez-Otin et al., (2013) attempted to categorise nine candidate hallmarks of ageing in order to identify therapeutic targets for pharmaceuticals. The authors identified the following hallmarks for ageing; genome instability, senescence, mitochondrial dysfunction, nutrient sensing deregulation, telomere shortening, loss of proteostasis, stem cell dysfunction, epigenetic alterations and change in cellular communication (Fig.1.12).

Fig.1.12 shows the proposed three groups for the hallmarks of ageing. Primary hallmarks principally cause harm and are negative. The antagonistic hallmarks are cellular processes that attempt to balance the harm caused by primary hallmarks. These antagonistic hallmarks such as senescence only become detrimental if there is an excess response to increasing harm. Together, these overcompensate leading to overall cellular dysfunction and signs of integrative hallmarks of ageing. Although there are strong interconnections between the processes involved in ageing, the authors conclude some hierarchal framework exists. Each of the hallmarks identified cover numerous biological pathways that could be useful targets of intervention that may increase human health and lifespan (Lopez-Otin et al., 2013).
Fig. 1.12 Hallmarks of ageing

The nine hallmarks of ageing proposed by Lopez-Otin et al., (2013) categorised into three groups. Primary hallmarks which are always negative, antagonistic hallmarks that try to compensate and integrative hallmarks and effect cellular outcomes (Lopez-Otin et al., 2013).

Ageing involves many pathways that converge on cellular outcomes such as senescence. As p53 regulates many cellular outcomes, it is not that surprising this gene is also implicated in the process of ageing. The first evidence came from p53 overexpression mutant mice which were shown to have high resistance to cancer but decreased longevity (Tyner et al., 2002); the authors concluded that p53 is pro-ageing. The general consensus was that excess apoptosis and senescence increased cancer resistance, but this led to major cellular dysfunction and decrease lifespan of the mice (Campisi, 2004). However, following the previous study, mice with extra wild-type p53 alleles and measureable increase in p53 activity were generated. These mice had a similar life span to wild-type mice but also had much greater cancer resistance. Moreover, overexpression of p53 with its upstream regulator enhanced longevity but cancer resistance was lost (Matheu et al., 2007). p53 both promotes and slows ageing subject to cellular context and the processes it regulates.
The mammalian target of rapamycin (mTOR) pathway regulates anabolism and growth rates. Upon activation of the mTOR receptor, a signal is passed through the cell which switches off mRNA translation and protein synthesis (Laplante & Sabatini, 2012). In animal models, a sustained signal in the mTOR pathway gives rise to tissue ageing (Harrison et al., 2009; Kapahi et al., 2004). This signal can be repressed by p53 in a number of ways (Tucci, 2012) including expression of AMP-activated protein kinase (AMPK) (Feng et al., 2007) and PTEN.

Autophagy, from the Greek for self-eating, is a cellular process that recycles/degrades damaged or defective components. The process occurs under normal conditions and is important for development and homeostasis. When under stress, such as starvation, autophagy is able sustain and balance energy levels through recycling of cell components which produces intracellular nutrients (Levine & Kroemer, 2008). One of the major genes associated with autophagy in mice in is Atg5. Mice with overexpression of Atg5 have unregulated autography and increased longevity (Pyo et al., 2013). The p53 protein is an important regulator of autophagy, and can both up-regulate and repress the process (Maiuri et al., 2010; Tasdemir et al., 2008; Liang, 2010). Furthermore, the repression of the mTOR signals by AMPK and PTEN and induce autophagy.

Many of the processes discussed earlier regarding p53 such as apoptosis, cell cycle arrest, senescence, oxidative stress, metabolism, reproduction and development (Bensaad et al., 2006; Pellegata et al., 1996; Strozyk & Kulms, 2013; Sablina et al., 2005; Gottlieb & Vousden, 2010; Levine et al., 2011; Kon et al., 2011), are connected to ageing. Though this is a complex picture with much interconnectedness, the current pharmaceuticals for therapeutic targets in ageing directly or indirectly involve p53-associated pathways. However, there is little research in this area and only a select few drugs have proven to work in animal models (de Cabo et al., 2014), and these are coupled with adverse health effects in humans. Nevertheless in my opinion, it is probable any novel drugs that affect longevity will be related to p53-associated pathways.
1.8 Aims and objectives of this study

This project is aimed at identifying genetic variation underlying UVR sensitivity, using *Drosophila melanogaster*, which is a powerful model system for studying molecular mechanisms, as well as population genetics and evolution (Mackay, 2001). Fruit flies are easy to handle, have a short life cycle and share approximately 50% of their genes with humans, but have much more compact genome, \(~174\) million bases\(^{15}\) compared to \(~3,400\) million\(^{16}\) in human. An extensive number of powerful genetic tools are available, including a number of UVR assays (Aguilar-Fuentes *et al.*, 2008; Besaratinia & Pfeifer, 2012; de Cock *et al.*, 1992; Karbaschi *et al.*, 2012; Sun *et al.*, 2010; Shen *et al.*, 2014; Zirkin *et al.*, 2013). The fly genome has been fully sequenced and is well annotated (BDGP), with several large and comprehensive databases (http://flybase.org/ and http://flyatlas.org/).

Natural populations of *D. melanogaster* are spread throughout the world. *D. melanogaster* makes a good model for studying natural variation as sample populations derived from many different regions and climates are available. Indeed, a large number of studies have used *D. melanogaster* as a model organism to gather evidence for natural variation, giving credence to my study (e.g. (Carbone *et al.*, 2006; Goenaga *et al.*, 2010; Harbison *et al.*, 2009; Sawyer *et al.*, 1997; Gomez *et al.*, 2013; Vaisnav *et al.*, 2014; Mackay *et al.*, 2012; Lupu *et al.*, 2004). In association studies, linkage disequilibrium (LD) plays an important role in the resolution of mapping causal variants. In some studies, loci that are known to be under linkage are considered as single markers, which increasing the power to identify associations (detection) but reduces the resolution (localisation). In fruit flies, analyses of quantitative trait loci have a relatively high resolution, which is a benefit of the relatively high recombination rate in *Drosophila*, causing a rapid decay in LD over short genomic distances, even between adjacent sites (Carbone *et al.*, 2006).

The first objective of this study was to investigate the natural genetic variation underlying UVR-sensitivity in *D. melanogaster*. I aimed to develop a suitable UVR-
A sensitivity assay was conducted to detect phenotypic variation amongst the Drosophila Genetic Reference Panel (DGRP). A suitable assay should meet the following criteria: it should be practically easy to perform, have a good degree of accuracy and reproducibility, and therefore be amenable to a high throughput or massively parallel approach. The phenotypic variation identified in this screen was carried forward to a genome-wide association study (GWAS). This was followed up by testing candidate polymorphisms found in this study and validating them with the use of complementation tests, as well as exploring possible enriched pathways.

Alongside this work, a candidate gene approach was also undertaken to explore natural genetic variation within the Drosophila p53, a gene wholly implicated in UV response in many studies (Ch 1.6, p.27). An allelic series of strains was created using a multiple back-crossing scheme (congenic strains, described in chapter 4), carrying natural Dmp53 variants from different geographical regions. These strains were sequenced to identify possible genetic variation. The goal was to use the phenotypic variation in UVR-response between the strains, together with the genotypic variation identified by genotyping, to perform (candidate locus) association study. The molecular polymorphism was also analysed, to identify the signatures of natural selection.

In addition, the p53 gene plays a major role in ageing (Ch 1.7, p.30-31). A recent study has identified two novel anti-ageing drugs (X and Y\textsuperscript{17}) that are able to extend longevity in human cell line models\textsuperscript{18}, in a p53-independent manner. My aim was to test the anti-ageing effect of these drugs in the fruit fly model. The drugs have not been tested in multi-cellular models yet. A selection of assays including motor function and weight gain of flies was employed to assess possible secondary/side-effects of drug treatment. Drug X is under clinical trials and any side-effects can be compared. Using the fly system allowed me to test whether Dmp53 is required to potentiate this effect.

\textsuperscript{17} As patents are currently pending for these drugs, they will be referred to as drugs ‘X’ and ‘Y’

\textsuperscript{18} M Althubiti and S Macip 2014, Pers. Comm., 1 February
2. General Methods

2.1 Fly stocks and maintenance

All fly stocks, including the laboratory wild-type strain Canton-S and other isogenic lines, were grown on maize meal food (Table 2.1) at either 18 or 25°C in constant light-dark cycle of 12:12 hr. The flies were maintained at 18°C for reduced growth rate (~21 days life cycle) and all flies used in experimental work were grown at 25°C (~10 day life cycle). The strains that were used are listed in each of the relevant chapters.

2.2 Mass production of maize meal food for vials, bottles and petri dishes

Each batch of maize meal food was made up to 7 litres and dispensed into vials, bottles or petri dishes. Maize meal, glucose, brewer’s yeast and agar were added to water and the mixture heated at maximum. Antifungal Nipagen and antibacterial propanoic acid were added.

<table>
<thead>
<tr>
<th>Table 2.1 Ingredients for maize food</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize meal (Quaker, UK)</td>
</tr>
<tr>
<td>Glucose (Fisher Scientific, UK)</td>
</tr>
<tr>
<td>Brewer’s yeast (MP Biomedicals, UK)</td>
</tr>
<tr>
<td>Agar (Biogene, USA)</td>
</tr>
<tr>
<td>Tap water</td>
</tr>
<tr>
<td>Nipagen in ethanol (20%)</td>
</tr>
<tr>
<td>Propanoic acid (&gt;99.5%, Sigma-Aldrich, UK)</td>
</tr>
</tbody>
</table>

2.3 Preparing and dispensing food with drug

The drugs were administered to male flies (Canton-S wild-type strain). A daily dose was administered to flies as a food additive. The drugs required suspension in DMSO which was also used as a control for the additive. As no acceptable daily intake (ADI) of DMSO has been identified for flies, the concentration used was informed by previous drug studies in *Drosophila* (1-5% (v/v) final conc.) (Mason et al., 2013; Spindler et al., 2012; Willoughby et al., 2013). Dmp53 null mutant (p53[5A-1-4], stock no. 6815) flies
were acquired from the Bloomington stock centre, USA, and Canton-S flies were a gift from a colleague; Mirko Pegoraro, Genetics department, University of Leicester.

The food was prepared one day before use and kept at 25°C. The first step in the protocol was to label pre-made cool (4 °C) maize food vials with either DMSO (vehicle) or drug/concentration dissolved in DMSO. A small amount of food mixed with DMSO/drug was then added to the pre-chilled labelled vials. Specific concentrations of drug were mixed with melted maize food as required (final conc. 2% (v/v)).

As the fly response to the drug was unknown, the drug concentration was trialled at 5-100µM. Previous studies have reported drugs used at doses between 0-300µM (Mason et al., 2013; Willoughby et al., 2013).
Chapter 3: Natural genetic polymorphism underlying variability in UVR-sensitivity in *D. melanogaster*

3.1 Introduction

As a complex trait, UVR-sensitivity is expected to be under the influence of both genetic and environmental factors, and interaction of the two. There is certainly a strong genetic component to UVR-sensitivity and many of the genes associated with this trait are evolutionary conserved with orthologous genes present in broad range of organisms (Reviewed in Chapter 1). UVR-sensitivity is a complex trait that has been found to be associated with genetic variation in various organisms such as yeasts (Jarosz & Lindquist, 2010), worms (Murakami & Johnson, 1996) and mice (Ueda *et al.*, 2005). In humans, natural variation in cellular regulators of p53, the oncoprotein *Mdm2* was found to be strongly correlated with UVR level (Shi *et al.*, 2009). This evidence for natural polymorphism in UVR related genes is intriguing, given that many of the genes/loci associated with this trait are important for normal cellular processes and would have been expected to be under purifying selection, and consequently show reduced variation.

A QTL study by Gomez *et al.*, (2013) is currently the only other investigation into natural variation in UVR-sensitivity in fruit flies. The authors detected several large QTLs across all three major chromosomes, some of which overlapped with QTL for thermotolerance (pericentromeric region of Chr2). However, the study tested the response of UVC, which is a rather biologically irrelevant wavelength of UV (Henderson, 1977), and relatively high, unnatural doses were used.

Recently, the genome of 205 *Drosophila* inbred strains derived from a wild population in Raleigh, North Carolina have been completely sequenced, allowing genetic analysis to be made at the SNP level (Mackay *et al.*, 2012). This community resource has been named the *Drosophila* Genetic Reference Panel (DGRP) and was developed by 20 generations of sibling inbreeding. The DGRP has been created as a tool for identifying natural variation of quantitative traits and exploring population genomics. Genome-wide association study (GWAS) can be carried out with high
resolution due to the high density of markers i.e. SNPs (Fig. 3.1). The increased power of the genotype-phenotype association using the DGRP comes from the relatively low linkage disequilibrium (LD) in *Drosophila*, which is a consequence of relatively high recombination rate compared to other model organisms (Carbone *et al.*, 2006). The reduced linkage between markers increases the number of informative markers for statistical analysis as there can be complete linkage equilibrium between adjacent SNPs (Carbone *et al.*, 2006). In addition, the accuracy of the phenotypic scores attained in studies is increased by the ability to phenotype many individuals with the same genotype, and this adds to the statistical power to detect associations.

The panel has been screened and verified to ensure substantial genotypic variation between the strains exists (Fig. 3.1). The Mackay lab has shown that the DGRP strains have phenotypic variation in traits such as sleep, starvation resistance and chill coma recovery (Mackay *et al.*, 2012; Harbison *et al.*, 2013).

**Fig. 3.1 Sample of SNPs in the DGRP strains**
Considerable variation in 110 SNPs occurring in ORFs covering the first 1 Mbp of Chr2R. The strains have been organised by similarity of SNPs. Base colour key: A- green, C- blue, G- yellow, T- red and missing data- white.
A recent study by Vaisnav et al., (2014) made use of the DGRP to assess ionising radiation resistance. They detected 32 SNPs, 31 of which were in novel genes, and identified 62 radioresistant and 92 radiosensitive strains. Their study focussed on the medical relevance of radiotherapy and the differences in response between individuals. The authors concluded that the DGRP is underpowered for this trait, and some traits may require more than the 154 DGRP strains available. In this study, ionising radiation resistance phenotype in flies is used to explore genetic variation underlying the trait, but selection at these loci during evolution was not expected or assessed (Vaisnav et al., 2014).

In this chapter, I describe a genome-wide association study (GWAS) that was used to identify loci underlying UVR-sensitivity in D. melanogaster. For this purpose, an assay for testing UVR-sensitivity in fruit flies was developed and a sample of the DGRP strains was screened. The methodology was carefully considered to ensure a natural dose of simulated solar light was used, to increase the likelihood of detecting genetic associations which are biologically relevant.
3.2 Methods

3.2.1 Fly stocks
105 DGRP strains were used to carry out a GWAS. A complete list of DGRP strains is shown in Table 7.1.2. Deficiency strains (Df) were used to carry out the complementation tests (see 3.2.4 below; Table 3.1; Table 7.1.3). All flies were acquired from the Bloomington stock centre, Indiana University, Indiana, USA, and genotypes are shown in the specific method sections.

3.2.2 Assay development for testing UVR-sensitivity
Developing an efficient high-throughput phenotypic screen is the key for successful GWAS. Inspired by a QTL study for UV-resistance in rice (Ueda et al., in 2005) that was based on irradiating rice grains (e.g. the early stage of the rice life cycle), I developed an assay based on embryo viability. Earlier study in Drosophila (Togashi and Okada, 1983) demonstrated that fly embryos are sensitive to UVR and irradiation can cause sterilisation. Furthermore, they were able to dissect UVR-sensitivity during embryonic development and show peaks at early stages (stages 1-3) (Togashi & Okada, 1983).

The use of embryos simplified the method for irradiation and allowed good control of extraneous variables such as temperature variation. Drosophila embryos can be observed under a dissecting microscope making the phenotype easy to distinguish. Unlike some other stages (such as larval), the embryos are not susceptible to changes in light-dark (Sehgal et al., 1992). A large sample size has been used (e.g. \( N_{\text{embryos}} = 1000/\text{strain} \)) for statistical analysis.

The first step of the protocol was to grow large numbers of flies to collect the embryos. Each DGRP strain was grown in bottles (described in 2.1.1) at 25°C to approximately 200-300 flies. These flies were then expanded to 8 bottles of 200-300 flies each, giving approximately 2000 flies. Once at this stage, the flies were moved to new food (bottles) for one day to allow embryos to be laid. These (12 days); the embryos (F2) from the F1 were used in the assay. In this way, the F1 adults were synchronised by age; in preliminary experiments I found that the age of F1 flies can
influence the results. The F1 flies were moved onto new food with a few grains of dried bakers’ yeast (Allinson flour) one day before the assay, to encourage more embryos to be laid for irradiation.

The F2 embryos were collected at specific times using collection tubes, the tubes were placed on to petri dishes with maize food (described in 2.1.1) and plugged with cotton flugs. The petri dishes were replaced between collections. The collections started at 12:00, and continued every 30 minutes for three hrs. A parallel approach was taken so strains could be repeated over 4 tubes (400-500 flies) with only 12 tubes at any one time (three genotypes a day/12 a week).

The irradiations were delivered using the CPS+ Suntest Solar simulator (Atlas materials, Leicester, UK), which harbours a xenon arc lamp that generates the complete spectrum of solar wavelengths. The system also has a filter that narrows the wavelengths to simulate surface irradiation (300-800 nm), and includes UVR wavelengths of natural light. The samples were placed on ice to minimise temperature variability. Each irradiation was carried out three hours after embryo laying, after the blastoderm stage (stage 6). The petri dishes were placed on ice for a 20 secs cooling/chilling period before each dose was delivered, in order to reduce the warming effect of UVR. The petri dishes were then place at 25 °C to develop. One day after irradiation (at ~18:00), the embryos were counted and scored for viability.

The viability was scored by counting the numbers of hatched and unhatched embryos. The totals from all collections were pooled together. In each experiment, half of the embryos were kept as controls and half were irradiated. For control experiments, the embryos were placed in the chamber covered with foil to screen the light. For the purpose of blind scoring, the petri dishes were numbered, mixed (by a colleague), scored and then decoded. In order to establish a critical dose that could be used to reveal differences in phenotype between the DGRP strains, the LD (lethal dose) 50 was identified (see below).
Finding the LD_{50} of Canton-S in the embryo viability assay

The optimal irradiation dose was determined by exposing Canton-S embryos, a standard laboratory strain, to various solar radiation doses. The intermediate dose (LD_{50}) that generates an observable phenotype (Fig. 3.2) was determined. There are two ways in which the dose of irradiation can be altered: time of dose and intensity of irradiation. At first, an experiment changing intensity of light was carried out with single replicates.

![Dose response for UVR in Canton-S strain.](image)

As the level of irradiation increases, the viability decreases. Each point is represented by ~300-500 embryos and the irradiation was delivered for five minutes. The curve shows initial decline in viability to 30% at 25mW/cm^2 and then reduces less sharply until 30mW/cm^2.

In Fig. 3.2, a dose response curve shows that an increase in irradiation reduces viability. In this experiment, the time of the irradiation was kept the same and the intensity of light was changed. Each point represents over 300 embryos and LD_{50} lies between 0 and 25mW/cm^2 (for five minutes). There is a 50% decrease in viability between the control and the first intensity of 25mW/cm^2. The range used in this
experiment was too high to get good resolution of the LD$_{50}$. In the next experiment, a smaller range was used to narrow down the dose required for the LD$_{50}$.

![Graph](image)

**Fig.3.3 Effect of stimulus duration (intensity 25mW/cm$^2$)**
The graph shows as the duration of irradiation increases, the viability decreases. At least 300 embryos represent each point. The irradiation dose is increasing via length of irradiation whilst the intensity is kept at 25mW/cm$^2$. The curve shows gradual decline in viability to approximately 75% to 25% with the LD$_{50}$ at 72 seconds or 18KJ/m$^2$.

In Fig.3.3, the intensity was fixed at 25mW/cm$^2$ and the time of irradiation was adjusted. The LD$_{50}$ for this experiment was calculated using Probit analysis, which returned 72 seconds at 25mW/cm$^2$, or 18KJ/m$^2$. However, since the maximum viability attained in these experiments was approximately 70%, the LD$_{65}$ dose, which is 50% viability relative to the control and should be used for the GWAS. This dose can discriminate against 70% of embryos that would have been viable in this assay under
control conditions. This lethal dose is strong enough to reveal the full extent of the sensitivity to the solar irradiation the strains.

In Fig.3.4 the repeatability of the UVR response was tested in two strains, *DGRP 25174* and *Canton S*. The flies were assayed on different days with independent flies, using over 300 embryos per replicate per strain. There was remarkably good reproducibility for both strains, the differences are non-significant using t-tests ($N_{\text{replicate}}=4$). The differences between the control and irradiated were significant for *DGRP 25174* in two independent experiments; $t(6)=9.96$, $p<0.0001$ and $t(6)=10.96$, $p<0.0001$. The differences between the two conditions were also significant for *Canton S* in two independent experiments; $t(6)=7.13$, $p=0.0004$ and $t(6)=9.91$, $p<0.0001$. An ANOVA across all the conditions confirms these results; $F(7,24)=63.33$, $p<0.0001$. The strains’ response to UV was significantly different between under both two conditions (control conditions; $F(3,12)=7.87$, $p=0.0036$, experimental conditions; $F(3,12)=28.77$, $p<0.0001$). These experiments confirm that a conservative threshold of over 300 embryos per sample/replicate can be used for this method to be reliable. The LD$_{65}$ of the DGRP strain (25174) is 60KJ/m$^2$. 
Fig. 3.4 Reproducibility of two strains in the embryo viability assay at a dose of 18KJ/m²
Each experiment was carried out with independent sample of flies. The grey bars represent the control conditions (no UVR) and the white bars experimental conditions of the two experiments. The experiments showed high reproducibility for each strain in the embryo viability assay over the two independent experiments. The error bars are SEM. At this dose, the DGRP seem to be slightly less resistance and the controls of the strains ranged between 67% and 73%, and conditions are significantly different (F(7,24)= 63.33, p<0.0001) (Full t-tests in Table.7.1.1).

3.2.3 Phenotypic screen of the DGRP for the GWAS
DGRP strains were screened using the embryo viability assay with a dose of 60KJ/m². This dose can be considered biologically relevant dose as the average solar irradiance in Raleigh, North Carolina can be as high as 750KJ/m²/hour in summer (http://www.cpc.ncep.noaa.gov/products/stratosphere/uv_index/uv_annual.shtml). In total, 105 DGRP strains were assayed for both UVR and control responses. For
normalisation of scores, the median score of the control condition were subtracted from the experimental scores.

The association analysis was performed on the University of Leicester’s high performance computing (HPC) service, using a Perl programming script that also called a second R script (the R statistics package, http://www.r-project.org/). To carry out the GWAS, a simple approach was taken where by each SNP was tested for association independently (single-marker analysis). The SNP data were filtered for biallelic SNPs, with a minimum count of 5 per allele. A Wilcoxon rank sum non-parametric test was performed on SNPs within genes where a minor allele frequency (MAF) of greater than 25% is observed. The test compared the two alleles of 778,427 SNPs where high W value correlates with the major allele conferring high median score, and low W value correlates with minor allele conferring high median score. In addition, Wilcoxon rank sum tests were carried out for each SNP after permutations (N=1000) to create a null distribution of the W values. The critical W values were identified using 0.0001 and 99.9999 quantiles of the distribution respectively (1774 and 120), and then applied to all tests carried out on the original data, as the 0.0001 false-discovery rate (FDR). All tests with W values above 1774 and below 120 were considered significant, and the permutations addressed the issue of multiple testing (Churchill & Doerge, 1994). The W values identified in the GWAS were represented in a histogram (at 1% and 99% quantiles) using the ‘R’ software.

The DGRP population was found to have *Wolbachia pipientis* infection in some of the strains (where *Wolbachia* DNA is present) (Mackay et al., 2012). Wolbachia infection has been shown to increase viral resistance and increase fitness in *D. melanogaster* (Teixeira et al., 2008). An association study for infection status (yes or no) was carried out using the embryo viability scores attained for the GWAS. Statistical analysis was carried out with a t-test, and a boxplot was made using R.
3.2.4 Validation of candidate SNPs identified in the GWAS

Significant SNPs were further validated by complementation tests (Box.1). The test involves comparing fly strains with a single a copy of each of the alleles. In order to create these strains to phenotype, deficiency strains that carry chromosomal deletions spanning those loci were used. The deficiency (Df) strains came from two collections, the Exelixis deletion collection (Parks et al., 2004) and Bloomington Stock Centre (BSC) collection (Cook et al., 2012, Roote and Russell, 2012). The collections have the same genetic background and the end-points of the chromosomal deletions have been molecularly mapped. Each chromosome carrying a deletion is paired with a balancer chromosome (a modified chromosomal that carry multiple inversions), and visible dominant markers to maintain viable stocks. The balancer chromosome acted as the wild-type chromosome in the experiments.

For the complementation test crosses, virgin females were collected from the Df strains and crossed with males from the DGRP strains (Box.1, Table.3.1). The same method used in the embryo viability assay was employed to irradiate the samples including the same number of embryos. In this protocol, the embryo viability cannot be calculated as the genotype of the individual can only be established later in development, at the pupa or adult stage. For this reason, the adult viability was scored instead. The adult viability was calculated as the percentage of adults developing from number of embryos.

The adult viability was scored for both genotypes (Df/+ and Bal+/) of each cross (Fig.7.1.1) under control and experimental conditions. Four DGRP strains were crossed for each allele of a SNP (crosses shown Table.7.1.3) and these data were then statistically analysed to test for quantitative failure to complement.
**Box.1 The complementation test**

The lack of complementation test was carried out as described by Mackay in 2001.

The DGRP strains were crossed with a deficiency strains (Df) that carry deletions covering the locus of the interesting SNP (from GWAS analysis). The DGRP strains can carry one of the two possible alleles (A1 or A2). The four resulting progeny of the cross are phenotyped for UVR-sensitivity. If the differences between the progeny carrying the Df chromosomes is greater than with the balancer (Bal) chromosome, there is evidence for lack of complementation. Using multiple DGRP strains allows for statistical analyses to determine quantitative complementation (for crosses in full see Table.7.1.3).
The statistical test for quantitative failure to complement outlined in Box.1 was analysed by a two-way analysis of variance (ANOVA):

$$Y = A + G + A \times G + E$$

Where A is the effect of the SNP (A1 or A2), G is the effect of genotype (Df or Bal), A × G is the interaction of the two, and E is the error variance. Where the difference between, A1/Df and A2/Df, is greater than the difference between, A1/Bal and A2/Bal, a significant interaction term is observed (ANOVA, P<0.05) (Box.1). This is interpreted as a quantitative failure to complement (Mackay, 2001).

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP loci</th>
<th>Alleles of SNPs</th>
<th>Deficiency Strain/ Stock no. (Bloomington stock centre, <a href="http://flystocks.bio.indiana.edu/">http://flystocks.bio.indiana.edu/</a>)</th>
<th>Deficiency strain phenotypic marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>bin</td>
<td>3L:6984893</td>
<td>A/G</td>
<td>w[1118];Df(3L)Exel6085,P{w[+mC]=XP-U}Exel6085/TM6B,Tb[1]/7564</td>
<td>Short fat pupa (Dominant marker)</td>
</tr>
<tr>
<td>Cyp4d8</td>
<td>3L:7487880</td>
<td>C/T</td>
<td>w[1118];Df(3L)BSC437/TM6C,Sb[1]cu[1]/24941</td>
<td>Short bristles on back of adults (Dominant marker)</td>
</tr>
<tr>
<td>vn</td>
<td>3L:5826900</td>
<td>C/A</td>
<td>w[1118];Df(3L)Exel8104/TM6B,Tb[1]/7929</td>
<td>Short fat pupa (Dominant marker)</td>
</tr>
<tr>
<td>MED14</td>
<td>3L:596317</td>
<td>C/T</td>
<td>w[1118];Df(3L)BSC224/TM6C,Sb[1]cu[1]/9701</td>
<td>Short bristles on back of adults (Dominant marker)</td>
</tr>
<tr>
<td>CG42342</td>
<td>3R:12383617</td>
<td>C/G</td>
<td>w[1118];Df(3R)Exel6270,P{w[+mC]=XP-U}Exel6270/TM6B,Tb[1]/7737</td>
<td>Short fat pupa (Dominant marker)</td>
</tr>
</tbody>
</table>
3.2.5 Analysis of enriched pathways network and gene ontology

The R spider software (Antonov et al., 2010) was used to find networks among the genes with significant associations. R spider is an online tool that computes data from Reactome and KEGG network databases to create statistically significant signalling and metabolic networks of input data/genes. Significance P-values for enriched pathways were calculated by R spider using Monte-Carlo simulations (N=1000) of random selection of genes (same number as input size). The data output was visualised and exported using the Cytoscape network analysis software (http://www.cytoscape.org/). The networks are annotated with rectangles representing the input genes and triangles representing intermediates.

The gene ontology (GO) Protein ANalysis THrough Evolutionary Relationships (PANTHER) webtool (Mi & Thomas, 2009) was used to identified pathways associated with GO of the top 200 genes with significant associations.
3.3 Results

3.3.1 GWAS of UVR-sensitivity

There was substantial variation in UVR sensitivity among the DGRP strains (Fig. 7.1.3) with median survivorship ranging from 29% to 47%. Because embryo viability differed between the strains under no irradiation ($F(104,314)=1.634$, $p=0.0007$), the response under experimental conditions was normalised using the controls. Importantly, the variability in control flies was substantially smaller than irradiated flies (Fig. 7.1.3), and the viability between the two conditions did not correlate ($t(103)=-0.0706$, $p=0.94$), indicating that the normalised scores truly represent the differential UV response of the DGRP strains. The variation in median survivorship (normalised scores) between the strains was highly significant ($F(104,315)=15.59$, $p<0.0001$). Furthermore, using the control conditions as a covariate, an ANCOVA was performed and revealed highly significant variation amongst the DGRP strains ($F(104,315)=15.72$, $p<0.0001$).
Fig.3.5 Variation in survivorship of embryos among DGRP strains (normalised scores)
A graph of normalised embryo viability scores for 105 DGRP strains. The strains have been arranged from lowest to highest median score (black dots). The red lines represent the range from minimum to maximum values of four replicates. There is substantial variation in median survivorship between the strains (F(104,315=15.59, p<0.0001), ranging from 29% to 47%.
At the 0.0001 and 99.9999 threshold level of the association analysis, 114 SNPs in 100 genes are significantly associated with UVR-sensitivity; a complete list is shown in Table.7.1.4. The W values range from 14 to 119 for low tail (61 SNPs), and 1776 to 1893 for high tail (53 SNPs). Some of the SNPs lie in notable UVR-associated genes such as Xpd, Ercc1 and Xpac (DNA repair genes). Many novel UVR-sensitivity SNPs have been identified including CG34114, CG14445 and CG43078, which are currently unannotated for gene ontology. The distribution of the normalized effect size for each SNP (the W values) is shown in Fig.3.6A. A scatterplot of W values shows the 114 significant associations over all three major chromosomes (X, Chr2 and Chr3) (Fig. 3.7). The number of sites tested decreases towards the centromeres reflecting the lower number of genes in these regions.

Examples of the most significant SNPs are represented in Fig.3.7, showing approximately 4-6% in median scores. Note however, that the maximal difference observed in the DGRP strains was only 19% (Fig. 3.8). Some of the SNPs examples exhibit strong differences between the alleles with non-overlapping interquartile ranges (in genes TwdlB, shep and alpha-cat). The SNPs are in genes novel for UV response, but have previously been implicated in a variety of other pathways such as development (TwdlB, raw, Bx and alpha-cat) and gravitaxis (shep).
Fig. 3.6 Distribution of W values in the GWAS

W values for each association along three *D. melanogaster* chromosomes (Chr2, 3, X). **A.** Histogram for W values identified using Wilcoxon tests in the GWAS. Distribution of W values represented with cut-offs (grey) at 1% and 99% quantiles. **B.** Distribution of W values represented as a scatterplot along the chromosomes. The 0.0001 and 99.9999 quantiles are shown (red line) with significant associations above and below these thresholds.
Fig. 3.7 Boxplots of significant SNPs associated with UVR-sensitivity
The effect size of six significant SNPs detected by the GWAS is shown. The boxplots represent the range in survivorship between the two variants of each SNP in genes TwdIB, raw, shep, alpha-cat, Bx and CG7556 (A to F). Lines inside each box represent the median, and the bottom and upper ends of the box represent the 25 and 75 percentiles respectively. The error bars represent the 1.5*IQR.
3.3.2 Validation of significant SNPs using complementation tests

Five SNPs that showed significant association were selected for validation by complementation tests. In Fig.3.8, a significant interaction between allele and genotype (AxG) for a SNP in gene CG42342 (F(1,12)=5.094, p=0.043) indicated quantitative failure to complement. The interaction plot has the classic ‘wedge’ shape for interaction and clearly illustrates the greater difference in scores (~15%) between two homozygous alleles (over the Df chromosome), compared to difference between the two alleles over the balancer chromosome (~1%). Fig.7.1.2 shows evidence for 4 significant SNPs and the deficiencies that did not fail to complement.

![Interaction plot representing a complementation test.](image)

**Fig.3.8 Complementation test of significant SNP 3R:12383617 in CG42342**

Interaction plot representing a complementation test. The significant interaction (p=0.04) between the strain (allele, A) and genotype (G); A x G. A significant quantitative failure to complement for SNP in gene CG42342, confirming the variants of the SNP detected in the GWAS affect UVR response.
The maternally inherited *Wolbachia pipiensis* infection is prevalent in the DGRP strains (Table.7.1.5) (Mackay *et al.*, 2012). *Wolbachia* infection has been shown to increase viral resistance and increase fitness in *D. melanogaster* (Teixeira *et al.*, 2008). In most cases the infection affects the reproduction of its host to increase fecundity and reduce male fitness (Stouthamer *et al.*, 1999), but the infection can also cause parthenogenesis and affect cellular processes such as apoptosis (Pannebakker *et al.*, 2007). Comparing the UVR sensitivity between the strains with *Wolbachia* (n=51 and strains without (n=49) revealed no significant difference (t=0.15, p=0.88, df=98) (Fig.7.1.4). Thus, the *Wolbachia* infection did not affect UVR-sensitivity.

### 3.3.3 Enriched pathways network analysis and gene ontology

The pathway network analysis was performed to test for enrichment of specific pathways among the genes that showed significant association, some of which may UVR-related. The network analysis inferred by R spider algorithm returned significant subnetworks from multiple gene input sizes.

The gene network inferred from the Reactome and KEGG databases using R spider algorithm is shown in Fig.3.9, a large number of input genes were computed (N=400) to detect a significant network. A number of genes in the network are implicated in axon guidance pathway (*lea, CDC42, DOCK, RHOGAP93B, SLI, fra, NetB and PAK*). Several genes are associated with regulation of cell shape (*CTA, ARM, ROK and ALPHA-SPEC*), although these genes are all intermediates and were not a part of the input list. Signal transduction (*Pka-C1 and SNF4Ay*) and protein phosphorylation genes are also enriched in this dataset. Two genes are implicated in G-protein signalling (*mAcR-60C and Gbeta13F*). This is consistent with previous studies showing evidence for G-protein coupled receptors involvement in UVR-responses (Seo *et al.*, 2004; Warpeha *et al.*, 2008). The analysis was repeated with top 200 genes that show significant associations (Fig.3.10). Two of the genes in this network, *decay* and *Dmp53*, are associated with apoptosis, a pathway wholly implicated in UVR response. In this analysis, two genes are again implicated in G-protein signalling. The gene *Gbeta13F* was detected in both analyses. When the analysis was repeated with the top 100
genes, the most prominent enriched pathway was nucleotide excision repair (NER) (Fig.7.1.5). NER is an important pathway for UVR-sensitivity with many evolutionarily conserved genes (Ch 1.4.2, p.16). DNA replication is a closely related pathway that was also enriched in this network. Proteins such as RPA (RPA-70) and RPA2 are part of the replication protein family, which has been shown to be important for UVR-induced DNA damage response in humans (Rodrigo et al., 2000; Zou et al., 2006), and some replication proteins are also involved in NER (de Laat et al., 1999).

Fig.3.9 Network of genes (N=400) for UVR-sensitivity
A gene network inferred from Reactome and KEGG databases using R spider algorithm. The rectangles represent input genes (N_{input}=400) and the triangles represent intermediate genes.

Fig.3.10 Network of genes (N=200) for UVR-sensitivity
A gene network inferred from Reactome and KEGG databases using R spider algorithm. The rectangles represent input genes (N_{input}=200) and the triangles represent intermediate genes.
Gene ontology analysis was carried out using gene ontology PANTHER webtool. The top 200 genes from the GWAS were imputed (155 genes were found in the database), the Wnt pathway (4.6%, Table 7.1.6) and apoptotic process (2.6%, Table 7.1.7) were detected. The Wnt pathway is associated with a number of cell fate decisions and development (Swarup & Verheyen, 2012) in flies including embryonic ectoderm (Bejsovec & Wieschaus, 1993; Dougan & DiNardo, 1992). The pathway has been shown to be particularly important for epidermal patterns in embryos (Bejsovec, 2013). It is plausible that the Wnt pathway has an effect on embryo viability - a measure of UVR-sensitivity.

The apoptotic process was identified by the PANTHER webtool and wholly implicated in UVR-sensitivity (Chapter 1.4). Apoptosis is also important process in embryonic development controlled by embryonic patterns (Mergliano & Minden, 2003), suggesting the interconnectedness of the two pathways detected.
3.4 Discussion

In this study I have investigated natural variation underlying UVR-sensitivity in strains derived from a natural population of *D. melanogaster*. I was able to identify 114 novel candidate SNPs at a FDR of 0.0001 and global network analysis revealed the data are significantly enriched for UVR-response pathways (Fig.7.1.5). The assays that I used here focused on the embryonic stage, which has been previously identified as UVR-sensitive (Togashi & Okada, 1983). In the lab, I verified the oviposition (laying) occurs on the food with embryos remaining half-buried until hatching, and therefore it is reasonable to assume that wild embryos are subjected to sunlight. Moreover, embryos require contact with the atmosphere for respiration until hatching. Embryos respire though filaments or ‘horns’ which protrude from the anterior end (Nezis *et al*., 2005).

The reproducibility of the embryo viability assay was critical to carrying out a successful GWAS. The assay was validated by repeating experiments on independent flies of the same strain and successfully reproducing results (Fig.3.4 and Table.7.1.1). However, this level of reliability in the developed assay requires multiple generations of flies and a large sample size; in the GWAS approximately 182,000 embryos were scored. The small difference between the replicates Fig.7.1.3 illustrates the robustness and reproducibility of the assay. Another important factor was the type of irradiation. A study by Gomez *et al*., (2013) accessed the UVC-sensitivity of *D. melanogaster* and some major QTLs were detected. Here, I was able to use biologically relevant doses and wavelengths of UVR. The CPS+ Solar simulator was able to generate a complete spectrum of light including UV and enabled a more natural response to be induced.

The viability scores under the control condition were not 100% and significant variation was observed between the strains (Fig.7.1.3). However, a fairly small range is observed for each strain (*N*_replicates=4, maximum range ~8%), indicating the method was successful in producing robust, reliable and repeatable data. The reduced viability observed could be a consequence of the methodology or unfertilised embryos. As viability was scored by embryo hatching, it is impossible to distinguish true loss of viability and non-fertilisation, indeed fertilisation may occur at ~70% in the DGRP strains (Fig.7.1.3).
Of the significant associations detected in the study, approximately half the SNPs associated with the major allele (53 SNPs), and half with the minor allele (61 SNPs), and most within non-coding regions. It is possible some of these SNPs are in multiple overlapping genes. Of the 100 genes associated with UVR-sensitivity, a number have previously been implicated in UV response including Xpd, Ercc1, mus205, RPA2, Xpac, hay, decay, mrn, and MAT1, which are involved pathways such as NER and apoptosis. In Fig.7.1.5, the effect of SNPs lying within six UVR-response genes is shown (TwdIB, raw, shep, alpha-cat, Bx and CG7556). The TwdIB gene is similar to the TwdID gene which involved in body and cuticle formation (Guan et al., 2006). It is possible that this gene affects the ability of embryos by blocking UVR. The raw gene, together with rib and puc plays a key role in dorsal closure, which is an important step during embryonic development occurring around 8-12 hr after embryo laying (Bates et al., 2008). Whether UVR affects this process is unknown, but it is entirely possible this gene was detected in the GWAS as it may affect embryo viability independent of UVR. The shep gene (named after the astronaut Alan Shepard), is associated with gravity response. Interesting, a link has been identified between gravitaxis and UVR in some marine organisms (Richter et al., 2007), and vertical migration in aquatic organisms is a form of UVR-avoidance (Ch 1.4.3, p.20). Furthermore, in fruit flies, newly hatched first-star larvae exhibit positive gravitaxis and negative phototaxis. The genes raw, alpha-cat and Bx are all involved in development (Bates et al., 2008; Berns et al., 2014; Zenvirt et al., 2008).

A range of five of the significant SNPs with given criteria were selected for validation using quantitative complementation tests (Box.1). Only one of the five SNPs that were analysed in complementation tests (SNP 3R:12383617 in CG42342 Fig.3.8) showed a quantitative failure to complement. The complementation tests thus may indicate that a substantial number of associations are false positives. However, it is also possible that the different genetic backgrounds between strains tested in this experiment may have masked the phenotypic influence of these loci. Indeed, several weaknesses have been indicated associated with complementation testing (Service, 2004). Ideally, validation would be carried by using genetically near-identical strains carrying different alleles at these loci (congenic strains), and then to test these strains
for differences in phenotype-functional tests. Furthermore, as adult viability phenotype was measured for the complementation tests, the 3R:12383617 locus in gene CG42342 can be further associated with complete adult development after embryonic irradiation. The CG42342 gene is 68,594bp long and has 10 isoforms ranging from 2966 to 5682bp. Although, the gene is currently unannotated, protein sequence similarities reveal several collagen-like domains in the protein (http://www.uniprot.org/uniprot/B7Z0K8).

Several pathways not commonly related to UV response were detected by network analysis. The most prominent pathway is protein phosphorylation which was detected in all analyses. As a post-translational modification, this is a common characteristic of proteins (kinases) which are often involved in signalling cascades (Keshet & Seger, 2010). It is plausible post-translational modification plays a large role in the response as opposed to transcriptional changes. These analyses suggest a complex biological response to UV involving overlapping of many pathways (Seo & Lee, 2004).

The network analysis also indicated some pathways commonly associated with UV responses were detected (Fig.3.9, Fig.3.10 and Fig.7.1.5) including those related to apoptosis and NER. These two pathways have been implicated in UV response in various organisms (Chapter 1), including D. melanogaster (Sun et al., 2010; Sandoval & Zurita, 2001). These results further validate the GWAS, and suggest that small contributions of multiple loci (SNPs) to the phenotype are concentrated in particular pathways, giving rise to the phenotypic variation observed (Fig.3.5).

Undeniably, this is a challenge in studies of quantitative traits such as UVR-sensitivity. The DGRP were created to meet some of these challenges behind the main components of genotype-phenotype mapping; localisation and detection (Mackay et al., 2009). The higher marker-density, as well as, the variation in the panel allowed for high resolution mapping in this study. The findings in this study are in parallel with previous studies which have used the panel to analyse complex traits e.g. sleep (Harbison et al., 2013) and starvation resistance (Mackay et al., 2012). As in this study, Harbison et al., (2013) found that the GWAS in sleep detect many genes that had been
previously identified as well as novel associations. Two radiation-based studies, QTL analysis of UVC-resistance (Gomez et al., 2013) and GWAS into effects of ionising radiation, both identified many loci contributing to the phenotypes tested. The multiple significant associations with small effects on the phenotype (e.g. in Fig. 3.8) in the GWAS indicate UVR-sensitivity has a highly complex genetic architecture in D. melanogaster. A recent study by Vaisnav et al., (2014), used the DGRP to assess ionising radiation resistance. They detected 32 SNPs at low threshold of p<10^{-5} and predicted that more DGRP strains are required (265-793) then are available (154). The authors acknowledged there is adequate power for other complex traits (e.g. oxidative stress (Weber et al., 2012) and tunicamycin-induced ER stress (Chow et al., 2013). Indeed, in this study, significant associations for UVR-resistance were detected at a FDR of 0.0001 many of which are implicated in UVR-associated pathways.

Previous studies, namely in human diseases, have identified genes associated with UVR-sensitivity (Brash et al., 1991; Cleaver & Thomas, 1993; Tanaka et al., 1990; Henning et al., 1995). However, in natural populations these genes may not exhibit any genetic variation. Some genes that are imperative for fitness may undergo robust purifying selection. This is a reasonable assumption for genes implicated in UVR-resistance as many are also essential for normal cellular processes e.g. DNA replication (Ch 1.4.2, p.16).

How can a polymorphism in UVR genes be explained? Shi et al., in 2009 illustrated a plausible explanation for genotypic variation underlying traits presumed to be under directional selection. Their study identified variants of two polymorphic sites in two different genes, under selection from different environmental factors. A polymorphism in the hp53 gene associated with winter temperature, and a polymorphism in MDM2 gene associated with low UVR-intensity (Ch 1.5.1, p.23-25). The variants under selection in both polymorphisms influence hp53 activity in opposite directions, which allows a constant the level of activity to be maintained in different geographical locations. Therefore, constant hp53 activity is maintained through environmental pressures acting on the variants of the two genes in the same pathway. The alleles of the two genes appear to be under balancing selection through this ‘trade-off’ phenomenon. The phenomenon offers an explanation for some of the
phenotypic variation observed in the DGRP, where one would expect purifying selection (e.g. in NER), considering the relatively high levels of UVR in Raleigh.
Chapter 4: Natural genetic polymorphism in *Drosophila p53* underlying variation in UVR-sensitivity

4.1 Introduction

The human *p53* (*hp53*, *TP53*) gene is involved in UV responses, as well as many processes including cellular outcomes, reproduction, development, metabolism, melanisation, ageing and tumour suppression (see Introduction). The *hp53* gene has a dual gene structure with an extra internal promoter within intron 4. Similarly, *Drosophila p53* has an evolutionarily conserved internal promoter upstream of exon 3 (Bourdon et al., 2005). The internal promoter of *Dmp53* produces the ΔN*p53* (N-truncated) isoform of the protein (Marcel et al., 2011). This isoform was the first to be identified and early publications designated this isoform as *Dmp53* (Jin et al., 2000; Ollmann et al., 2000). The nomenclature was later changed to *Dmp53* (*Dp53*) for the longer isoform. *Drosophila p53* (*Dmp53*) is a clear structural homolog of *hp53* (Ollmann et al., 2000).

![Drosophila p53 isoforms](image)

**Fig.4.1 Isoforms of Dmp53**

Three isoforms of *Dmp53* labelled with transactivation domain (TAD), DNA binding domain (DBD) and oligomization domain (OD). Adapted from Marcel et al., 2011.

Dmp53 is involved in similar UV responses as *hp53*, such as apoptosis, cell cycle arrest, longevity and oxidative stress (Lo et al., 2004; Geyer et al., 2000; Renzing et al., 1996). As with *hp53*, overexpression of ΔN*p53* induces apoptosis (Jin et al., 2000; Ollmann et al., 2000). Upon ionising radiation, the ΔN*p53* protein activates the expression of Reaper, a pro-apoptotic protein (Jin et al., 2000; Sogame et al., 2003). Moreover, ΔN*p53* knockout flies have increased retinal sensitivity to UVR-induced apoptosis (Jassim et al., 2003). ΔN*p53* protein together with MNK/Chk2 also
regulates DNA repair by activating the expression of DNA repair genes. Two of the genes, *Ku70* and *Ku80*, are involved in non-homologous end joining repair (Brodsky et al., 2000).

Dmp53 also regulates cell cycle arrest as a response to metabolic stress. Studies have shown that G1/S arrest can occur when cells have low ATP/AMP ratio and this is mediated through cyclin E degradation (Mandal et al., 2005; Owusu-Ansah et al., 2008). The degradation of cyclin E is controlled by Dmp53 which activates the ubiquitin–proteasome system (Mandal et al., 2010).

Dmp53 is also involved in the sensitivity of flies to oxidative stress (Ortega-Arellano et al., 2013). The chemical additive paraquat was used as a food supplement in order to induce oxidative stress. The study showed a strong response in females with a reduction in survivorship, and a Dmp53 RNAi knockdown conferred some rescue of the wild-type phenotype. The knockdown also rescued climbing phenotype which is known to decrease under oxidative stress.

Previous studies have shown that the hp53 protein plays an important role in UVR-induced histone modification (Allison & Milner, 2003; Rubbi & Milner, 2003). Rebollar et al., in 2006, detected a similar response in wild-type *Drosophila* larvae where an increase H3 acetylation was observed after UV irradiation. Dmp53 knockout flies have decreased level of basal H3 acetylation of the K14 and do not respond to UV irradiation (Rebollar et al., 2006). This change in both basal and UVR-induced histone modification suggests a link between histone modification, Dmp53 and nucleotide excision repair (NER); there is an evidence for a similar mechanism in humans (Adimoolam & Ford, 2003). Thermodynamics analysis of hp53 and Dmp53 protein variants also demonstrated a major overlap in function (Herzog et al., 2012). This study concluded that most known structural modifications caused by point mutations in *hp53* (which are related to cancer), have a similar effect on Dmp53.

The *p53* gene has also been implicated in response to radiation in flies (Brodsky et al., 2000; Brodsky et al., 2004; Jassim et al., 2003; Rebollar et al., 2006). Overall, the evidence suggests that *Dmp53* would make a good candidate for an association study to dissect which loci within the gene, if any, contribute to variability in UVR-sensitivity.
In this chapter, I describe set of experiments using strains carrying natural \textit{Dmp53} variants within a similar genetic background (congenic strains) that were tested for their UVR-sensitivity phenotypes in order to identify associated loci.

\textbf{4.2 Methods}

\textit{4.2.1 Fly stocks}

Flies strains were maintained as described in Chapter 2. 27 Isofemale fly strains from natural populations across Europe were used (Fig.4.1 and Table.4.1). A deficiency strain spanning \textasciitilde229Kb including the \textit{p53} locus (stock no. 27375), \textit{Dmp53} null mutant (\textit{p53}\textasciitilde5A-1-4, stock no. 6815), and all other genotypes were acquired from the Bloomington stock centre, USA. The full genotypes are noted in the corresponding methods below.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{map.png}
\caption{Geographical origin of isofemale strains}
\end{figure}
Table 4.1 List of isofemale strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>24_M-BUR 3</td>
<td>Burgundy, France</td>
</tr>
<tr>
<td>91_M-CAV 30</td>
<td>Cavarzere, Italy</td>
</tr>
<tr>
<td>89_M-CAV 4</td>
<td>Cavarzere, Italy</td>
</tr>
<tr>
<td>63_M-COR 25</td>
<td>Corces St Agidius, Italy</td>
</tr>
<tr>
<td>1_M-COSTA 2</td>
<td>Corces St Agidius, Italy</td>
</tr>
<tr>
<td>64_M-4 LUF</td>
<td>Fulda, Germany</td>
</tr>
<tr>
<td>55_M-GOT F1-3</td>
<td>Goteborg, Sweden</td>
</tr>
<tr>
<td>30_M-GOT F1-5</td>
<td>Goteborg, Sweden</td>
</tr>
<tr>
<td>41_M-GOT F1-8</td>
<td>Goteborg, Sweden</td>
</tr>
<tr>
<td>116_M-HOJ 1</td>
<td>Højbjeng, Denmark</td>
</tr>
<tr>
<td>111_M-HOJ 17</td>
<td>Højbjeng, Denmark</td>
</tr>
<tr>
<td>117_M-HOJ 18</td>
<td>Højbjeng, Denmark</td>
</tr>
<tr>
<td>105_M-HOJ 19</td>
<td>Højbjeng, Denmark</td>
</tr>
<tr>
<td>119_M-HOJ 3</td>
<td>Højbjeng, Denmark</td>
</tr>
<tr>
<td>107_M-HOJ 39</td>
<td>Højbjeng, Denmark</td>
</tr>
<tr>
<td>120_M-HOJ 42</td>
<td>Højbjeng, Denmark</td>
</tr>
<tr>
<td>14_M-KNO 29</td>
<td>Knossos, Greece</td>
</tr>
<tr>
<td>103_M-KOR 28</td>
<td>Korpilahti, Finland</td>
</tr>
<tr>
<td>99_M-KOR 31</td>
<td>Korpilahti, Finland</td>
</tr>
<tr>
<td>102_M-KOR 9</td>
<td>Korpilahti, Finland</td>
</tr>
<tr>
<td>93_M-NN 6.15</td>
<td>Nahal Oren, Israel</td>
</tr>
<tr>
<td>29_M-REN 1414</td>
<td>Rende, Italy</td>
</tr>
<tr>
<td>56_M-REN 56-22</td>
<td>Rende, Italy</td>
</tr>
<tr>
<td>13_M-RUT 16</td>
<td>Rutigliano, Italy</td>
</tr>
<tr>
<td>10_M-RUT 4</td>
<td>Rutigliano, Italy</td>
</tr>
<tr>
<td>3_M-STO- F1-10</td>
<td>Stockholm, Sweden</td>
</tr>
<tr>
<td>2_M-STO- F1-8</td>
<td>Stockholm, Sweden</td>
</tr>
</tbody>
</table>

4.2.2 Creating congenic strains carrying natural Dmp53 variants.

Each of the 27 isofemale strains (Table 4.1) were backcrossed with the p53 deficiency strain (Dfp53) carrying a deletion in the Dmp53 region on chromosome 3R (Box 2), resulting in 27 congenic strains that carry a different wild-type Dmp53 allele. Dfp53 strain carries deletion chromosome over a balancer chromosome (described in chapter 3.2, disrupted/deleted genes shown in Table 7.2.1). The congenic strains were backcrossed for six generations, resulting in set of strains whose genetic background is 98% identical.
Box 2: Generating allelic series by backcrossing

Step 1: Crossing Natural strain males with Df strain females.

Multiple natural population males were crossed with females from the Df strain. The resulting progeny carried a natural allele of Dmp53 over a Df or balancer chromosome. The Df chromosome carries a mini-white phenotypic maker with a mutant (null) white background, and the balancer a dominant Stubble marker. In the first cross the dominant Stubble marker allowed for selection against progeny carrying the balancer chromosome. A female progeny carrying a wild-type chromosome from the natural strain over a DF chromosome was selected. In female germ line, recombination occurs throughout the genome except the region spanning the deletion in the Df chromosome. Although the natural strains could be assumed to be isogenic, precaution was taken to select only one female, hence only one allele.
Box continued...

Step 2: Multiple backcross

A single female from step is crossed with a Df strain male resulting in progeny with the following genotypes: (i) the natural \textit{Dmp53} allele chromosome over the balancer or (ii) the Df chromosomes, (iii) the balancer over the Df chromosome from the female (The double Df chromosome progeny are \textit{not viable}), and progeny carrying the balancer can be distinguished by dominant phenotypic marker \textit{Stubble}. The progeny carrying the balancer chromosome are selected against for females carrying the natural \textit{Dmp53} allele and the Df chromosome from the male parent. Recombination occurs once again in the female germ line and the natural allele chromosome increases in similarity to the Df chromosome. The female is then crossed with a Df male and this is repeated six times.

Select a virgin female w/o Bal

Further recombination occurs after each backcross to the Df strain and chromosome carrying the natural allele becomes more similar to Df strain (red)
Step 3: Creating isogenic strains

The final step was achieved by selecting a male and a female from the cross in step 2. A male and a female carrying both the balancer chromosome with dominant maker Stubble and the natural \( Dmp53 \) allele chromosome were inbred. The flies carrying the balancer and the Df could be selected against using the 'mini-white' marker which gives the flies orange eyes on a mutant (null) white background. The progeny of this cross which is homozygous for the natural \( Dmp53 \) allele were inbred to create an isogenic stock.

Each backcrossed strain was given a unique ID and the geographic location recorded.

The final step was achieved by selecting a male and a female from the cross in step 2. A male and a female carrying both the balancer chromosome with dominant maker Stubble and the natural \( Dmp53 \) allele chromosome were inbred. The flies carrying the balancer and the Df could be selected against using the 'mini-white' marker which gives the flies orange eyes on a mutant (null) white background. The progeny of this cross which is homozygous for the natural \( Dmp53 \) allele were inbred to create an isogenic stock.
4.2.3 Sequencing the p53 alleles
For each congenic strain, the complete Dmp53 gene (4432bp) and additional small upstream sequence (307bp) was amplified and sequenced. The details of each of the steps are given below.

4.2.3.1 DNA extraction from multiple flies
DNA was extracted from groups of ten flies using the Insect DNA kit (Omega Bio-Tek). The kit was used to produce long high quality DNA fragments for the downstream sequencing application. The manufacture’s protocol was used with some following changes. Male flies were collected in groups of 10 and placed into 1.5ml microcentrifuge tubes. The flies were frozen using liquid nitrogen and then stored in -20°C. The flies were homogenised in 100µl of lysis buffer (CTL with Proteinase K) using autoclaved micropestles with hand motorised homogeniser. Once the flies were homogenised, 275µl of lysis buffer was added. The homogenate was left to incubate at 60°C for 30 minutes. On the final step 50µl of pre-warmed (70°C) elution buffer was used. The entire protocol was carried out on ice and the eluted DNA (100µl) was stored at -80°C.

4.2.3.2 DNA amplification by Polymerase Chain Reaction (PCR)
The PCR reactions were carried out using the G-STORM GS4 Multi Block Thermal Cycler. A high fidelity polymerase kit (Phusion® High Fidelity DNA polymerase) was used in all the experiments. PCR reactions were set up with the following reagents: Ultrapure Water\textsuperscript{19} (34µl), Buffer 5x (10µl) dNTPs (10mM, 1µl), Forward Primer F2 (10µM, 2.5µl), Reverse Primer R2 (10µM, 2.5µl), DNA (genomic extraction, 0.5µl) and Phusion® polymerase (1 unit, 0.5µl). Primers are listed in Table.7.2.1. For the primary PCR, the following PCR touchdown program was used: Initial denaturation at 98°C for 30 secs, touchdown (4 cycles) at 67°C-64°C for 30 secs, PCR cycles (35x)- denaturation at 98°C for 10 secs, annealing at 63°C for 20 secs, elongation at 72°C for 90 secs and

\textsuperscript{19} Ultrapure water: Double stilled water purified to 18.2ohms, Type 1 water.
final elongation at 72°C 5 min. The PCR fragments were visualised on an agarose gel and purified (Ch 4.2.3.3, p.74).

A secondary PCR was carried out to create 3 smaller fragments: 5’ (2.54kbp, primers R2/RS3), M (2.77kbp, primers FS3/RS5) and 3’ (2.88kbp, primers FS4/R2) (Table.7.2.1). Three 50µl reactions were run for each fragment using a high fidelity polymerase (Phusion® High Fidelity DNA polymerase) with the same reagents as the primary PCR. The primers are shown in Table.7.2.1 and Fig.4.3A. For the secondary PCR, the following PCR program was used: Initial denaturation at 98°C for 30 secs, followed by 35 PCR cycles: denaturation at 98°C for 10 secs, annealing at 62°C for 20 secs, elongation at 72°C for 90 secs, and a final elongation at 72°C for 5 min.
Fig. 4.3 Amplification and sequencing Dmp53.
Schematic of primers used for PCR and sequencing along the Dmp53 gene region of congenic strains (blue). A. The primers for the primary PCRs (F2/R2 primers, 5.2kbp) which amplified the entire gene and ~400bp flanking regions. Also shown are the primers for the secondary PCRs which amplified 5’ region (primers R2/RS3, 2.54kbp), M region (primers FS3/RS5, 2.77kbp) and 3’ region (primers FS4/R2, 2.88kbp). B. The sequencing primers for the 5’ fragment. C The sequencing primers for the M fragment. D. The sequencing primers for the 3’ fragment. Various combinations of sequencing primers were used until clean sequence traces were received (Bechman Coulter genomics); it is likely some gene fragments had polymorphisms in primer binding sites. These polymorphisms may interfere with primer binding and reduce PCR efficiency.
4.2.3.3 Visualisation of DNA on an agarose gel
The PCR fragments from the primary/secondary PCR were run on an agarose gel to check the amplification. The gel was made by melting a mixture of 0.5g of agarose with 50ml of 0.5x TBE buffer (1x TBE buffer - Tris 0.089M, EDT-Na salt 0.002M, Boric acid 0.089M). Once melted, the mixture was allowed to cool, and 1.5µl of ethidium bromide (1%) was added. The gels were run at 100V for 1 hour and then visualised under UV light in the trans-illuminator (Synegene). The size of the fragments were estimated using the DNA ladder.

4.2.3.4 Pooling DNA amplifications and sequencing
The PCR products from the secondary PCR were pooled together and then column purified (Cycle Pure Kit, Omega Bio-tek). The purified DNA products were stored in elution buffer at -20°C. The sequencing of the fragments (by Bechman Coulter genomics) required a minimum of 10µl of 40ng/µl DNA template for each fragment. For quality control, each PCR product (pooled DNA) was run on an agarose gel (Ch 4.2.3.3, p.74) and the concentration was measured using a nanodrop (Nanodrop 2000, Thermo Scientific), and Qubit 2.0 (Invitrogen) systems. The 27 congenic strains were sequenced using the three fragments mentioned earlier (Fig.4.3 and Table.7.2.1). Various combinations of primers were used as unexpected SNPs were discovered in some primer binding sites.

4.2.4 Assembling and aligning the sequences
The short 800-1000 bp sequences generated by Sanger sequencing method (Beckman Coulter genomics) were stored in trace (.ab1) files. The sequences were trimmed and assembled using CodonCode Aligner software (http://www.codoncode.com/aligner/). The sequences were visually checked and amended. The sequences were all assembled successfully using a reference sequence (BDGP, 2006) and were exported as consensus fasta files.
The sequences were aligned using *emma*, a wrapper which is part of the EMBOSS package (http://emboss.sourceforge.net/) that runs the ClustalW multiple sequence alignment program (Thompson *et al.*, 1994). The data were uploaded and run on a high performance computing cluster (SPECTRE\textsuperscript{20}) at the University of Leicester. The fasta sequences were concatenated into one file and *emma* was used to create an alignment file (.aln). The file was then visualised in the program Mega5.2 (Tamura *et al.*, 2011), sequence ends were trimmed and SNPs were extracted (minimum count of 4). A diagram of SNPs was created using Excel (Microsoft) (Table.7.2.3), and the SNPs were then annotated using data from ensemble genome browser (http://www.ensembl.org/) (Table.7.2.2).

### 4.2.5 Phenotyping the congenic strains

The 27 congenic strains were phenotyped using three different UVR-sensitivity assays. The assays used both the CPS+ Solar simulator and a narrowband UVB lamp (311nm, PHILIPS TL 2ft, 20W), which was sourced from a local distributor. The assays are described below.

#### 4.2.5.1 Adult viability Assay (Solar simulator)

An adult viability assay was also established. The adult viability assay uses fully grown adults under the same set-up as the embryo viability assay. F2 males and females were separated as virgins and placed into fresh media vials (Chapter 2). The next day 30-50 flies were placed into vials without food.

The experiment was always carried out at 10:00 (2 hr after lights-on). An optimal dose of 485W/m\textsuperscript{2} for 30 min was established. After irradiation, the flies were transferred to vials (15 in each) and incubated at 25°C. The vials were changed after 3 days and viability scored every day. The viability was calculated as a percentage of viable flies over starting total (15 flies) and followed for 5 days (there is little change in...

\textsuperscript{20} SPECTRE- Special Computational Teaching and Research Environment, a high performance computer at the University of Leicester
viability after this time). Data was collected for control and experimental conditions (N replicates=5).

4.2.5.2 UVB assays

UVB assays were carried using **PHILIPS (TL 2ft 20W) 311 nm lamp**. The lamp produces narrowband light and is small enough for experiments to be carried out in an incubator. The assays were carried out for reasons of practicality. The UVB assays were carried out in same way as the adult viability assay (see above), but with a UVB lamp and using only males. In a preliminary experiment, adult viability was tested to check for a UV response using the UVB lamp. The flies were placed into a 25°C incubator with a UVB lamp (12KJ/m², 311nm, PHILIPS TL 2ft, 20W) for 3hrs. The lamp was pre-warmed for 20mins and the controls were kept in a light box within the incubator, the temperature consistency was checked with a data logger (±1°C). Fig.4.7 shows flies exposed to 12KJ/m² of UVB-supplemented irradiation respond to the light. There is phenotypic variation between CS (Canton-S), Af (strain derived from natural Zambian population), p53[5A-1-4] (Dmp53 null mutant) and Df(3R)BSC803 (strain used to create congenic strains). Notably, variation in response was observed between two congenic strains, **KNO29** (Knossos, Greece) and **MHOJ17** (Højbjerg, Denmark) (Fig.4.7B), which is a good indicator that phenotypic variation in UVB-sensitivity may be detected amongst the congenic strains.
Fig. 4.7 Adult viability after UVB irradiation

Adult viability after 12KJ/m2 irradiation of UVB with strains CS (Canton-S), Af (strain derived from natural Zambian population), p53[5A-1-4] (Dmp53 null mutant), Df(3R)BSC803 (strain used to create congenic strains), KNO29 (congenic strain 14), and MHOJ17 (congenic strain 111). The error bars are the SEM of 5 replicates per strain. A. The homozygous Dmp53 knockout (p53[5A-1-4]) was the most sensitive to UVB irradiation and the hemizygous Dmp53 strain (Df(3R)BSC803) was less sensitive than the wild-type strains (CS and Af). B. Two congenic strains show variable response to UVB irradiation, KNO29 (Knossos, Greece), and MHOJ17 (Højbjerg, Denmark).
Oxidative stress assay (UVB)

The oxidative stress of flies was measured using a kit (OxiSelect™ Protein Carbonyl Fluorometric Assay, CELL BIOLABS, INC). **Males** were separated into vials of 15 flies for both control and experimental conditions ($N_{replicates}=3$). The flies were exposed to 12KJ/m$^2$ UVB-supplemented light for 3hrs (as described earlier). After the incubation, the flies were transferred to microcentrifuge tubes, frozen using liquid nitrogen and stored at -80°C.

The manufacturer’s instructions were followed to measure oxidative stress by calculating the protein carbonyl concentration using a fluorometric assay. The assay works on the principal that proteins may undergo oxidation in environments with high ROS. These protein carbonyl derivatives are useful biomarkers for oxidative stress as they are fairly stable. The assay requires large amounts of protein (10mg/ml) which is extracted and mixed with a fluorophore. This fluorophore binds at a 1:1 ratio and is used to measure the amount of protein carbonyl in the samples. The levels of oxidative stress are measured as a ratio of protein carbonyl content, over the total protein in the samples. The protein carbonyl levels are measured using a fluorometer at 485/538 nm and the assay was validated for UVB-response ($F(3,8)=192.5$, $p<0.0001$) using wild-type (*Canton-S*) and *Dmp53* knockout flies (*p53[5A-1-4]*) (Fig.4.8).
Fig.4.8 Validation of oxidative stress assay.
Oxidative stress measured using Protein Carbonyl concentration (nM/mg) in wild-type CS (Canton-S) flies and Dmp53 knockout flies (p53[5A-1-4]), under control (CS-C and p53[5A-1-4]-C) and UVB irradiated conditions (12KJ/m²). There is significant difference between the conditions (F(1,8)=525.2, p<0.0001) and genotypes (F(1,8)=52.33, p<0.0001) but there is no significant interaction (F(1,8)=0.002084, p<0.96). SEM of 3 replicates is shown for each strain.

**Real-time PCR assay (UVB)**

Real-time PCR was carried out on male flies irradiated with UVB, following the same protocol as in the oxidative stress assay (see above). The samples were frozen using liquid nitrogen as quickly as possible after irradiation. The Dmp53 transcript was quantified for each congenic strain in both control and experimental conditions (N_repeats=3). The extraction of RNA, cDNA synthesis and qPCR are described below.

**RNA extraction**

A kit was used to extract whole body RNA from frozen samples (Promega SV total RNA Isolation). The flies (15 males) were homogenised with 100µl of lysis buffer and 75µl of
lysis buffer was added to the samples and kept on ice. After extraction of RNA (40µl), the samples were stored in -80°C. The quality of the RNA samples was checked using the NanoDrop (as in Ch 4.2.3.4, p.74). The samples were all of good quality with 260/280 ratio of ~2.0.

First-strand cDNA synthesis

The reverse transcription of RNA samples was carried out using SuperScript™ II RT kit (Invitrogen). The kit produced first-strand cDNA synthesis using random primers (Random Hexamers, 50 µM, Invitrogen™) with 20µl of RNA.

The RNA was first treated with a DNase (TURBO™ DNase). 2µl of TURBO™ DNase were added to 20µl RNA samples with 2.4µl of 10x TURBO DNase buffer. The reaction was mixed gently by pipetting up-and-down and incubated at 37°C for 30 min. A volume of 5µl of inactivating reagent was added to stop the reaction. The reaction was mixed again and allowed to incubate at room temperature for 5mins. The reaction was then centrifuged at 10,000g for 1.5 min and the clear liquid was transferred to a new microcentrifuge tube.

cDNA synthesis reaction was carried out in 200µl microcentrifuge strips. Each reaction was set up as follows: 1µl of random primers (Random Hexamers (50 µM), Invitrogen), 1µg of RNA, 1µl of dNTPs (10 mM each). The reaction was heated to 65°C for 5mins and then chilled quickly on ice. 4µl 5x first-strand synthesis buffer was added with 2µl 0.1M DTT and 1µl of RNase OUT (40units/µl Invitrogen). The strips were centrifuged briefly. After the strips were incubated at 45°C for 2 min, 1µl of SuperScript™ II reverse transcriptase was added and the reaction volume was made up to 20µl using ultrapure water.\(^{21}\) The reactions were mixed by pipetting up-and-down and then centrifuged briefly. The strips were then incubated at 25°C for 10 min, followed by 42°C for 50 min and then at 70°C for 15 min, to inactivated the reverse transcriptase.

\(^{21}\) Ultrapure water- Double stilled water purified to 18.2ohms, Type 1 water.
To remove the RNA, 0.5µl of RNase H (NEB) was added to the mixture and incubated at 37°C for 30 mins and inactivated by incubating at 65°C for 20 min. The cDNA was then diluted 10x using elution buffer.

**Real-time PCR reaction**

The standard curves for the real-time PCR were made using primers listed in Table.4.2. Standard curves were made for both the target and the reference genes by running 6 50µl PCR reactions, separating the PCR products on an agarose gel, cutting and then purifying the DNA using a kit (as in 4.2.3.2 and 4.2.3.3). The standard curves included DNA standards that ranged from 300,000 to 30 copies of the amplicons (5 standards) and were created by serial dilution.

The primers were designed to overlap exon-exon junctions, and the primers for Dmp53 (R1-tr1-p53/F1-tr1-p53) were designed to include all known transcripts. The Rp49 gene transcript was used as a reference (primers RpL32_F/RpL32_R) (Table.4.2). The real-time PCRs were set-up with duplicates for each sample including the standard curves. The reactions were made using the following reagents (total volume 20µl): Brilliant II QPCR Low ROX Master Mix (Agilent technologies) 12.5ul, cDNA template 5ul, forward primer 1ul, reverse primer 1µl and ultrapure water22 5.5ul. The reaction kept on ice through the setup. The Opticon System (MJ Research) real-time PCR machine was used to run the reactions using the following program: Initial denaturation at 94°C for 15 min, followed by 42 cycles of: denaturation at 94°C for 15 secs, annealing at 60°C for 30 secs, and elongation at 72°C for 30 secs. A melting curve taking readings every 0.2°C between 50-95°C has also been produced.

**Table.4.2 Primers for real-time PCR**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>RpL32_F</td>
<td>CACTTCATCCGCCACCAGT</td>
</tr>
<tr>
<td>RpL32_R</td>
<td>CGCTTGTTCGATCCGTAAACC</td>
</tr>
<tr>
<td>R1-tr1-p53</td>
<td>AGCTCTCGCGCATTTTTT</td>
</tr>
<tr>
<td>F1-tr1-p53</td>
<td>CGTGTGTTCTTTGCTTCTC</td>
</tr>
</tbody>
</table>

22 Ultrapure water: Double stilled water purified to 18.2ohms, Type 1 water.
Once the reference gene (*rpL32*) and the target gene (*p53*) were quantified, a target/reference ratio was calculated per sample (i.e. normalised *p53* expression values) using the Opticon Software (Monitor 3, Bio-Rad) and Excel 2010 (Microsoft).

### 4.2.6 Association analysis

The *p53* allele sequences and the phenotypic information were used in association analysis that was carried out for the UVR responses: adult viability (males and females), oxidative stress (males) and change in *Dmp53* expression (males). Two different algorithms were used. The first analysis was carried out using a permutated ANOVA method (Anderson & Braak, 2003) implemented by the TASSEL5.0 software (Bradbury et al., 2007). TASSEL5.0 uses general linear models and performs association analyses according to the permutation method by Anderson and Ter Braak (2003). The software used data from all replicates and 1000 permutations were carried in each analysis.

The second method for association analysis was based on identifying and comparing major haplotypes (Templeton et al., 2005). A gene tree of the *p53* alleles was using Mega5.2 and FastML (Ashkenazy et al., 2012) software. Two major clades (haplotypes) were identified, and the strains consisting each of the haplotype group were compared using a nested ANOVA (implemented in 'R'). For each test, the scores were resampled across the groups/clusters and the test repeated many times (N=1000). The 95th percentile of the F-scores was used as a critical value and any observed F-score below this value was not considered significant. Mega5.2 and FastML software were used to infer ancestral sequences (haplotypes) at the divergence of the major clades.

The SNPs identified using TASSEL 5.0 and haplotype reconstruction (Mega.5.2) were drawn onto a schematic of the *Dmp53* gene (4432bp), using Geneious software (Kearse et al., 2012) (http://www.geneious.com/) and PowerPoint 2010 (Microsoft).
Testing for the signature of selection

Various neutrality tests were used for testing whether polymorphisms in p53 is driven by natural selection in an African population (Rwanda, DPGP) and the congenic strains. These included Tajima’s D (Tajima, 1989), McDonald–Kreitman (MK) (McDonald & Kreitman, 1991) and the Hudson-Kreitman-Aguade (HKA) (Hudson et al., 1987) tests. For the MK test, three D. simulans p53 sequences were used (3R:18,875,687-18,879,730). A further three sequences were available for analysis but were not included in the analysis due to high level of missing data. There is little variation amongst the six strains (68 SNPs), only one site has no missing data and can inform the analysis (Table.7.2.4). For the HKA test the following D. melanogaster regions were used: p53 (18,875,379-18,879,804), intergenic region (3R:18,842,092-18,845,622) using BDGP5. The following D. simulans sequences were used: GD18411 (3R:18,875,687-18,879,730), and intergenic region (3R:18,656,713-18,659,322) using WUGSC 1, (INSDC Assembly GCA_000259055.1, Apr 2005).

The analysis was carried out using DNAsp software (Librado & Rozas, 2009) (http://www.ub.edu/dnasp/). For the Tajima’s D, 95% confidence intervals were calculated by coalescent simulations using 1000 permutations, and by feeding in recombination rates and θ values using Tajima’s D and recombination analysis, per population with DNAsp. The tests were carried out on sequence data from a single outbred Rwandan population (27 strains), and from the congenic strains. The graphical representation of Tajima’s D was carried out in Excel 2010 (Microsoft) and overlaid with the Dmp53 gene schematic, the LD was also visualised in Excel using this data.

4.2.8 Visualisation and analysis of 3D protein structure of Dmp53

The protein structure of Dmp53 was predicted by Phyre2 online protein prediction tool (Kelley & Sternberg, 2009), and analysed using Phyre2 and SIFT (Sorting Intolerant From Tolerant, http://sift.bii.a-star.edu.sg/) (Kumar et al., 2009). The data was visualised and annotated using PyMOL software (http://pymol.org/).
4.3 Results

4.3.1 Adult viability Assay (Solar simulator)
Substantial phenotypic variation was observed between the congenic strains in both males and females (Fig.4.9). The median values of adult viability ranged from 63.33% to 100% for both males and females. For both sexes, the majority of the strains have a median score of ~95%. The variation between the strains was tested using ANOVA, which showed significance for males \(F(24,95)= 15.65, \ p<0.001\) and females \(F(25,99)=12.66, \ p<0.001\). As there is phenotypic variation between the strains under experimental conditions, an association study was carried out.
Fig. 4.9. Adult viability following solar irradiation
The adult viability of males (A) and females (B) from 26 congeneric lines carrying different p53 natural alleles was scored after irradiation using the CPS+ solar simulator (485W/m²) for 30 min. The viability was scored as the number of adults that survived for 5 days (Survivorship). The range (blue/red lines) and median scores (black dots) of 5 replicates are shown, and the strains in each graph have been ordered from lowest to highest median score (Note, A and B are not in the same order).
4.3.2 Oxidative stress assay (UVB)

The protein carbonyl levels of the congenic strains (only males were tested) vary under both control ($F(24,50)=20.34$, $p<0.001$) and experimental conditions ($F(24,50)=10.7$, $p<0.001$) (Fig.4.10). In control experiments, the flies were kept in the same incubator with the UVB lamp, but not exposed to the light. Under experimental conditions, the flies were exposed to $12KJ/m^2$ of UVB in 3hrs. The congenic strains showed similar variation in oxidative stress as the controls (Fig.4.10B). The only noticeable difference between the control and experimental conditions is the levels of stress, which are $\sim1.3x$ higher after irradiation. The similar variation observed in the two conditions suggests that they might be correlated, but this is not the case ($t =1.08$, $df=23$, $p$-value$=0.29$), indicating that part of the variation is due to variability in the UVR response. Furthermore, using the control conditions as a covariate, an ANCOVA was performed and also detected significant variation amongst the DGRP strains ($F(24,49)=10.51$, $p<0.0001$).

To account for the variance under control conditions, normalised scores were calculated. These were based on the difference between each score under experimental conditions and the median score under control conditions. Fig.4.10C shows considerable phenotypic variation between the strains ($F(24,47)=6.41$, $p<0.001$). In contrast to the non-normalised scores (Fig.4.10B), which illustrate a 2-fold difference in the range of median scores, the normalised median scores (Fig.4.10C) show a 7-fold difference in the range from $\sim200-\sim1400nM/mg/ml$. As there is phenotypic variation between the strains, an association study was carried out.
Fig. 4.10. Oxidative stress in the congenic strains

The oxidative stress of 24 congenic strains (males only) under (A) control (no UV) and (B) experimental conditions (UVB light (12KJ/m²) for 3hrs). The oxidative stress was measured using protein Carbonyl content. The range (green/red lines) and median scores (black dots) of 3 replicates are shown. The strains have been ordered from lowest to highest median score (note: strains in A, B and C are not in the same order).

A. Protein carbonyl content of males under control conditions. There is a 2-fold range between the median scores and substantial variation (F(24,50)=20.34, p<0.001). These scores represent the background stress under control conditions.

B. Protein carbonyl content of males under experimental conditions. There is a 2-fold range between the median scores and the variation is similar to the controls (F(24,50)=10.7, p<0.001) (Fig.4.10). These scores show a similar profile to the controls except the levels of oxidative stress detected are higher (~500nM/mg/ml).

C. The normalised oxidative stress scores. For each strain, the scores under experimental conditions were normalised with the median score of the control. The normalised median scores for protein carbonyl content illustrate substantial range (7-fold) (F(24,47)=6.41, p<0.001) and variation to oxidative stress.
4.3.3 Real-time PCR assay (UVB)

The real-time PCR assays were carried out to assess the expression of Dmp53 transcripts after irradiation (males only). Under control conditions, the expression levels did not vary significantly ($F(24,43)=1.41$, $p=0.16$) between the congenic strains (Fig.4.12A). The background levels of Dmp53 (in whole body) are fairly low with many strains displaying little or no expression.

After irradiation, there were some changes in p53 expression levels, but there was significant ($t(23)=2.62$, $p=0.015$) but weak correlation (0.48) with the control levels. The expression differed significantly between the strains ($F(24,47)=3.08$, $p<0.001$). Furthermore, using the control conditions as a covariate, an ANCOVA was performed and also detected significant variation amongst the DGRP strains ($F(24,39)=3.73$, $p<0.0001$). To account for the basal transcriptional variation, normalised scores were calculated (Fig.4.13).

The normalised expression scores were derived from the difference between each score under experimental conditions and the median score under control conditions. Fig.4.13C shows that the normalised expression levels of Dmp53 after irradiation decreased in some strains, and increased in others, though approximately half the strains show no/little change. Given the considerable phenotypic variation between the strains that was detected ($F(24,50)=3.19$, $p<0.001$), this phenotype has been further analysed by association analysis (see below).
Fig. 4.12 Relative expression of Dmp53 in congenic strains

The relative expression levels of Dmp53 transcripts in 24 congenic strains (males only) under A. control (no UVR), and B. experimental conditions using a UVB light (12KJ/m²) for 3 hr. The expression levels were measured by real-time PCR using RP49 transcript as a reference, the range (green/red/blue) lines and median scores (black dots) of 3 replicates are shown, and the strains have been ordered form lowest to highest median score (A, B and C are not in the same order). A. Relative expression levels of Dmp53 transcript under control conditions. There is a small variation in background levels of Dmp53 transcripts in the congenic strains; but this is not significant (F(24,43)=1.41, p=0.16). B. Relative expression levels of Dmp53 transcripts under experimental conditions. There is some variation of Dmp53 expression under experimental conditions and this is significant F(24,47)=3.08, p<0.001). C. The normalised relative expression levels of Dmp53 transcripts in 24 congenic strains. The expression levels show variation between the congenic strains (F(24,50)=3.19, p<0.001). Approximately half the strains showed no/little change in expression after irradiation.
4.3.4 Association analysis

Association analysis of oxidative stress using normalised scores for protein carbonyl content revealed a single SNP, A4123C (Table.7.2.2) associated with oxidative stress in males (p= 0.012). This polymorphism in exon 8 results in a missense mutation coding for an amino acid replacement from aspartic acid to glutamic acid (D305E). The oxidative stress-associated SNP 4123 showed a considerable difference in the size of median scores between the variants (Fig.4.14), with the D305 variant showing a 4–fold higher level of protein carbonyl content (i.e. increased oxidative stress). In contrast, neither adult viability, nor the differential expression of Dmp53 showed a significant association with SNP 4123 (Appendix Fig.7.2.3).

![Fig.4.14 UVR induced oxidative stress associated with SNP 4123 in p53.](image)

**Fig.4.14 UVR induced oxidative stress associated with SNP 4123 in p53.**

Boxplot of protein carbonyl content (nM/mg/ml) scores of congenic strains carrying variants of SNP 4123, which was found to show significant association using the TASSEL algorithm. The A allele encode the 305E variant (glutamic acid) and the C allele encode the D305 (aspartic acid). The bold line in the box indicates the median, the bottom of the box indicates the 25th percentile, and the upper edge of the box the 75th percentile. The whiskers indicate 1.5 times the IQR.

4.3.5 Haplotype association

The Dmp53 gene tree was constructed using the sequences of the congenic strains using the maximum likelihood algorithm, and is shown in Fig.4.16. The p53 alleles were
clustered into two major clades. Some alleles from the same population appear in both clades. For example, most of the alleles from the MHOJ population (Højbjeng, Denmark) were grouped in clade A, but one allele is present in clade B. A and B are sister clades diverging near the base of the tree, with strains CAV 30 (Cavarzere, Italy) and MKOR 28 (Korpilahti, Finland) acting as an outgroup.

Using the scores of each of the three UVR-sensitivity phenotypes, an association analysis (ANOVA, Ch 4.2.6, p.82) was performed by comparing the phenotype of strains of clades A with those in clade B (Fig. 4.16). Adult viability in females and oxidative stress were significantly different between the two haplotypes (F(1,118)=3.93, p=0.049 and F(1,62)=6.08, p=0.016 respectively). The effect size of adult viability was rather small, with ~5% difference in medians between clusters A and B. Adult viability in males and p53 expression level did not show any significant association.

**Fig.4.16 Phylogenetic tree of Dmp53 gene of the congenic strains.** Gene tree of Dmp53 alleles of the congenic strains, with two labelled clades (A and B). A and B, are sister clades diverging at the base of the tree. The tree was constructed using maximum-likelihood method (Mega5.2). The arrows indicate the internal nodes for which the ancestral sequences have been inferred and bootstrap values.
The ancestral sequences of the two sister clades A and B were inferred and compared, revealing five differences (SNPs) between the two haplotypes (Table.4.3). As the SNPs are inferred using ML, they are a best guess of the ancestral state of the gene under the model.

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>1069</th>
<th>1604</th>
<th>3940</th>
<th>4364</th>
<th>4388</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>G</td>
</tr>
</tbody>
</table>

### Table.4.3 Haplotype sequence inferred by ML

4.3.6 Summary of SNPs associated with UVR-sensitivity

The analyses of SNPs in the Dmp53 gene with UVR-sensitivity phenotypes suggested one SNP (TASSEL) and two possible haplotypes (consisting of five SNPs, see Fig. 4.19, Table 4.5) are under association. The oxidative stress-associated SNP 4123, lies with exon 8 and results in non-synonymous change (D305E). The Gr94a gene, a candidate for taste perception in Drosophila (Clyne et al., 2000), lies within intron 5 of Dmp53. SNP 4123 is an upstream variant of Gr94a and may affect its transcription, but this gene is unlikely to be involved in an UV response.

The two haplotypes identified through inferring ancestral sequences of clades A and B, consist five SNPs which are located in introns. Interestingly, SNPs 4364, 4388, 1069 and 3940 show linkage across multiple introns; SNP 1604 does not show any linkage.

**Fig. 4.19. Schematic of Dmp53 gene with UVR-sensitivity associated SNPs**

The SNP in the 8th exon of the Dmp53 gene associated with oxidative stress and results in a non-synonymous change (diamond arrow, aspartic acid (D)/glutamic acid (E)). The other 5 SNPs are intron variants and were inferred by ancestral sequences using a maximum likelihood tree of Dmp53 alleles. The SNPs represent the divergence of the sequences in clades A and B (Fig.4.16).
### Table 4.4 Significant SNPs in Dmp53 associated with UVR response

<table>
<thead>
<tr>
<th>SNP</th>
<th>Genomic position</th>
<th>Annotation in Dmp53</th>
<th>Alleles (complement)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1069</td>
<td>3R:18879043</td>
<td>Intron variant (2\textsuperscript{nd})</td>
<td>T/C</td>
</tr>
<tr>
<td>1604</td>
<td>3R:18878508</td>
<td>Intron variant (5\textsuperscript{th})</td>
<td>A/G</td>
</tr>
<tr>
<td>3940</td>
<td>3R:18876172</td>
<td>Intron variant (7\textsuperscript{th})</td>
<td>A/T</td>
</tr>
<tr>
<td>4123</td>
<td>3R:18875989</td>
<td>Missense variant (8\textsuperscript{th} exon)</td>
<td>G/T</td>
</tr>
<tr>
<td>4364</td>
<td>3R:18875748</td>
<td>Intron variant (8\textsuperscript{th})</td>
<td>G/A</td>
</tr>
<tr>
<td>4388</td>
<td>3R:18875724</td>
<td>Intron variant (8\textsuperscript{th})</td>
<td>C/G</td>
</tr>
</tbody>
</table>

#### 4.3.6 Analysis of Dmp53 sequence polymorphism

Neutrality tests were performed on the Dmp53 locus of the congenic strains and also on sequences from a single African population (Rwanda). The congenic strains carry Dmp53 alleles from multiple European populations, though this is not ideal, comparisons between the two populations allowed for some interesting insights.

The African and congenic population’s overall *Tajima’s D* are not significantly deviating from a neutral expectation; $D=-0.97$, $p>0.10$ and $D=-0.38$, $p>0.10$ respectively. However, significant peaks were observed along the gene, detecting both positive and negative values of $D$ (Fig. 4.20). In the African population (Fig. 4.20A), the second intron variant (SNP 1604) lies in a region of positive *Tajima’s D* peak that may indicate the signature of balancing selection (or sudden population contraction) (Hartl, 2000), and the remaining intron variants lie in regions of negative *Tajima’s D* that may suggest it is under purifying selection (or a population expansion after a recent bottleneck) (Hartl, 2000). The missense mutation (4123) lies in region of significant negative *Tajima’s D* that may indicate purifying selection. By contrast, in the congenic population, SNP 1604 lies in a region showing no deviation from neutrality, and the missense mutation lies in a region of a positive Tajima $D$ (Fig. 4.20B); these regions do not show significant deviation from neutrality. The *Tajima’s D* in the congenic population also revealed a large region of positive $D$ (balancing selection) across the 5\textsuperscript{th} intron, whilst largely negative $D$ (purifying selection) was detected in the African population at this locus. It seems that the non-synonymous SNP 4123, lies within regions that are under different selection pressures in the two populations, possibly purifying selection in the African population and balancing selection in the congenic
population. The higher levels of UVR in Africa may contribute to the different selection pressures at this locus.

![Graph of Tajima's D along Dmp53](image)

**Fig. 4.20 Tajima's D along Dmp53**

Tajima’s D plotted along Dmp53 using data from (A) a Rwandan population (27 strains) and (B) the congenic strains. The 95% confidence intervals were calculated by coalescent simulations. Tajima’s D is calculated every 25bp with a sliding window of 200bp and the graph has been overlaid with Dmp53 gene schematic with UVR-sensitivity-associated SNPs.
Three *p53* sequences from *D. simulans*, a closely related species, was used to perform the McDonald-Kreitman test of neutrality on both populations. In the African population, I observed 29 fixed and 25 polymorphism sites as silent (synonymous) substitutions (*Ds,Ps*), and 46 fixed and 8 polymorphic sites as non-synonymous substitutions (replacement). The neutrality index (NI), ratio of (*Pn/Ps*)/*(Dn/Ds)*, indicates deviation from neutrality in this test. Under neutrality, the ratio of non-synonymous to synonymous variation within (*Dn/Ds*) and between (*Pn/Ps*) species is considered equal, i.e. NI=1 (Stoletzki & Eyre-Walker, 2011). The test revealed significant low NI (NI= 0.2402, *p*<0.001) which suggests positive selection is acting on the *Dmp53* locus in the African population. An equation derived from this test (Smith & Eyre-Walker, 2002) produces an *α* value which indicates the level of positive selection (*α* = 1-*((DsPn/DnPs)). The values of *α* can range from $-\infty$ to 1, positive values of *α* are observed under positive selection, and *α* is zero under neutrality (negative values occur through error). The *α* value for the African population is relatively high (*α* =0.7598), suggesting strong positive selection at the *Dmp53* locus.

In the congenic population, I observed 31 fixed and 23 polymorphic sites as silent (synonymous) substitutions (*Ds,Ps*), and 47 fixed and 12 polymorphic sites as non-synonymous substitutions. The test suggests the *Dmp53* locus of the congenic population is also under positive selection with a significantly low NI (NI= 0.344, *p*= 0.01), and a moderate *α* value of 0.656.

As the complete genome sequences of the African population are available, the Hudson-Kreitman-Aguade (HKA) test was performed. The HKA test compared two loci from the African population and *D. simulans*, the locus of interest- *Dmp53*, and an intergenic locus which is presumed to be under neutrality (see Methods for sequence coordinates). Neutrality is tested by comparing divergence between species with polymorphisms within species (Hartl, 2000). The *Dmp53* locus did not show any significant deviation from a neutral expectation.

Interestingly, the two populations show alternative pairwise linkage disequilibrium (LD) patterns. The pairwise LD matrix of the African population reveals several blocks of significance (Fig.4.21A). A small block of LD at the N-terminus breaks
down at SNP 1069, and this is followed by several adjacent SNPs showing linkage. A larger LD block is observed at the C-terminus which is also broken up by a single SNP. The oxidative stress-associated-SNP 4123 shows linkage with both the N- and C-terminus blocks. The LD block encompasses UVR-sensitivity-associated SNPs 3940, 4123, 4364 and 4388. The pairwise LD pattern of the congenic population reveals a single block of significance in the middle of the gene, and the UVR-sensitivity-associated SNPs flank this region (Fig.4.21B). The oxidative stress-associated-SNP 4123 is completely free from linkage. This region also overlaps with balancing selection as measured by Tajima’s $D$ (Fig.4.20B). The fact that the linkage between sites extends across the gene, even though LD usually decays over relatively short distances (Carbone et al., 2006; Long et al., 1998) in Drosophila, may further suggest the signature of balancing selection (Charlesworth et al., 1997).
Fig. 4.21. Linkage disequilibrium across Dmp53

Significant LD amongst Dmp53 SNPs is shown as matrices of pairwise comparisons using informative SNPs from (A) a single Rwandan population (27 strains) and (B) congenic population. The total physical distance is ~4kb and the significance is shaded from blue to yellow (see key). The arrows represent SNPs summarised in Fig. 4.19.

The non-synonymous change caused by SNP 4123 (D305E) was analysed further, testing for possible functional changes to the Dmp53 protein. The Phyre2 algorithm for protein prediction suggested that the amino acid replacement occurs in a region of high confidence of intrinsic disorder (Fig. 4.22). Intrinsic disorder (ID) refers to the stability of 3D conformation of proteins, which may be dependent on local conditions, and evidence suggests that these regions have functional importance (Xie et al., 2007; Vucetic et al., 2007; Xie et al., 2007; Tompa, 2005; Dunker et al., 2002).
The Dmp53 protein is a functional and structural homolog of hp53 (Ollmann et al., 2000), and as the functional properties of hp53 have been well studied, an alignment between long isoform of Dmp53 and hp53 was carried out for comparison. SNP 4123 results in an amino acid change in Dmp53, and maps to amino acid position 314 in hp53. The amino acid position 314 in hp53 lies in the CARM1 interaction domain (An et al., 2004), and forms part of the bipartite nuclear localization signal motif (O'Keefe et al., 2003), but has not yet been reported in mutagenesis studies. CARM1 (coactivator-associated arginine methyltransferase 1) protein is involved in methylation of arginine and is associated with diseases such as breast cancer (Wang & Roberts, 2014) and spina bifida (Lu et al., 2010). The bipartite nuclear localization signal motif acts as a tag for nuclear import of the protein.

Fig.4.22 3D model of Dmp53
The predicted protein structure of Dmp53 by Phyre2 server annotated with amino acid 305 (the D305E polymorphism). The secondary structure is coloured in blue with regions of high confidence for intrinsic disorder in yellow. Amino acid position 305 (red) lies in a region of intrinsic disorder.
4.4 Discussion

Previous studies have associated UVR-sensitivity with Dmp53 (Jassim et al., 2003; Ortega-Arellano et al., 2013; Rebollar et al., 2006). Natural variation in this locus that may drive evolution of this gene has not been explored yet. Here, sequencing of the Dmp53 gene from natural European populations has revealed variation in both coding and non-coding regions. In this chapter, I have presented evidence that resolves the consequences of this genotypic variation on some UV responses, and in doing so, identified a strong novel association.

The phenotypic variation in viability observed in the congenic population is relatively small, and the majority of the strains scored an average survivorship of 95% independent of sex (Fig.4.9). Males and females have similar distribution of scores, and the TASSEL5.0 software did not detect any association between adult viability and Dmp53 SNPs. Some significant associations were detected, but rejected after applying the permutation method for false discovery (Anderson and Ter Braak, 2003). As the method is fairly stringent, some these associations may be true positives and may merit further investigation.

Adult flies were exposed to UVB conditions when measuring oxidative stress and Dmp53 expression. The level of UVB exposure in the protocols was equivalent to high levels in Europe, which generally lie closest to the equator. Fig.4.7B shows that longevity of natural strains decreases post irradiation for up to 15 days. Dmp53 null mutants are highly sensitive to solar light and hemizygous flies show an intermediate response (Fig.4.7A). This delay in UVB response is similar to the delay observed after irradiation of simulated solar radiation mentioned earlier.

The congenic population has substantial variation underlying oxidative stress when exposed to UVB. The normalised scores for protein carbonyl content have ~7-fold difference in range between congenic strains (Fig.4.11). TASSEL5.0 software detected several significant associations, one of which remained significant after permutation. More importantly, the large range between the median scores (~4-fold) suggests a reasonably large contribution to the phenotypic variance by this locus (Fig.4.14). Furthermore, the associated-SNP (4123) has been annotated as a missense
variant in exon 8 (Fig.4.19, Table.4.4). The difference in the primary structure of Dmp53 between strains carrying variants of SNP 4123, may explain the difference in the UV response. The change in the structure could alter the function of the protein giving rise to differential resistance to UVR-induced oxidative stress.

Even though the single significant association detected in the three screens of the congenic population tightly correlated with oxidative stress, the mechanism behind the UVR-sensitivity may involve one or multiple ROS-associated pathways. For example, UVB-induced ROS triggers production of anti-oxidants and DNA damage repair pathways, both of which have been shown to be regulated by \( p53 \) (Chapter 1). As Dmp53 is a major transcription factor and interacts with many proteins (Guruharsha et al., 2011), variants of the SNP 4123 may differentially induce any such ROS related pathways. Moreover, the \( Gr94a \) gene, a candidate for taste perception in \( Drosophila \) (Clyne et al., 2000), lies within intron 5 of Dmp53 (Fig.4.19). SNP 4123 is an upstream variant of \( Gr94a \) and may affect its transcription, but this gene is unlikely to be involved in an UV response.

As the UVR-induced oxidative stress is a complex trait, the genetic background of the congenic strains may contribute to the phenotype. This contribution includes possible genetic interactions (epistasis) between Dmp53 and other loci. Therefore, a combination of the genetic background and the point mutation (4123) resulted in the strong UV response, and it remains to test SNP 4123 for UVR resistance in other genetic backgrounds.

The neutral theory of molecular evolution maintains that most polymorphic sites are not under selection but are neutral (Kimura, 1984). In this model, neutral loci do not add to the genetic load of populations as fitness is unaffected. The neutrality tests simply estimate the deviation from the expectation under the neutral model (with constant population size). The tests revealed the Dmp53 gene is under different selection pressures, with peaks of positive and negative Tajima’s \( D \) throughout the locus, in both the African and congenic populations. The overall levels of \( D \) were negative for both populations but the tests were not significant. The analysis is complicated by the presence of the \( Gr94a \) gene, in intron 5 (Fig.4.19). In addition, the
region of negative Tajima’s D in intron 5 of the congenic population corresponds to the large LD block shown in Fig.21B.

The UVR-associated SNP 4123 appears to be under selection in the two populations. The sliding window analysis reveals a significant negative peak for SNP 4123 in the African population indicating a recent selective sweep (Hartl, 2000). Negative Tajima D reflects an excess of low frequency polymorphisms, which is often indicates purifying selection. This interpretation is in contradiction with the genotypic variation observed in the population with a relatively high minor allele frequency (MAF =19%). As the association between the SNP and UVR-sensitivity is fairly strong, such a high frequency would not be expected under purifying/negative selection. Conversely, the LD pattern in Fig.21A, shows linkage of SNP 4123 locally and across the entire gene; indicating possible balancing selection (Charlesworth et al., 1997). The alternative interpretation is population expansion, after a bottleneck, which can also lead to excess low allele frequency polymorphisms. On the other hand, in the congenic population, SNP 4123 lies in a region of balancing selection, which is consistent with the high frequency of the minor allele (MAF of 22%). Further evidence for balancing selection can be gathered by creating near-isogenic strains which are heterozygous at this locus, to test for over-dominance.

The 5 intronic SNPs identified through ancestral sequence inference showed some linkage. Even though, LD decays rapidly in Drosophila (Carbone et al., 2006; Long et al., 1998), SNPs 1069, 3940, 4364 and 4388 showed linkage across multiple introns. Generally, Fig.4.20 shows that these SNPs all lie within peaks of negative Tajima’s D, although they are not significant. SNP 1604 is an outlier and does not show any linkage. Furthermore, the site is within a neutral region in the African population and a region under balancing selection (or population contraction) in the congenic population. Although the other four SNPs are linked, it is possible 1604 has the greater influence on the association between the clusters. Given the large distance between the linked SNPs, it is likely any influence on the phenotype requires all four the linked variants. Further evidence needs to be gathered to dissect the association between the SNP 1604 and the linked SNPs.
The non-synonymous SNP 4123 seems to lie within loci under different selection pressures in the two populations, possibly purifying selection in the African population and balancing selection in the congenic population. The higher levels of UVR in Africa may contribute to the different selection pressures at this locus. Indeed, the differential selection between the populations across the gene is likely due to geographical differences in temperature and UVR, which have both been shown to be associated with the gene in human (Chapter 1).

The McDonald-Kreitman test revealed significant positive selection of the p53 gene in both populations. The positive selection suggests divergence in non-silent sites at the Dmp53 locus driven by adaptive evolution (Hartl, 2000). However the HKA test, which has been shown to have more power in resolving positive selection (Zhai et al., 2009), was not significant. One major drawback of the HKA test is that a true ‘neutral’ locus is required. The neutral locus selected for HKA test may have not been under neutrality in both species, and this may explain the lack of power of the test.

The UVR-associated polymorphism 4123, changes the primary structure of the Dmp53 protein (D305E). Although this SNP has been previously identified in the DGRP, the functional consequence of the polymorphism is unknown. In hp53, the amino acid change corresponds to the CARM1 interaction domain and the bipartite nuclear localisation signal motif. A study by Wang et al., found reduced CARM1 protein levels associated with oxidative stress in motor neuron cells (Wang et al., 2013). hp53 and CARM1 interaction may be involved in protection against oxidative stress in neuronal disease.

The Phrye2 prediction of Dmp53 showed amino acid position 305 lies within a large region of ID (Fig.4.22), which have been shown to have functional importance (Xie et al., 2007; Vucetic et al., 2007; Xie et al., 2007; Tompa, 2005; Dunker et al., 2002). Even though, regions of ID have flexibility in conformation, evolutionary studies have shown they exhibit sequence length conservation amongst divergent species (Daughdrill et al., 2007). Some regions of a protein under ID may be stabilised or form particular conformations under certain conditions, and disordered proteins have been shown to have preference for particular structures under native conditions (Paz et al.,
There is also evidence that intrinsically disordered proteins have some residual structure (Zhu et al., 2008). Functional properties of ID in proteins include protein interactions, DNA binding and binding to other molecules (e.g. antibodies) (Xie et al., 2007; Vucetic et al., 2007; Xie et al., 2007; Tompa, 2005). Furthermore, amino acids in disordered regions are often targets for post-translational modification (Xie et al., 2007) altering structure and function, and a large body of evidence implicates ID with proteins that are central in interaction networks (Patil & Nakamura, 2006; Haynes et al., 2006; Dosztányi et al., 2006; Ekman et al., 2006), including a study on hp53 (Oldfield et al., 2008).

As UVR-induced oxidative stress is a complex trait, and the genetic backgrounds of the congenic strains may contribute to the phenotype. This contribution includes possible genetic interactions (epistasis) between Dmp53 and other loci. Therefore, a combination of the genetic background and the SNP (4123) resulted in the strong UV association, and it is possible that SNP 4123 may not contribute to variation in UVR resistance in other genetic backgrounds.

In summary, the putative UVR-associated polymorphisms identified in the Dmp53 gene, warrant further investigation. SNP 4123 is a particularly strong candidate, which could be investigated in functional assays using congenic or transgenic strains (with a single point-mutation). The mode of selection at the site and the mechanisms that drive the evolution of this locus merit further investigation. The spatial distribution of the variants in wild populations can also be explored, testing for possible UVR-associated clines. In East Asian human populations, a polymorphism in p53 (R72P) was shown to follow a latitudinal cline, and driven by winter temperatures (Shi et al., 2009), although a recent study suggested that the spatial distribution is driven by demographic processes rather than selection (Sucheston et al., 2011). Given the resemblance in population history between human and flies (recent population expansion), variation in Dmp53 is likely to be shaped by both selection and demography.
Chapter 5: Increased longevity in *Drosophila* is induced by inhibitor of Bruton’s tyrosine kinase (BTK) and requires p53

5.1 Introduction

Upon solar UV irradiation, various protein kinases, including MAPK, are activated (Bode & Dong, 2003). Another kinase that is being triggered by UVR is Bruton’s tyrosine kinase (BTK) (Kawakami *et al.*, 1998). Recently, an inhibitor of BTK (Ibrutinib) has been demonstrated to be an effective treatment against cancers, including mantle cell lymphoma (Wang *et al.*, 2013) and chronic lymphocytic leukaemia (Byrd *et al.*, 2013) in clinical trials, and has been approved for use by the US FDA (Food and drug administration). In treating chronic lymphocytic leukaemias, the inhibitor has shown to reduce cell survival and induce apoptosis in patients (Brown, 2013; Ponader *et al.*, 2012). The inhibitor is also a promising drug treatment for autoimmune diseases such as rheumatoid arthritis (Chang *et al.*, 2011), reducing inflammation and cartilage destruction (mice models). Overall, inhibitors of BTK represent promising therapeutic drugs for cancer and inflammatory diseases.

Cellular senescence occurs when a cell stops dividing as telomere length decreases to a critical point of chromosomal instability, and this process has shown to be protective against tumourigenesis (Chapter 1). A study in mice by Baker *et al.*, (2011), revealed clearance/removal of senescent cells throughout an organism extends lifespan and healthspan. This study proved, in principal, that senescence can drive lifespan and age-related diseases (Baker *et al.*, 2011). In human cell lines, BTK presence has been shown to be specifically up-regulated in membrane proteins of senescent cells using mass spectrometry. BTK expression was found to be high in these cells, compared to replicating control cells, and further evidence revealed BTK knockdown bypasses senescence. Hence, BTK is a cellular marker for senescence, and its disruption prevents the accumulation of senescent cells. Drugs X and Y were used as inhibitors of BTK to validate this mechanism; X and Y target different sets of proteins but overlap in their mechanism of action (MOA) as inhibitors of BTK. The p53-null EJp53 cell line

---

23M Althubiti and S Macip 2013, Pers. comm., 1 August
carries a tet-regulatable *hp53* expression system, and was used to show that this response is dependent on *hp53*, a protein wholly implicated in ageing-related pathways (Chapter 1.7). The *Drosophila* homolog Btk29A, shares 37% protein identity (79% coverage) with human BTK, and human BTK partially rescues *Btk29A*− phenotype in null *Btk29A* flies (Hamada *et al.*, 2005). In addition, studies of ageing in mammalian models have found that postmitotic cells can exhibit senescence-related characteristics (Jurk *et al.*, 2012; Sedelnikova *et al.*, 2004).

*C. elegans*, another model used in ageing research, has recently been exploited to screen compounds that affect longevity and oxidative stress. The study identified 60 lifespan–extending compounds, of which 33 also increased resistance to oxidative stress. One of the compounds LFM-A13, an inhibitor of BTK, increased longevity by 27% and reduced oxidative stress by 60% relative to DMSO-control condition (Ye *et al.*, 2014). Experiments in human cell lines mentioned earlier, together with these data, suggested that BTK inhibition using drugs X and Y represents a plausible anti-ageing therapy that could be tested in *D. melanogaster*.

Ageing is the accumulation of changes over time characterised by progressive dysfunction of biological processes and increased susceptibility to pathologies. In recent years, much of the research in ageing has focussed on the discovery of cellular pathways and their genetic basis. The physiological changes associated with ageing have been categorised into nine tentative cellular hallmarks (Ch 1.7, p.29). Although these indicators of ageing were identified with respect to mammals, many of the cellular pathways associated with ageing (Ch 1.7, p.30) are evolutionary conserved (Kenyon, 2010), and may be applied to other organisms such as insects.

Anti-ageing therapies have so far been categorised into behavioural and pharmacological interventions. Behavioural interventions such as dietary restriction have been proven to effective in many models (Fontana *et al.*, 2010) including *D. melanogaster* (Min *et al.*, 2007). However, pharmacological interventions have been less effective and reproducible across different animal models (de Cabo *et al.*, 2014). As well as the obvious benefits of extended lifespan, some interventions are primary
prescribed to increase healthspan and treat disease (Mercken et al., 2012) e.g. a drug
treatment for diabetics which also extends healthy lifespan (Berstein, 2012).

Recently de Cabo et al., (2014) reviewed the current anti-ageing interventions
using strict criteria. The authors singled out molecular compounds that extend lifespan
in at least three models, as well as being independently verified in multiple
laboratories. Only a select few met these criteria: Rapamycin, Spermidine, Metformin
and Resveratrol. These potent anti-ageing compounds are known to cause adverse
effects and the effect of long term exposure is unknown. Rapamycin, an antifungal
agent, targets the TOR kinase and regulates ageing through anabolic and growth rate
pathways (Ch1.7). The compound extends lifespan in mice, fruit flies, worms and yeast
(Bjedov & Partridge, 2011; Kapahi et al., 2004). However, the compound has not been
used as an ageing intervention because of its immunosuppressive properties, and
additional negative effects of anaemia and pneumonitis. Resveratrol has also proven
to extend lifespan in multiple models (Bauer et al., 2004; Baur & Sinclair, 2006; Howitz
et al., 2003). The compound has many health benefits such as protection insulin
resistance which reduces risk of cardiovascular dysfunction (Baur et al., 2006; Lagouge
et al., 2006). As an immunosuppressant, Resveratrol has been tested in multiple
clinical trials (Timmers et al., 2012) which found the compound can cause headaches
and dizziness at higher concentrations (150mg/six times a day) (Almeida et al., 2009).
Spermidine, a naturally occurring compound in humans, decreases in cellular
concentration with age. A study comparing human nonagenarians/centenarians to
quadragenarians and septuagenarians, found higher levels of spermidine in long living
individuals (Pucciarelli et al., 2012). Metformin, a drug treatment for type 2-diabetes,
effects energy balance and inhibits complex I of the electron transport chain (El-Mir et
al., 2000). This inhibition simulates oxidative stress, up regulating antioxidant
pathways and extending healthy lifespan (Berstein, 2012). The direct target of
Metformin is unknown, but mice studies have shown expression profiles of animals
under Metformin treatment, overlap with animals under dietary restriction (Anisimov
et al., 2011; Martin-Montalvo et al., 2013).

*Drosophila* is an excellent model organism for many studying complex biological
processes and many powerful genetic approaches are available (Ch 1.8, p.32). Fruit
flies have been successfully used to study ageing (Bauer et al., 2004; Ford et al., 2007; Shen et al., 2009; Spindler et al., 2012), and high stringency protocols have been developed (Linford et al., 2013). Some of the advantages of using fruit flies, as well as the general properties of the model organism outlined in chapter 1.8 include: (i) relatively short lifespan of two or three months, (ii) the use of mature/fully developed adults and (iii) the availability of previous ageing data. Mature adults of Drosophila species are generally postmitotic (Wolf & Austad, 2010) as only a few cell types divide, for example gut and gonads (Bozcuk, 1972; Ito & Hotta, 1992), and this allows for investigation of post mitotic ageing. However, experimental studies of life extending compounds, including those mentioned earlier have shown these compounds are not only effective in mammals, (Baur & Sinclair, 2006; Harrison et al., 2009; Martin-Montalvo et al., 2013; Soda et al., 2009) which generally have a lower ratio of postmitotic cell types, but are also effective in fruit flies (Bauer et al., 2004; Eisenberg et al., 2009; Kapahi et al., 2004), with the exception of Spermidine (Slack et al., 2012). In addition, fruit flies have been a successful model for anti-ageing drug discovery studies (Pandey & Nichols, 2011; Spindler et al., 2012).

In this chapter, I present a study of two drugs which have been identified in ageing experiments carried out by colleagues24 using human cell lines (human dermal fibroblast: HDF), cancer bladder cell line (EJp53), cancer prostate cell line (Pc3p21). As patents are currently pending for these drugs, they will be referred to as drugs ‘X’ and ‘Y’. The drugs are able to potentiate increased longevity in cells by inhibiting BTK.

---

24 M Althubiti and S Macip 2013, Pers. comm., 1 August
5.2 Methods

5.2.1 Longevity assay
The flies in this assay were grown as in the embryo viability assay (Ch 3.2.2, p.38) to synchronise the age of the flies and to control for potential effects of parental stocks. Male flies were allowed to mature for one-two days AEL on normal maize food prior to each experiment. For each condition, 10 vials of 10 flies were setup and incubated at 25°C throughout the experiments. The flies were transferred to new food twice a week and the number of viable flies was scored each day. Flies that inadvertently escaped or accidentally lost, were right censored and recorded to inform the analysis; as little anaesthetic as possible was used in the experiments. Combinations of DMSO and drug concentration conditions were used. Survival curves for longevity data were produced in Graphpad Prism 6 software (http://www.graphpad.com/). The statistical analysis was carried out using the log-rank (Mantel-Cox test) and the Gehan-Breslow-Wilcoxon (which gives more weight to earlier events) tests in Graphpad Prism 6. For a multiple curve comparison, the number of pairwise comparisons was considered as a number of individual tests and p-values were adjusted using Bonferroni correction (p-value multiplied by no. of tests).

To assess the temporal effect of the drugs, the flies were assayed under DMSO conditions until day 30. At this time, the drug intervention using 10µM drug was started.

5.2.2 Climbing assay
The motor function of the flies after drug administration was tested using the climbing assay. The same experimental procedure was followed as the in longevity assay (see above), however rather than scoring viability, a climbing assay was carried out prior to transfer of flies to new food. The climbing assay was carried out in an isolated temperature controlled room (25°C).

The first step in the protocol was to transfer flies into a tube made of two empty vials. Tape was used to hold the vials together and a line was drawn 8cm from
one side. The flies were allowed to acclimatise to the new environment for one minute. After this time, the vial was tapped firmly on the bench and a 10 secs countdown started. The climbing was scored as a percentage of flies that touch or pass the line/threshold after 10 secs. The flies were then transferred to conditioned food and kept at 25°C. The assay was carried out on flies under DMSO (control) and drug (treatment) (10µM X) conditions. The difference between the two conditions was tested using the Mann-Whitney test and adjusted for multiple testing (Benjamini & Hochberg, 1995).

5.2.3 Weight of flies after drug administration

On completion of the climbing assay (day 30), the remaining flies were weighed. The wet weight of the flies was measured in groups of five flies (one replicate). Flies under DMSO-control and experimental conditions were weighed and compared using Mann-Whitney test. The weights were graphically represented using a bar chart in Graphpad Prism 6.
5.3 Results

In a preliminary experiment, the longevity of the flies was affected by both DMSO and drug additive (10µM drug X) (see appendix, Fig.7.3.1). The longevity of flies was reduced by DMSO intake (median survival= 23 days) compared to the DMSO-free control (median survival= 26 days); the log-rank test showed that the difference is significant (χ2(1, N=200)=9.826, p=0.0017). Similarly, the longevity of flies under DMSO conditions was reduced compared to the drug treatment (median survival= 31 days) (χ2(1, N=200)=8.27, p=0.004). The drug treatment extended the median survival by ~19% compared to the DMSO-free control phenotype, but the difference between the conditions is not statistically significant (χ2(1, N=200)=0.18, p=0.6714). The preliminary results suggest drug X can extend lifespan when administered to mature adults.

The following experiments were carried out using the methodology described in Chapter 5.2. The comparison of survival curves between flies under the DMSO-control and drug X treatment conditions revealed a concentration-dependent response (Fig.5.1). Concentration of 5µM of drug X additive extended the median lifespan by ~17% compared to DMSO-control, and this difference is statistically significant for both the log-rank test (χ2(1, N=200)=20.06, p<0.0001) and Gehan-Breslow-Wilcoxon test (χ2(1, N=200)=17.60, p<0.0001; note that the Gehan-Breslow-Wilcoxon test gives more weight to earlier events). The pairwise comparison between 20µM drug X and DMSO-control condition shows no difference in median lifespan (35 days), and the curves are not statistically different (χ2(1, N=200)=0.56, p=0.45). The overall comparison using the log-rank test reveals that the curves are statistically different (χ2(2, N=300)=28.89, p<0.0001) after Bonferroni correction (p<0.0001). The effectiveness of the intervention is reduced at higher concentrations of drug X, as 5µM concentration is able to extend lifespan, whereas 20µM and 100µM do not (Fig.7.3.1). The 100µM drug concentration was not tested to completion as no effect was higher concentrations of drug X are toxic or ineffective.
The survival curves for male Canton-S flies (N_replicates=10) under DMSO-control and drug X (5µM and 20µM) conditions. The overall comparison using the log-rank test reveals the curves are statistically different (χ²(2, N=300)=28.89, p<0.0001) after Bonferroni correction (p<0.0001). Flies under DMSO-control and drug X 20µM conditions have similar median survival (35 days) and the curves are not statistically different (χ²(1, N=200)=0.56, p=0.45). Drug X extended median lifespan by ~17% when administered at 5µM concentration. The pairwise comparison between the DMSO-control and the 5µM drug treatment conditions were significant for both log-rank test (χ²(1, N=200)=20.06, p<0.0001) and Gehan-Breslow-Wilcoxon test (χ²(1, N=200)=17.60, p<0.0001).
The survival curves for male Canton-S flies ($N_{\text{replicates}}=10$) under DMSO-control and 10µM drug Y conditions. The drug treatment extends the median lifespan by ~14% compared to the DMSO-control condition (35 days); the curves are statistically different for both the log-rank test ($\chi^2(1, N=200)=9.46$, $p=0.0021$) and the Gehan-Breslow-Wilcoxon test ($\chi^2(1, N=200)=10.45$, $p=0.0012$).

The pairwise comparison of flies under DMSO-control and drug Y treatment conditions revealed a similar response as drug X (5µM). 10µM of drug Y additive extends median lifespan by ~14% compared to the DMSO-control (Fig.5.2) and the curves are statistically different for both the log-rank test ($\chi^2(1, N=200)=9.459$, $p=0.0021$) and the Gehan-Breslow-Wilcoxon test ($\chi^2(1, N=200)=10.45$, $p=0.0012$). The curves have a similar shapes but the drug appears to delay the first death event by 10 days which is considerable in comparison to the lifespan of the flies (DMSO 1- 54 days and Y 10µM- 59 days).
Fig. 5.3 Temporal analysis of drug intervention
The survival curves for male Canton-S flies ($N_{\text{replicates}}=10$) under DMSO-control conditions with 10µM drug X intervention after 30 days (arrow). The median lifespan with intervention was the same as without (35 days); the curves are not statistically different ($\chi^2(1, N=200)= 0.037$, $p=0.85$).

The temporal effect of intervention using drug X (10µM) was tested by administering the drug after 30 days (approx. median lifespan, Fig.5.3). Male Canton-S flies under DMSO-control and delayed intervention conditions, have the same median survival (35 days) and the curves are not statistically different ($\chi^2(1, N=200)= 0.037$, $p=0.85$). But, these results indicate good reproducibility of the experiment under DMSO-control conditions. The drug treatment is not effective when administered at approximately 50% through the prospective lifespan of the flies.
Drug intervention using 5µM drug X was ineffective on Dmp53 null mutant (p53[5A-1-4]) flies (Fig.5.4). The drug treatment extended median lifespan by ~10% compared to DMSO-controls, but the curves were not statically different ($\chi^2(1, N=200)= 0.44, p= 0.51$). Under DMSO-control conditions, Canton-S flies have ~15% higher median survival compared to Dmp53 null mutant flies (figure not shown), and this difference is statistically significant ($\chi^2(1, N=200)=21.95, p<0.0001$; note however that the two strains have different genetic background).

Fig.5.4 Survival curves for drug intervention in Dmp53 null mutant flies
The survival curves for male null Dmp53 (p53[5A-1-4]) flies ($N_{\text{replicates}}=10$) under DMSO-control, 5µM drug X conditions. The drug treatment extends the median lifespan by just over 10% compared to the DMSO-control condition (42 days); the curves are not statically different ($\chi^2(1, N=200)= 0.44, p= 0.51$).

The motor function of Canton-S male flies was measured using the climbing assay (Ch 5.2.2, p.107). The flies under drug treatment showed increased motor function compared to DMSO-control condition from day 4 to 26 day (Fig.5.5).
The statistical analysis for each time point was carried out using the non-parametric Mann-Whitney test adjusted for multiple testing (Table.7.3.1). At day 29, the two conditions do not show significance and this is most likely due to the small number of flies which were able to reach the threshold (zero in some vials). The decrease in overall motor function over 29 days may be due to ageing or DMSO toxicity.

**Fig.5.5 Motor function during drug X intervention**
Climbing assay for 100 male Canton-S flies ($N_{replicates}$=10) under DMSO-control and drug X (10µM) conditions, over 29 days. The drug treatment increases motor function in the flies from day 4 to day 26. The comparisons are statistically different until day 29, when few flies are able to reach the threshold (Table.7.3.1). A general trend of decreasing motor function occurs over time. The statistical comparisons were carried using Mann-Whitney tests, adjusted for multiple testing using the Benjamini and Hochberg (1995) procedure. Key: *** $P \leq 0.001$, ** $P \leq 0.01$, * $P \leq 0.05$, ns $P > 0.05$. 
Clinicians prescribing drug $X$ to patients suffering from cancers/autoimmune diseases suspect drug treatment increases weight\textsuperscript{25}. To test if the drug has this side effect in flies, the wet weight of flies at day 30 of the climbing assay was recorded by averaging randomly selected flies (five flies per replicate) under the two conditions (Fig.5.6). Flies under drug treatment measured higher wet weight (median weight=0.97g) compared to the DMSO-control condition (median weight=0.84g), and this increase is significant using the Mann-Whitney test ($U(7.433)= 3.17$, $p=0.018$).

![Fig.5.6 Wet weight of flies after treatment with drug X](image)

Fig.5.6 Wet weight of flies after treatment with drug $X$

The wet weight of male Canton-$S$ flies from the climbing assay at day 30. The bar chart represents the average weight of five flies per replicate, the number of replicates are shown. The flies under the drug treatment weigh more on average (DMSO median weight=0.84g, Drug $X$ (10µM) median weight=0.97g); and this was statistically different using the Mann-Whitney test ($U(7.433)= 3.17$, $p=0.018$). Key: $^* P \leq 0.05$.

\textsuperscript{25}Martin J. S. Dyer 2013, Pers. comm., July 1
5.4 Discussion

The main driving force behind antiageing research is the contribution to advances in human medical research, such as research into age related diseases. Some drugs for chronic or age related disease have been found to affect longevity when analysed retrospectively (Blagosklonny, 2009), and individuals that are unaffected by such diseases tend to live longer (Hitt et al., 1999). As clinical studies on humans are not appropriate for the majority of novel drug interventions, they are often tested in mammalian models, for example rodents. These models may not be suitable for screening novel or putative drugs when evidence is scarce, as they are relatively expensive and long-lived. Under these circumstances fruit flies represent a powerful alternative and a steppingstone to gather further evidence.

In this chapter, I have shown evidence for two novel anti-ageing drugs, X and Y, being effective at increasing longevity in D. melanogaster. These drugs share a common MOA (in human) as BTK inhibitors and potentiate increased longevity and healthspan in fruit flies. Drug X was tested using 5µM, 10µM (preliminary), 20µM and 100µM doses. The experiment was able to establish a response to both the drug and DMSO. Fig.5.1 clearly shows the toxicity of DMSO, an oxidant, when compared to the control condition. Even though flies are exposed to the same levels of DMSO under intervention, the drug treatment extends median survival above the control (non-DMSO) phenotype.

The higher concentrations, 20µM and 100µM, were not effective in extending lifespan, and appear to be toxic. It is possible that as the concentration of drug X increases, so does the hyper-inhibition of Btk29A and activation of cellular off-targets. The undesired pharmacological effects may be due to hyper-inhibition of Btk29A or by the increased action at unknown sites. In parallel with these results, X is currently a prescription medication and under investigation in many clinical trials, and dose-dependent adverse pharmacological effects have been reported by clinicians\(^{26}\) (e.g. nausea, anaemia and pneumonia); the drug is prescribed to patients at the highest tolerable dosage. In fruit flies, higher concentrations of X did not yield a reduction in

\(^{26}\) Reference names drug and is withheld
median survivorship greater than that caused by DMSO. Thus, the toxicity of drug \( X \) at higher concentrations does not reduce longevity than that caused by DMSO.

There is a possibility that the drugs are interacting with/or alleviating the DMSO responses. Conversely, a drug discovery screen (Pandey & Nichols, 2011) for extended lifespan in *Drosophila*, applied the same concentration of DMSO as used here (2% v/v). This issue can be addressed by performing a longevity assay using DMSO-free controls.

Drug \( X \) at 5µM did not only delay ageing in flies compared to DMSO-control condition, but was effective right from the beginning of the experiment. The effectiveness of drug \( X \) was greater at lower concentrations and the efficacy (most effective dose) most likely lies below 10µM. In addition, the temporal analysis of drug \( X \) (10µM) suggests that early intervention is required to extend lifespan (Fig.5.3). Drug intervention after 30 days did not elicit a response in longevity, and the median lifespan was identical for the two conditions (35 days). The early intervention may be required to activate a specific MOA in order to potentiate the extension in lifespan, or it is possible that an accumulation of the drug over time is critical for the response.

Another (human) BTK inhibitor, drug \( Y \), was tested to validate the MOA of \( X \). The drug was administered as an additive at 10µM concentration. A similar response to 5µM drug \( X \) was observed using 10µM drug \( Y \). Drug \( Y \) delays ageing in flies compared to DMSO-control condition and is similarly effective from the beginning of the experiment.

Although, drugs \( X \) and \( Y \) have multiple cellular targets, they have both been shown to target the human BTK protein\(^{27}\). Both drugs are effective in extending lifespan in male adult flies (Fig.5.1 and Fig.5.2) and the MOA maybe involves clearance of senescent cells. Even though adult fruit flies are mostly postmitotic, the effectiveness of drugs \( X \) and \( Y \) is not surprising. Previously discovered potent antiageing compounds have been found to be effective in multiple models including *Drosophila* (Eisenberg *et al.*, 2009; Kapahi *et al.*, 2004). Furthermore, studies of ageing in mammalian models have found postmitotic cells can exhibit senescence-related

\(^{27}\) Reference names drug and is withheld
characteristics (Jurk et al., 2012; Sedelnikova et al., 2004). As yet, there have been no studies to investigate possible cellular off-targets of the drugs in *Drosophila* species.

A daily dose of 10µM drug *X* had a positive effect on the mobility of male flies, tested by using the climbing assay. Fig.5.5 shows that male flies have consistently better motor function under drug treatment compared to DMSO-control conditions. It is possible that oxidative stress caused by the DMSO additive may negatively affect mobility, however a study inducing oxidative stress (using paraquat) in male flies, did not detect a response in climbing (Wang et al., 2011). The increase in mobility under drug treatment conditions is statistically significant for all time points except at 29 days. At this time point, the majority of flies in both conditions do not show a climbing response. The difference between both conditions is consistent throughout the experiment but a general trend of decreasing response is observed. The decrease in mobility observed in both conditions appears to correlate with ageing. However, this may be a consequence of the accumulation of oxidative stress over the course of the experiment. Performing the assay with a DMSO-free control flies may help elucidate this trend. In clinical trials drug *X* has shown to have little adverse effects, but no positive effect on mobility has been observed, in fact, common symptoms include fatigue and musculoskeletal pain. One possible side effect noted by clinicians is weight gain28, but these claims have not been investigated. An increase in weight was observed in fruit flies after 30 days of drug treatment (Fig.5.6). The possible mechanism or basis of this phenotype is untested.

In human cell lines, BTK inhibitors seem to prevent the induction of hp53, a master regulator of senescence29. Drug *X* intervention using 5µM does not potentiate an increase in longevity in *Dmp53* null mutant flies, suggesting that the effects of BTK inhibition on longevity, observed in flies, is mediated by the p53 pathway. *Dmp53* protein has been associated with ageing-related pathways present in humans and flies (Chapter 1). As drug *X* is effective in *Canton-S* flies, it is possible *Dmp53* modulates or is required for a response. *Canton-S* and *Dmp53* null mutant flies do not share a common genetic background, and this hinders interpretation of these results. In addition, the

---

28 Martin J. S. Dyer 2013, Pers. comm., July 1
29 M Althubiti and S Macip 2013, Pers. comm., 1 August
two strains respond differently under DMSO-control conditions. The difference may be due to the homozygous knockout of \textit{Dmp53} gene, the genetic background or a combination of the two. In chapter 4, I detected association between \textit{Dmp53} gene and oxidative stress. It is therefore unsurprising that the two genotypes have different sensitivities to oxidative stress.

In summary, drugs X and Y are effective anti-ageing treatments in fruit flies. The adverse effects caused by the drugs are dose-dependent and the exact MOA of the drugs needs to be investigated further using biochemical assays. The life-extending capacity of drug X in flies requires Dmp53, a protein implicated in many ageing-related pathways (Chapter 1). Drug X also increased mobility and caused weight gain in flies, when compared to DMSO-controls, but the mechanism underlying this response is yet to be determined.
Chapter 6: General discussion

UVR has been a major environmental factor throughout the evolution of life. The history of life on earth suggests UVR-sensitivity has evolved over many billions of years. The first aim of this project was to identify the extent of natural genetic variation underlying UVR-sensitivity in wild populations of *Drosophila*. For this purpose, a panel of *Drosophila* strains (DGRP) derived from a single wild population (Raleigh, N. Carolina) were used to perform a GWAS. The considerable phenotypic variation observed in the DGRP, together with the genotypic/marker data available was exploited in the analysis. The study detected 114 significant associations with equal and rather small contribution to the phenotype, suggesting a highly complex genetic architecture for UVR-sensitivity in flies. These results mirror previous radiation-based studies, a QTL analysis of UVC-resistance (Gomez *et al.*, 2013) and a GWAS into effects of ionising radiation (Vaisnav *et al.*, 2014).

The comparisons of some of the most significant associations (SNPs) revealed that they have a small contribution to the phenotype (~5%, Figure.3.11). These small differences in the phenotype are expected when studying a naturally derived population. Some of the significant associations were analysed by complementation testing and a single SNP showed quantitative lack of complementation (SNP in gene CG42342). The gene is not associated with UVR-sensitivity or found in any UVR-related pathway. The CG42342 protein contains some collagen-like domain domains which could possibly be involved in screening UVR, and merits further studying.

Ideally, the validation of significant associations should be carried out using congenic strains which differ at the single SNP. The method to create congenic strains however takes ample time and is not suitable to screen multiple SNPs for validation. As 100+ SNPs strongly associated at the FDR 0.0001 level, with small contributions to the phenotype, it is difficult to speculate which candidate SNPs should be selected for further validation using congenic strains. Furthermore, a number of the candidates are novel or not currently associated with UVR-related pathways.
The UVR-associated polymorphisms that were detected in this study (chapters 3 and 4), would be interesting candidates for testing for latitudinal clines or any other geographical patterns in allele frequencies. Identifying loci/SNPs tightly correlated with UVR distributions will further implicate those loci to UVR-sensitivity. The strongest association detected in my experiments was SNP 4123 (3R:18875989), which would make the most prudent choice for population genotyping. Several studies have identified allelic clines in fruit flies e.g. temperature-resistance (McColl & McKechnie, 1999), biological rhythms (Costa et al., 1992), and diet (Fry et al., 2008). However, a weakness in some of these studies is the assumption of a linear correlation between geographical distances and environmental conditions, i.e. clines relating to longitude and latitude. Instead, correlations can be tested (regression analysis) between allele frequencies of UVR-associated loci and precise UVR data. Vast amounts of global UV irradiance data\(^{30}\) has been collected by a number of instruments including TOMS (Total Ozone Mapping Spectrometer) and OMI (Ozone Monitoring Instrument) which are collated in the NASA GES DISC repository (available at http://disc.sci.gsfc.nasa.gov). At first glance, the data suggest large differences (~2-fold) in averaged annual UV irradiance recordings (Erythemal\(^{31}\) Daily Dose) throughout Europe, with a latitudinal gradient for Erythemal\(^{31}\) UV levels. The correlation between the UV irradiance data and allele frequency of associated loci can be tested using regression analysis. Population samples of *D. melanogaster* derived from wild European populations are available for genotyping, and complete genome sequence data is available from nine French strains (DPGP) and 100+ Portuguese strains (frequency only- population data). The best method for mass genotyping of strains would be to use an automated PCR-based allelic discrimination protocol. Some commercial kits such as TaqMan® Genotyping Master Mix (life technologies) are available for this purpose (albeit expensive). However, as the fly strains are all homozygous, a cheaper qPCR-based method could be developed.

---

\(^{30}\) Global UV irradiance data- data including UVI, Erythemal UV and clear-sky Erythemal UV

\(^{31}\) Erythemal UV- Measure of potential biological damage caused by sunlight based on the erythemal action spectrum
Evolutionary drivers of variation in UVR genes

Shi et al., (2009) offers one mechanism by which SNPs that are expected to be under purifying selection, may be maintained through a ‘trade-off’ phenomenon (Ch 3.4, p.61). Their study focused on polymorphism in two genes driven by selection, where an allele of one of these genes, MDM2, is under selection pressure from UVR. The other gene, hp53, is under strong selection from cold (winter)-temperatures, but the ‘trade-off’ phenomenon maintains variation at both loci.

Another plausible mechanism for the phenotypic variation observed is antagonistic pleiotropy (AP) model (Williams, 2001). Interactions between pleotropic genes play a major role in complex traits such as UVR-sensitivity (Mackay, 2001). AP occurs when an underlying mutation is both beneficial and detrimental to different phenotypic traits, and one example of AP has been described in the p53 gene which may suppress cancer with a consequence of increased ageing. As fitness is affected by both processes, the underlying variants are not subject to purifying selection (Ungewitter & Scrable, 2009). In terms of UVR-sensitivity, the AP model can be applied where a mutation affecting different UVR responses has antagonistic effects on fitness. For example, the Dmp53 gene is implicated in UVR-induced DNA-damage and oxidative stress responses (Chapter 1), and a variant which is under AP related to these phenotypes may not be under negative selection. In this example, increasing levels of UVR exposure may exert different selection pressures on the variants resulting in a ‘trade-off’, and preserving the genetic variation; the model may be maintained with only one UVR-associated phenotype as long as the effects on the phenotypes involved are antagonistic for fitness. If the pleotropic effects are effects occur in the same direction, genetic correlation between the phenotypes is expected (Falconer & Mackay, 1996). In addition, it is possible for loci under this mechanism to be detected by a single-marker analysis performed in chapters 3 and 4, as the experimental setup may bias a particular pleiotropic phenotype over another, when compared to natural UV conditions i.e. abolishing the pleiotropic effect of a variant in the GWAS which is usually pleiotropic under natural conditions. There could also be epistasis at such loci which could modify the effect of all or only some of the phenotypes involved.
In the context of climate change, the evolutionary past is important as it may inform the future as many fundamental cellular mechanisms associated with UVR-sensitivity are evolutionary conserved. For example, DNA repair mechanisms were most likely more efficient in the past when UVR levels were higher. As there is a cost to DNA repair, the range of selection pressures found amongst and between ecosystems has led to a diversity of efficiencies. Given the cost of many UVR-sensitivity pathways, UVR-resistance is likely constrained to the levels that are required by selection pressures in different environments, and variation may be observed between populations with small differences in UVR (e.g. congenic strains derived from Italian populations Rende and Rutigliano, see Chapter 4). The natural variation maintained by selection pressures on loci associated with such mechanisms may contain important genetic information about UVR-resistance in the past, demonstrating the response to past higher levels of UVR (Chapter 1). Allele reconstruction (as described in Ch 4.2.6, p.82) of important UVR-associated genes from ancient DNA could lead to the discovery of new phenotypes related to the high levels of UVR experienced by the organisms of that time. The technique has already been used to dissect to the timing of pigmentation lightening in Europeans as a response to reduced UVR (Beleza et al., 2012). Furthermore, in a bacterial study by Gaucher et al., (2003), an ancient phenotype was ‘resurrected’ in the laboratory by creating transgenic organisms carrying alleles inferred through ancestral reconstruction (Gaucher et al., 2003). This approach may yield important genetic information about the ability of ancient organisms to cope with high UVR levels.

The accumulation of knowledge about the nature and evolution of UVR-sensitivity, together with the increased understanding of future climates and global distributions of UVR, may help direct conservation efforts of wild populations. The recent advances in climate change research have promoted new investigations into wildlife conservation (Tingley et al., 2013). One emerging and sometimes controversial area is genetic conservation. The main aim of genetic conservation is the preservation of biodiversity and protection of species from extinction; it has been reported that 15-40% of species will be extinct by 2050 though global warming alone (Thomas & Williamson, 2012). The controversy of genetic conservation centres on the idea of
genetic engineering of wild populations. In the last three decades, genetic engineering research has been focussed on genetically modified (GM) crops, which are planted in 12% of arable land worldwide (ISAAA Brief 46-2013). A recent review by Redford et al., (2013), outlined the increasing interest in GM organisms by conservation biologists in order to facilitate adaption (artificial adaption). There are three methods of facilitated adaption; firstly selected breeding using better-adapted populations, secondly the use of transgenics within the same species, and thirdly the use of cross species transgenics. In respect to a complex trait such as UVR-sensitivity, the latter two methods are not likely to be useful as many genes make a contribution to the phenotype. An exception is the CPD photolyase gene which was found to have natural variants in rice conferring large differences in UVR-resistance (Hidema et al., 2007). The best approach for conservation of UVR-sensitive species is to utilise breeding across populations (increased variation). This type of intervention has been successful in increasing heterogeneity in vulnerable populations (Johnson et al., 2010; Madsen et al., 1996), but there are many potential risks including the spread of diseases and dilution of unforeseen beneficial variants in the donor population. Currently there is no comprehensive database of UVR-sensitive populations and it is most likely these populations will be discovered when species become endangered. In addition, extinction is no longer permanent (Folch et al., 2009) and many species may be resurrected in the future. With this in mind, an increasing number of conservation efforts are focusing on storing samples of animal tissues in frozen arches, in the hope of resurrecting organisms that may become extinct as well as those that already are (e.g. The Long Now Foundation- http://rare.longnow.org).

Exploring non-additive effects

The DGRP data could also be further analysed to explain the genetic variation underlying UVR-associated loci using non-additive effect models (for example epistasis, pleiotropy, and phenotypic plasticity). Epistasis has been shown to contribute to variation in complex traits in several model species (Koller et al., 2008; Kroymann & Mitchell-Olds, 2005), and using gene-gene interaction models may reveal UVR-sensitivity associations that are not detectable using single-marker analysis. These methods are computationally expensive but several efficient tools are available for
large genome-wide datasets (Gyenesei et al., 2012; Wan et al., 2013). In particular, fastANOVA and COE programs which are appropriate for GWAS using homozygous genotypes (Zhang et al., 2011), may yield interesting results. On the other hand, pleiotropy in UVR-sensitivity cannot be analysed as only one phenotype was measured, and plasticity cannot be analysed as the flies were only exposed to one experimental condition.

Similarly, the association analysis of the p53 gene (Chapter 5) was carried using an additive model, but epistasis may be present at this locus, and that the manifestation of the phenotype may be context-dependant. The genetic background may influence the contribution of this site to the phenotype. In some cases, these epistatic effects can be extreme and ‘toggle’ fitness\(^{32}\), examples have been identified in worms, fruit flies and mice (Burnett et al., 2011; Naya et al., 2002). Therefore, it is possible the UVR-sensitivity-associated SNP (4123) identified in the TASSEL analysis may not confer the same association in a different background. The congenic strains were created to mitigate this issue between gene-gene interactions by focussing on the variation of one gene. However epistasis can also occur between sites within genes and the effects could be tested for Dmp53. The epistasis tests would not be computationally expensive and could be carried out exhaustively. Testing the epistasis interaction between Dmp53 SNPs could reveal possible intragenic interactions, and shed further light on the Dmp53 protein evolution (Breen et al., 2012).

---

\(^{32}\)‘Toggle’ fitness- act as a binary fashion on fitness from one extreme to another.
Chapter 7: Appendices

7.1 Appendix for Chapter 4

7.1.1 Reproducibility of DGRP strains in embryo viability assay

Table 7.1.1 P-values for t-tests comparing independent embryo viability experiments

<table>
<thead>
<tr>
<th>Exp. 2</th>
<th>Strain</th>
<th>25174 Control</th>
<th>25174 Irradiated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. 1</td>
<td>25174</td>
<td>0.4476 (ns)</td>
<td>&lt;0.0001 ****</td>
</tr>
<tr>
<td></td>
<td>25174</td>
<td>&lt;0.0001 ****</td>
<td>0.4252 (ns)</td>
</tr>
</tbody>
</table>

Exp. 2

<table>
<thead>
<tr>
<th>Strain</th>
<th>Condition</th>
<th>Canton S Control</th>
<th>Canton S Irradiated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. 1</td>
<td>Canton S</td>
<td>0.4602 (ns)</td>
<td>0.0004 ***</td>
</tr>
<tr>
<td></td>
<td>Canton S</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7.1.2 Lack of complementation assay

\[
\begin{array}{c}
+; +; -; + \\
+ + + + \\
\end{array}
\xrightarrow{X}
\begin{array}{c}
W; -; \frac{Df}{+} ; + \\
+ \frac{Bal}{+} ; + \\
\end{array}
\]

DGRP Strain

Deficiency Strain

\[
\begin{array}{c}
W/+; +; +; + \\
+ \frac{Df}{+} ; + \\
\end{array}
\]

DGRP chr with Df deletion

\[
\begin{array}{c}
W/+; +; +; + \\
+ \frac{Bal}{+} ; + \\
\end{array}
\]

DGRP chr with Bal chr

Fig. 7.1.1 Crosses for complementation tests
### Table 7.1.2 DGRP strains used in the embryo viability assay for the GWAS

<table>
<thead>
<tr>
<th>Stock no.</th>
<th>Strain ID</th>
<th>normalised Score</th>
<th>Stock no.</th>
<th>Strain ID</th>
<th>normalised Score</th>
<th>Stock no.</th>
<th>Strain ID</th>
<th>normalised Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>25185</td>
<td>RAL_358</td>
<td>29.73</td>
<td>28139</td>
<td>RAL_105</td>
<td>37.78</td>
<td>28234</td>
<td>RAL_801</td>
<td>42.14</td>
</tr>
<tr>
<td>25189</td>
<td>RAL_379</td>
<td>31.12</td>
<td>25193</td>
<td>RAL_427</td>
<td>38.01</td>
<td>28176</td>
<td>RAL_646</td>
<td>42.18</td>
</tr>
<tr>
<td>25198</td>
<td>RAL_555</td>
<td>31.50</td>
<td>28200</td>
<td>RAL_461</td>
<td>38.05</td>
<td>28154</td>
<td>RAL_217</td>
<td>42.20</td>
</tr>
<tr>
<td>28229</td>
<td>RAL_776</td>
<td>32.35</td>
<td>28179</td>
<td>RAL_359</td>
<td>38.42</td>
<td>25202</td>
<td>RAL_730</td>
<td>42.28</td>
</tr>
<tr>
<td>28232</td>
<td>RAL_790</td>
<td>34.21</td>
<td>28188</td>
<td>RAL_381</td>
<td>38.58</td>
<td>28231</td>
<td>RAL_787</td>
<td>42.33</td>
</tr>
<tr>
<td>28173</td>
<td>RAL_338</td>
<td>34.22</td>
<td>28143</td>
<td>RAL_138</td>
<td>38.72</td>
<td>28125</td>
<td>RAL_38</td>
<td>42.41</td>
</tr>
<tr>
<td>28183</td>
<td>RAL_371</td>
<td>34.39</td>
<td>28136</td>
<td>RAL_91</td>
<td>38.83</td>
<td>25209</td>
<td>RAL_852</td>
<td>42.45</td>
</tr>
<tr>
<td>25174</td>
<td>RAL_208</td>
<td>34.90</td>
<td>28140</td>
<td>RAL_109</td>
<td>38.86</td>
<td>28202</td>
<td>RAL_491</td>
<td>42.54</td>
</tr>
<tr>
<td>28244</td>
<td>RAL_822</td>
<td>34.96</td>
<td>28130</td>
<td>RAL_69</td>
<td>38.87</td>
<td>28240</td>
<td>RAL_812</td>
<td>42.64</td>
</tr>
<tr>
<td>28196</td>
<td>RAL_426</td>
<td>35.00</td>
<td>28145</td>
<td>RAL_149</td>
<td>38.87</td>
<td>28207</td>
<td>RAL_531</td>
<td>42.70</td>
</tr>
<tr>
<td>28238</td>
<td>RAL_808</td>
<td>35.41</td>
<td>28177</td>
<td>RAL_352</td>
<td>39.33</td>
<td>28182</td>
<td>RAL_370</td>
<td>42.94</td>
</tr>
<tr>
<td>25207</td>
<td>RAL_799</td>
<td>35.43</td>
<td>28166</td>
<td>RAL_309</td>
<td>39.46</td>
<td>28223</td>
<td>RAL_738</td>
<td>43.06</td>
</tr>
<tr>
<td>28208</td>
<td>RAL_535</td>
<td>35.57</td>
<td>28144</td>
<td>RAL_142</td>
<td>39.47</td>
<td>28176</td>
<td>RAL_350</td>
<td>43.10</td>
</tr>
<tr>
<td>28211</td>
<td>RAL_563</td>
<td>35.75</td>
<td>25194</td>
<td>RAL_437</td>
<td>39.91</td>
<td>28220</td>
<td>RAL_721</td>
<td>43.65</td>
</tr>
<tr>
<td>28194</td>
<td>RAL_392</td>
<td>35.77</td>
<td>28171</td>
<td>RAL_332</td>
<td>40.01</td>
<td>28170</td>
<td>RAL_325</td>
<td>43.82</td>
</tr>
<tr>
<td>28187</td>
<td>RAL_378</td>
<td>35.91</td>
<td>28239</td>
<td>RAL_810</td>
<td>40.07</td>
<td>28138</td>
<td>RAL_101</td>
<td>43.95</td>
</tr>
<tr>
<td>28216</td>
<td>RAL_642</td>
<td>36.06</td>
<td>28195</td>
<td>RAL_398</td>
<td>40.19</td>
<td>25201</td>
<td>RAL_712</td>
<td>44.01</td>
</tr>
<tr>
<td>28132</td>
<td>RAL_75</td>
<td>36.13</td>
<td>28209</td>
<td>RAL_554</td>
<td>40.20</td>
<td>25200</td>
<td>RAL_707</td>
<td>44.03</td>
</tr>
<tr>
<td>28142</td>
<td>RAL_136</td>
<td>36.17</td>
<td>25210</td>
<td>RAL_859</td>
<td>40.31</td>
<td>28241</td>
<td>RAL_818</td>
<td>44.14</td>
</tr>
<tr>
<td>28161</td>
<td>RAL_239</td>
<td>36.57</td>
<td>25197</td>
<td>RAL_517</td>
<td>40.39</td>
<td>28213</td>
<td>RAL_589</td>
<td>44.39</td>
</tr>
<tr>
<td>25190</td>
<td>RAL_380</td>
<td>36.58</td>
<td>28186</td>
<td>RAL_377</td>
<td>40.41</td>
<td>25199</td>
<td>RAL_639</td>
<td>44.66</td>
</tr>
<tr>
<td>28198</td>
<td>RAL_441</td>
<td>36.87</td>
<td>28178</td>
<td>RAL_356</td>
<td>40.45</td>
<td>28160</td>
<td>RAL_237</td>
<td>44.72</td>
</tr>
<tr>
<td>28197</td>
<td>RAL_440</td>
<td>36.93</td>
<td>25204</td>
<td>RAL_765</td>
<td>40.67</td>
<td>28131</td>
<td>RAL_73</td>
<td>45.00</td>
</tr>
<tr>
<td>28128</td>
<td>RAL_45</td>
<td>36.99</td>
<td>28245</td>
<td>RAL_832</td>
<td>40.75</td>
<td>28148</td>
<td>RAL_161</td>
<td>45.01</td>
</tr>
<tr>
<td>25744</td>
<td>RAL_705</td>
<td>37.01</td>
<td>28150</td>
<td>RAL_177</td>
<td>40.76</td>
<td>28190</td>
<td>RAL_383</td>
<td>45.18</td>
</tr>
<tr>
<td>28237</td>
<td>RAL_805</td>
<td>37.05</td>
<td>25180</td>
<td>RAL_313</td>
<td>40.95</td>
<td>28157</td>
<td>RAL_228</td>
<td>45.76</td>
</tr>
<tr>
<td>28164</td>
<td>RAL_280</td>
<td>37.19</td>
<td>28228</td>
<td>No seq</td>
<td>41.02</td>
<td>28236</td>
<td>RAL_804</td>
<td>46.31</td>
</tr>
<tr>
<td>28230</td>
<td>RAL_783</td>
<td>37.21</td>
<td>25445</td>
<td>RAL_365</td>
<td>41.10</td>
<td>25184</td>
<td>RAL_357</td>
<td>46.38</td>
</tr>
<tr>
<td>28163</td>
<td>RAL_272</td>
<td>37.36</td>
<td>28153</td>
<td>RAL_195</td>
<td>41.31</td>
<td>25191</td>
<td>RAL_391</td>
<td>46.99</td>
</tr>
<tr>
<td>28246</td>
<td>RAL_837</td>
<td>37.53</td>
<td>28134</td>
<td>RAL_83</td>
<td>41.32</td>
<td>25208</td>
<td>RAL_820</td>
<td>47.04</td>
</tr>
<tr>
<td>28215</td>
<td>RAL_595</td>
<td>37.59</td>
<td>28146</td>
<td>RAL_153</td>
<td>41.33</td>
<td>28124</td>
<td>RAL_28</td>
<td>47.41</td>
</tr>
<tr>
<td>28126</td>
<td>RAL_41</td>
<td>37.60</td>
<td>25188</td>
<td>RAL_375</td>
<td>41.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28235</td>
<td>RAL_802</td>
<td>37.63</td>
<td>28165</td>
<td>RAL_287</td>
<td>41.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28167</td>
<td>RAL_317</td>
<td>37.65</td>
<td>28151</td>
<td>RAL_181</td>
<td>41.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25745</td>
<td>RAL_714</td>
<td>37.69</td>
<td>28184</td>
<td>RAL_373</td>
<td>41.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28156</td>
<td>RAL_227</td>
<td>37.70</td>
<td>25205</td>
<td>RAL_774</td>
<td>41.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28141</td>
<td>RAL_129</td>
<td>37.74</td>
<td>25203</td>
<td>RAL_732</td>
<td>42.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>DGRP strain</td>
<td>Allele</td>
<td>Deficiency strain</td>
<td>Dominant phenotypic marker on Balancer chromosome</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>--------</td>
<td>-------------------</td>
<td>-----------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin</td>
<td>28124</td>
<td>A</td>
<td>9701</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin</td>
<td>25208</td>
<td>A</td>
<td>9701</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin</td>
<td>25191</td>
<td>A</td>
<td>9701</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin</td>
<td>25184</td>
<td>A</td>
<td>9701</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin</td>
<td>25185</td>
<td>G</td>
<td>9701</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin</td>
<td>25198</td>
<td>G</td>
<td>9701</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin</td>
<td>28229</td>
<td>G</td>
<td>9701</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin</td>
<td>28232</td>
<td>G</td>
<td>9701</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4d8</td>
<td>28124</td>
<td>C</td>
<td>7929</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4d8</td>
<td>25208</td>
<td>C</td>
<td>7929</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4d8</td>
<td>25191</td>
<td>C</td>
<td>7929</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4d8</td>
<td>25184</td>
<td>C</td>
<td>7929</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4d8</td>
<td>25189</td>
<td>T</td>
<td>7929</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4d8</td>
<td>25198</td>
<td>T</td>
<td>7929</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4d8</td>
<td>28173</td>
<td>T</td>
<td>7929</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp4d8</td>
<td>28183</td>
<td>T</td>
<td>7929</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vn</td>
<td>25189</td>
<td>C</td>
<td>24941</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vn</td>
<td>25198</td>
<td>C</td>
<td>24941</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vn</td>
<td>28229</td>
<td>C</td>
<td>24941</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vn</td>
<td>28232</td>
<td>C</td>
<td>24941</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vn</td>
<td>28124</td>
<td>A</td>
<td>24941</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vn</td>
<td>25191</td>
<td>A</td>
<td>24941</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vn</td>
<td>25184</td>
<td>A</td>
<td>24941</td>
<td>Short bristles on back of adults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED14</td>
<td>25208</td>
<td>C</td>
<td>7564</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED14</td>
<td>28236</td>
<td>C</td>
<td>7564</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED14</td>
<td>28157</td>
<td>C</td>
<td>7564</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED14</td>
<td>28190</td>
<td>C</td>
<td>7564</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED14</td>
<td>25185</td>
<td>T</td>
<td>7564</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED14</td>
<td>25189</td>
<td>T</td>
<td>7564</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED14</td>
<td>28232</td>
<td>T</td>
<td>7564</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MED14</td>
<td>28173</td>
<td>T</td>
<td>7564</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG42342</td>
<td>25185</td>
<td>C</td>
<td>7737</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG42342</td>
<td>25189</td>
<td>C</td>
<td>7737</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG42342</td>
<td>28229</td>
<td>C</td>
<td>7737</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG42342</td>
<td>28173</td>
<td>C</td>
<td>7737</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG42342</td>
<td>25191</td>
<td>G</td>
<td>7737</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG42342</td>
<td>25184</td>
<td>G</td>
<td>7737</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG42342</td>
<td>28190</td>
<td>G</td>
<td>7737</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG42342</td>
<td>28220</td>
<td>G</td>
<td>7737</td>
<td>Short fat pupa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.1.3 Crosses for quantitative complementation tests
Fig. 7.1.2 Complementation tests of significant SNPs
The complementation is represented as interaction plots. The interaction is between the strain (DGRP strain allele, A) and genotype (Bal/Df, G), A x G. In these complementation tests none of the SNPs showed a failure to complement.
7.1.2 GWAS in UVR-sensitivity

Fig.7.1.3 Variation in survivorship of embryos for DGRP strain under control and experimental conditions

105 DGRP strains have been assayed for embryo viability under control (A) and experimental (B) conditions (using the Suntest Solar simulator (Ch 3.2.2, p.38)). The strains have been arranged from lowest to highest median score (black dots); the strains of the two conditions are not in the same order. The lines represent the range from minimum to maximum values of four replicates (N=~1000 embryos). A. Under control conditions, the graph shows that there is little variation in median survivorship between the strains (F(104,314)=1.634, p<0.0006), ranging from 67% to 74%. B. Under experimental conditions, the graph shows that there is large variation in median survivorship between the strains (F(104,315)=15.59, p<0.0001), ranging from 23% to 38%, and this is reflected in the shape of the graph.
Table 7.1.4 Significant associations detected in the GWAS

<table>
<thead>
<tr>
<th>Gene</th>
<th>W values</th>
<th>Gene</th>
<th>W values</th>
<th>Gene</th>
<th>W values</th>
</tr>
</thead>
<tbody>
<tr>
<td>TwdlB</td>
<td>14</td>
<td>Fhos</td>
<td>104</td>
<td>Dhc64C</td>
<td>1801</td>
</tr>
<tr>
<td>shep</td>
<td>45</td>
<td>CG43078</td>
<td>105</td>
<td>Dhc64C</td>
<td>1801</td>
</tr>
<tr>
<td>raw</td>
<td>48</td>
<td>Indy</td>
<td>106</td>
<td>Dhc64C</td>
<td>1801</td>
</tr>
<tr>
<td>alpha-Cat</td>
<td>59</td>
<td>Gprk2</td>
<td>106</td>
<td>Dhc64C</td>
<td>1801</td>
</tr>
<tr>
<td>Pka-C1</td>
<td>61</td>
<td>luna</td>
<td>107</td>
<td>Dhc64C</td>
<td>1801</td>
</tr>
<tr>
<td>CG15817</td>
<td>70</td>
<td>bin3</td>
<td>109</td>
<td>Dhc64C</td>
<td>1801</td>
</tr>
<tr>
<td>CG15817</td>
<td>70</td>
<td>mus205</td>
<td>110</td>
<td>l(3)psg2</td>
<td>1802</td>
</tr>
<tr>
<td>SNF4Ay</td>
<td>71</td>
<td>CG18278</td>
<td>112</td>
<td>Fancd2</td>
<td>1803</td>
</tr>
<tr>
<td>CG18747</td>
<td>73</td>
<td>CG15475</td>
<td>114</td>
<td>CG5704</td>
<td>1804</td>
</tr>
<tr>
<td>Argk</td>
<td>75</td>
<td>bab1</td>
<td>114</td>
<td>RpA-70</td>
<td>1805</td>
</tr>
<tr>
<td>CG6024</td>
<td>76</td>
<td>lea</td>
<td>114</td>
<td>Rp6p</td>
<td>1808</td>
</tr>
<tr>
<td>CG42342</td>
<td>77</td>
<td>CG33144</td>
<td>115</td>
<td>Unc-13-4B</td>
<td>1811</td>
</tr>
<tr>
<td>Gyc88E</td>
<td>82</td>
<td>Ten-a</td>
<td>117</td>
<td>mnn</td>
<td>1813</td>
</tr>
<tr>
<td>CG4558</td>
<td>83</td>
<td>nAChRalpha7</td>
<td>117</td>
<td>CG11073</td>
<td>1813</td>
</tr>
<tr>
<td>kst</td>
<td>85</td>
<td>RAF2</td>
<td>117</td>
<td>Ssl1</td>
<td>1817</td>
</tr>
<tr>
<td>CG40045</td>
<td>86</td>
<td>Src64B</td>
<td>117</td>
<td>CG1998</td>
<td>1817</td>
</tr>
<tr>
<td>CG9682</td>
<td>88</td>
<td>Ugt86De</td>
<td>118</td>
<td>CG1998</td>
<td>1817</td>
</tr>
<tr>
<td>crc</td>
<td>88</td>
<td>CG3104</td>
<td>118</td>
<td>hay</td>
<td>1820</td>
</tr>
<tr>
<td>CG34114</td>
<td>89</td>
<td>Fkbp14</td>
<td>118</td>
<td>Neto</td>
<td>1820</td>
</tr>
<tr>
<td>RpS7</td>
<td>89</td>
<td>CG30490</td>
<td>119</td>
<td>MED14</td>
<td>1820</td>
</tr>
<tr>
<td>TrissinR</td>
<td>90</td>
<td>Muc30E</td>
<td>1776</td>
<td>Con</td>
<td>1824</td>
</tr>
<tr>
<td>CG12065</td>
<td>91</td>
<td>gry</td>
<td>1776</td>
<td>vn</td>
<td>1827</td>
</tr>
<tr>
<td>CG45049</td>
<td>92</td>
<td>form3</td>
<td>1777</td>
<td>gry</td>
<td>1829</td>
</tr>
<tr>
<td>ninaC</td>
<td>92</td>
<td>disco-r</td>
<td>1777</td>
<td>Cyp4d8</td>
<td>1839</td>
</tr>
<tr>
<td>Tfb4</td>
<td>93</td>
<td>CG7246</td>
<td>1779</td>
<td>CG2930</td>
<td>1846</td>
</tr>
<tr>
<td>nAChRalpha7</td>
<td>94</td>
<td>Elo68beta</td>
<td>1784</td>
<td>Ten-a</td>
<td>1851</td>
</tr>
<tr>
<td>msta</td>
<td>94</td>
<td>l(3)psg2</td>
<td>1786</td>
<td>CG6424</td>
<td>1853</td>
</tr>
<tr>
<td>SNF4Ay</td>
<td>94</td>
<td>Flo2</td>
<td>1786</td>
<td>Eip63E</td>
<td>1866</td>
</tr>
<tr>
<td>mus201</td>
<td>95</td>
<td>RPA2</td>
<td>1788</td>
<td>Bx</td>
<td>1870</td>
</tr>
<tr>
<td>PCNA</td>
<td>96</td>
<td>Ncc69</td>
<td>1788</td>
<td>CG6197</td>
<td>1873</td>
</tr>
<tr>
<td>mbl</td>
<td>96</td>
<td>CG43901</td>
<td>1791</td>
<td>CG7556</td>
<td>1881</td>
</tr>
<tr>
<td>Mat1</td>
<td>97</td>
<td>Cdk7</td>
<td>1793</td>
<td>bin</td>
<td>1893</td>
</tr>
<tr>
<td>GEfmeso</td>
<td>97</td>
<td>Sep4</td>
<td>1793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>decay</td>
<td>97</td>
<td>Xpac</td>
<td>1794</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fen1</td>
<td>99</td>
<td>Or63a</td>
<td>1795</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fog</td>
<td>99</td>
<td>NtR</td>
<td>1795</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xpd</td>
<td>100</td>
<td>msn</td>
<td>1795</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ten-a</td>
<td>101</td>
<td>Dhc64C</td>
<td>1796</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG42313</td>
<td>101</td>
<td>sn</td>
<td>1797</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG18170</td>
<td>103</td>
<td>CG14445</td>
<td>1800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ercc1</td>
<td>104</td>
<td>CG32278</td>
<td>1801</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGRP strain</td>
<td>Wolbachia infection</td>
<td>DGRP strain</td>
<td>Wolbachia infection</td>
<td>DGRP strain</td>
<td>Wolbachia infection</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>28138</td>
<td>No</td>
<td>25445</td>
<td>Yes</td>
<td>25200</td>
<td>Yes</td>
</tr>
<tr>
<td>28139</td>
<td>No</td>
<td>28182</td>
<td>Yes</td>
<td>25201</td>
<td>Yes</td>
</tr>
<tr>
<td>28140</td>
<td>No</td>
<td>28183</td>
<td>No</td>
<td>25745</td>
<td>No</td>
</tr>
<tr>
<td>28141</td>
<td>No</td>
<td>28184</td>
<td>No</td>
<td>28220</td>
<td>Yes</td>
</tr>
<tr>
<td>28142</td>
<td>Yes</td>
<td>25188</td>
<td>No</td>
<td>28131</td>
<td>Yes</td>
</tr>
<tr>
<td>28143</td>
<td>No</td>
<td>28186</td>
<td>No</td>
<td>25202</td>
<td>Yes</td>
</tr>
<tr>
<td>28144</td>
<td>Yes</td>
<td>25189</td>
<td>No</td>
<td>25203</td>
<td>No</td>
</tr>
<tr>
<td>28145</td>
<td>Yes</td>
<td>28125</td>
<td>No</td>
<td>28223</td>
<td>Yes</td>
</tr>
<tr>
<td>28146</td>
<td>Yes</td>
<td>25190</td>
<td>Yes</td>
<td>28132</td>
<td>Yes</td>
</tr>
<tr>
<td>28148</td>
<td>No</td>
<td>28188</td>
<td>No</td>
<td>25204</td>
<td>No</td>
</tr>
<tr>
<td>28150</td>
<td>No</td>
<td>28190</td>
<td>Yes</td>
<td>25205</td>
<td>No</td>
</tr>
<tr>
<td>28151</td>
<td>Yes</td>
<td>25191</td>
<td>No</td>
<td>28229</td>
<td>Yes</td>
</tr>
<tr>
<td>28153</td>
<td>No</td>
<td>28194</td>
<td>No</td>
<td>28230</td>
<td>Yes</td>
</tr>
<tr>
<td>25174</td>
<td>No</td>
<td>28126</td>
<td>No</td>
<td>28231</td>
<td>Yes</td>
</tr>
<tr>
<td>28154</td>
<td>No</td>
<td>28196</td>
<td>No</td>
<td>28232</td>
<td>Yes</td>
</tr>
<tr>
<td>28156</td>
<td>Yes</td>
<td>25193</td>
<td>No</td>
<td>25207</td>
<td>No</td>
</tr>
<tr>
<td>28157</td>
<td>No</td>
<td>25194</td>
<td>No</td>
<td>28234</td>
<td>Yes</td>
</tr>
<tr>
<td>28160</td>
<td>Yes</td>
<td>28197</td>
<td>Yes</td>
<td>28235</td>
<td>Yes</td>
</tr>
<tr>
<td>28161</td>
<td>No</td>
<td>28198</td>
<td>Yes</td>
<td>28236</td>
<td>Yes</td>
</tr>
<tr>
<td>28124</td>
<td>No</td>
<td>28128</td>
<td>No</td>
<td>28237</td>
<td>Yes</td>
</tr>
<tr>
<td>28164</td>
<td>Yes</td>
<td>28200</td>
<td>Yes</td>
<td>28238</td>
<td>No</td>
</tr>
<tr>
<td>28165</td>
<td>Yes</td>
<td>28202</td>
<td>No</td>
<td>28239</td>
<td>No</td>
</tr>
<tr>
<td>28166</td>
<td>No</td>
<td>25197</td>
<td>No</td>
<td>28240</td>
<td>No</td>
</tr>
<tr>
<td>25180</td>
<td>No</td>
<td>28207</td>
<td>Yes</td>
<td>28241</td>
<td>Yes</td>
</tr>
<tr>
<td>28167</td>
<td>Yes</td>
<td>28208</td>
<td>Yes</td>
<td>25208</td>
<td>Yes</td>
</tr>
<tr>
<td>28170</td>
<td>No</td>
<td>25198</td>
<td>Yes</td>
<td>28244</td>
<td>Yes</td>
</tr>
<tr>
<td>28171</td>
<td>No</td>
<td>28211</td>
<td>No</td>
<td>28134</td>
<td>No</td>
</tr>
<tr>
<td>28173</td>
<td>Yes</td>
<td>28213</td>
<td>Yes</td>
<td>28245</td>
<td>Yes</td>
</tr>
<tr>
<td>28176</td>
<td>No</td>
<td>28215</td>
<td>Yes</td>
<td>28246</td>
<td>Yes</td>
</tr>
<tr>
<td>28177</td>
<td>Yes</td>
<td>25199</td>
<td>Yes</td>
<td>25209</td>
<td>Yes</td>
</tr>
<tr>
<td>28178</td>
<td>Yes</td>
<td>28216</td>
<td>No</td>
<td>25210</td>
<td>Yes</td>
</tr>
<tr>
<td>25184</td>
<td>No</td>
<td>28217</td>
<td>Yes</td>
<td>28136</td>
<td>No</td>
</tr>
<tr>
<td>25185</td>
<td>No</td>
<td>28130</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28179</td>
<td>No</td>
<td>25744</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 7.1.4 Association of *Wolbachia pipiensis* infection with UVR-sensitivity
Boxplot representing association of *Wolbachia* infection status (Yes-Y or No-N) in DGRP strains with UVR-sensitivity as measured by embryo viability (Survivorship).
7.1.3 Analysis of enriched pathways and gene ontology

Fig. 7.1.5 Network of genes for UVR-sensitivity
A gene network inferred from Reactome and KEGG databases using R spider algorithm. The rectangles represent input genes (N<sub>input</sub>=100) and the triangles represent intermediate genes.

Table. 7.1.6 PANTHER GO terms

<table>
<thead>
<tr>
<th>GO term</th>
<th>No. of genes</th>
<th>Percentage genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>cellular component organization or biogenesis (GO:0071840)</td>
<td>8</td>
<td>5.20%</td>
</tr>
<tr>
<td>cellular process (GO:0009987)</td>
<td>52</td>
<td>33.50%</td>
</tr>
<tr>
<td>localization (GO:0051179)</td>
<td>21</td>
<td>13.50%</td>
</tr>
<tr>
<td>apoptotic process (GO:0006915)</td>
<td>7</td>
<td>4.50%</td>
</tr>
<tr>
<td>reproduction (GO:0000003)</td>
<td>3</td>
<td>1.90%</td>
</tr>
<tr>
<td>biological regulation (GO:0065007)</td>
<td>18</td>
<td>11.60%</td>
</tr>
<tr>
<td>response to stimulus (GO:0050896)</td>
<td>9</td>
<td>5.80%</td>
</tr>
<tr>
<td>developmental process (GO:0032502)</td>
<td>24</td>
<td>15.50%</td>
</tr>
<tr>
<td>multicellular organismal process (GO:0032501)</td>
<td>22</td>
<td>14.20%</td>
</tr>
<tr>
<td>biological adhesion (GO:0022610)</td>
<td>11</td>
<td>7.10%</td>
</tr>
<tr>
<td>metabolic process (GO:0008152)</td>
<td>54</td>
<td>34.80%</td>
</tr>
<tr>
<td>immune system process (GO:0002376)</td>
<td>13</td>
<td>8.40%</td>
</tr>
<tr>
<td>cellular component organization or biogenesis (GO:0071840)</td>
<td>8</td>
<td>5.20%</td>
</tr>
<tr>
<td>cellular process (GO:0009987)</td>
<td>52</td>
<td>33.50%</td>
</tr>
<tr>
<td>localization (GO:0051179)</td>
<td>21</td>
<td>13.50%</td>
</tr>
<tr>
<td>apoptotic process (GO:0006915)</td>
<td>7</td>
<td>4.50%</td>
</tr>
<tr>
<td>reproduction (GO:0000003)</td>
<td>3</td>
<td>1.90%</td>
</tr>
<tr>
<td>GO term</td>
<td>No. of genes</td>
<td>Percentage genes</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>N-acetylglucosamine metabolism (P02756)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Metabotropic glutamate receptor group III pathway (P00039)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Histamine H2 receptor mediated signaling pathway (P04386)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Alzheimer disease-presenilin pathway (P00004)</td>
<td>2</td>
<td>1.30%</td>
</tr>
<tr>
<td>Integrin signalling pathway (P00034)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Corticotropin releasing factor receptor signaling pathway (P04380)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Enkephalin release (P05913)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Inflammation mediated by chemokine and cytokine signaling pathway (P00031)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Dopamine receptor mediated signaling pathway (P05912)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Ubiquitin proteasome pathway (P00060)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Endothelin signaling pathway (P00019)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>PI3 kinase pathway (P00048)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Nicotinic acetylcholine receptor signaling pathway (P00044)</td>
<td>3</td>
<td>1.90%</td>
</tr>
<tr>
<td>Muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043)</td>
<td>3</td>
<td>1.90%</td>
</tr>
<tr>
<td>Cadherin signaling pathway (P00012)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Muscarinic acetylcholine receptor 1 and 3 signaling pathway (P00042)</td>
<td>2</td>
<td>1.30%</td>
</tr>
<tr>
<td>Metabotropic glutamate receptor group I pathway (P00041)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Metabotropic glutamate receptor group II pathway (P00040)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Beta2 adrenergic receptor signaling pathway (P04378)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Huntington disease (P00029)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Beta1 adrenergic receptor signaling pathway (P04377)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Pyruvate metabolism (P02772)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction (P00028)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Wnt signaling pathway (P00057)</td>
<td>4</td>
<td>2.60%</td>
</tr>
<tr>
<td>Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026)</td>
<td>2</td>
<td>1.30%</td>
</tr>
<tr>
<td>SHT1 type receptor mediated signaling pathway (P04373)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>Toll receptor signaling pathway (P00054)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>GABA-B_receptor II_signaling (P05731)</td>
<td>1</td>
<td>0.60%</td>
</tr>
<tr>
<td>FAS signaling pathway (P00020)</td>
<td>1</td>
<td>0.60%</td>
</tr>
</tbody>
</table>
7.2 Appendix for Chapter 5

7.2.1 Primers for sequencing and PCR for Dmp53 congenic strains

<table>
<thead>
<tr>
<th>Table 7.2.1 Primers for PCR and sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2 &amp; TGTGCCAGTAGCGCTTTATG</td>
</tr>
<tr>
<td>F2-2 &amp; CATTGCTCGACCATTCAAAG</td>
</tr>
<tr>
<td>R1 &amp; CCCTGCTTGAAGTTGTTTTT</td>
</tr>
<tr>
<td>R1-2 &amp; GTTGCACCTCTGAGTGCATTG</td>
</tr>
<tr>
<td>R2 &amp; TGGCTGAATTGGAGACACA</td>
</tr>
<tr>
<td>FS2 &amp; TTCCGATGGCAATAACAAAA</td>
</tr>
<tr>
<td>FS2-2 &amp; GCGCTCAAATCGATAAACAT</td>
</tr>
<tr>
<td>FS2-3 &amp; CAGCTGGTCAGTTCATCCG</td>
</tr>
<tr>
<td>FS2-4 &amp; GCGCTCAAATCGATAAACAT</td>
</tr>
<tr>
<td>RS3 &amp; TGAACATTACGGGTGGCTGTT</td>
</tr>
<tr>
<td>FS4 &amp; ACCTGCTGTCAGACGGGAAT</td>
</tr>
<tr>
<td>FS5 &amp; ACTCGAGCAAGCTGGAACAT</td>
</tr>
<tr>
<td>FS3 &amp; TTCAGCCAGGTCGGTGTAACAT</td>
</tr>
<tr>
<td>RS5 &amp; TATTGGGCCACGTAATAGCC</td>
</tr>
</tbody>
</table>
## 7.2.2 SNP annotations for Dmp53 congenic strains

### Table. 7.2.2 SNP annotations (BDGP 5) for congenic strains

<table>
<thead>
<tr>
<th>SNP</th>
<th>Absolute position (Chr3R)</th>
<th>Variant type (ensembl)</th>
<th>Alleles</th>
<th>SNP</th>
<th>Absolute position (Chr3R)</th>
<th>Variant type (ensembl)</th>
<th>Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>63</td>
<td>18880049</td>
<td>Upstream</td>
<td>A/G</td>
<td>1988</td>
<td>18878124</td>
<td>Intron</td>
<td>G/A</td>
</tr>
<tr>
<td>64</td>
<td>18880048</td>
<td>Upstream</td>
<td>A/G/C</td>
<td>2000</td>
<td>18878112</td>
<td>Intron</td>
<td>C/T</td>
</tr>
<tr>
<td>70</td>
<td>18880042</td>
<td>Upstream</td>
<td>A/G</td>
<td>2107</td>
<td>18878005</td>
<td>Intron</td>
<td>G/A</td>
</tr>
<tr>
<td>165</td>
<td>18879947</td>
<td>Upstream</td>
<td>A/G</td>
<td>2177</td>
<td>18879335</td>
<td>Intron</td>
<td>T/C</td>
</tr>
<tr>
<td>179</td>
<td>18879932</td>
<td>Upstream</td>
<td>A/G</td>
<td>2264</td>
<td>18877848</td>
<td>Intron</td>
<td>T/C</td>
</tr>
<tr>
<td>191</td>
<td>18879921</td>
<td>Upstream</td>
<td>T/C</td>
<td>2291</td>
<td>18877821</td>
<td>Intron</td>
<td>A/G</td>
</tr>
<tr>
<td>192</td>
<td>18879920</td>
<td>Upstream</td>
<td>T/A</td>
<td>2383</td>
<td>18877729</td>
<td>Intron</td>
<td>C/T</td>
</tr>
<tr>
<td>205</td>
<td>18879907</td>
<td>Upstream</td>
<td>T/A</td>
<td>2432</td>
<td>18877680</td>
<td>Intron</td>
<td>A/G</td>
</tr>
<tr>
<td>238</td>
<td>18879874</td>
<td>Upstream</td>
<td>C/T</td>
<td>2560</td>
<td>18877552</td>
<td>Intron</td>
<td>T/A</td>
</tr>
<tr>
<td>280</td>
<td>18879832</td>
<td>Upstream</td>
<td>G/A</td>
<td>2596</td>
<td>18877156</td>
<td>Intron</td>
<td>C/G</td>
</tr>
<tr>
<td>379</td>
<td>18879733</td>
<td>5 prime UTR</td>
<td>G/A</td>
<td>2780</td>
<td>18877332</td>
<td>Intron</td>
<td>A/C</td>
</tr>
<tr>
<td>584</td>
<td>18879528</td>
<td>5 prime UTR</td>
<td>G/A</td>
<td>2824</td>
<td>18877288</td>
<td>Intron</td>
<td>G/A</td>
</tr>
<tr>
<td>602</td>
<td>18879510</td>
<td>5 prime UTR</td>
<td>C/T</td>
<td>3030</td>
<td>18877082</td>
<td>Intron</td>
<td>C/A</td>
</tr>
<tr>
<td>647</td>
<td>18879465</td>
<td>5 prime UTR</td>
<td>C/T</td>
<td>3058</td>
<td>18877054</td>
<td>Intron</td>
<td>C/G</td>
</tr>
<tr>
<td>760</td>
<td>18879352</td>
<td>5 prime UTR</td>
<td>A/T</td>
<td>3466</td>
<td>18876646</td>
<td>Synonymous</td>
<td>C/T</td>
</tr>
<tr>
<td>855</td>
<td>18879257</td>
<td>5 prime UTR</td>
<td>G/T</td>
<td>3475</td>
<td>18876637</td>
<td>Synonymous</td>
<td>A/G</td>
</tr>
<tr>
<td>990</td>
<td>18879122</td>
<td>5 prime UTR</td>
<td>A/T</td>
<td>3487</td>
<td>18876625</td>
<td>Synonymous</td>
<td>A/G</td>
</tr>
<tr>
<td>1069</td>
<td>18879043</td>
<td>Intron</td>
<td>A/G</td>
<td>3529</td>
<td>18876583</td>
<td>Synonymous</td>
<td>G/C</td>
</tr>
<tr>
<td>1417</td>
<td>18878695</td>
<td>Intron</td>
<td>A/T</td>
<td>3725</td>
<td>18876387</td>
<td>Synonymous</td>
<td>G/A</td>
</tr>
<tr>
<td>1425</td>
<td>18878687</td>
<td>Intron</td>
<td>A/T</td>
<td>3770</td>
<td>18876342</td>
<td>Synonymous</td>
<td>C/T</td>
</tr>
<tr>
<td>1529</td>
<td>18878583</td>
<td>Intron</td>
<td>G/A</td>
<td>3940</td>
<td>18876172</td>
<td>Intron</td>
<td>T/A</td>
</tr>
<tr>
<td>1604</td>
<td>18878508</td>
<td>Intron</td>
<td>T/C</td>
<td>4123</td>
<td>18875989</td>
<td>Missense</td>
<td>C/A</td>
</tr>
<tr>
<td>1637</td>
<td>18878475</td>
<td>Intron</td>
<td>A/G</td>
<td>4210</td>
<td>18875902</td>
<td>Synonymous</td>
<td>T/C</td>
</tr>
<tr>
<td>1640</td>
<td>18878472</td>
<td>Intron</td>
<td>A/G</td>
<td>4264</td>
<td>18875848</td>
<td>Synonymous</td>
<td>G/A</td>
</tr>
<tr>
<td>1667</td>
<td>18878445</td>
<td>Intron</td>
<td>G/C</td>
<td>4364</td>
<td>18875748</td>
<td>Intron</td>
<td>C/T</td>
</tr>
<tr>
<td>1844</td>
<td>18878268</td>
<td>Intron</td>
<td>C/G</td>
<td>4388</td>
<td>18875724</td>
<td>Intron</td>
<td>G/C</td>
</tr>
<tr>
<td>1898</td>
<td>18878214</td>
<td>Intron</td>
<td>T/G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.2.3 SNPs in Dmp53 from European (congenic) strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Position in gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>64_M-4 LUF</td>
<td></td>
</tr>
<tr>
<td>24_M-BUR 3</td>
<td></td>
</tr>
<tr>
<td>91_M-CAV 30</td>
<td></td>
</tr>
<tr>
<td>89_M-CAV 4</td>
<td></td>
</tr>
<tr>
<td>63_M-COR 25</td>
<td></td>
</tr>
<tr>
<td>1_M-COSTA 2</td>
<td></td>
</tr>
<tr>
<td>55_M-GOT F1-3</td>
<td></td>
</tr>
<tr>
<td>30_M-GOT F1-5</td>
<td></td>
</tr>
<tr>
<td>41_M-GOT F1-8</td>
<td></td>
</tr>
<tr>
<td>116_M-HOJ 1</td>
<td></td>
</tr>
<tr>
<td>111_M-HOJ 17</td>
<td></td>
</tr>
<tr>
<td>117_M-HOJ 18</td>
<td></td>
</tr>
<tr>
<td>105_M-HOJ 19</td>
<td></td>
</tr>
<tr>
<td>119_M-HOJ 3</td>
<td></td>
</tr>
<tr>
<td>107_M-HOJ 39</td>
<td></td>
</tr>
<tr>
<td>120_M-HOJ 42</td>
<td></td>
</tr>
<tr>
<td>14_M-KNO 29</td>
<td></td>
</tr>
<tr>
<td>103_M-KOR 28</td>
<td></td>
</tr>
<tr>
<td>99_M-KOR 31</td>
<td></td>
</tr>
<tr>
<td>102_M-KOR 9</td>
<td></td>
</tr>
<tr>
<td>93_M-KN 6.15</td>
<td></td>
</tr>
<tr>
<td>29_M-REN 1414</td>
<td></td>
</tr>
<tr>
<td>56_M-REN 56-22</td>
<td></td>
</tr>
<tr>
<td>13_M-RUT 16</td>
<td></td>
</tr>
<tr>
<td>10_M-RUT 4</td>
<td></td>
</tr>
<tr>
<td>3_M-STO-F1-10</td>
<td></td>
</tr>
<tr>
<td>2_M-STO-F1-8</td>
<td></td>
</tr>
</tbody>
</table>
**7.2.4 UVR response in larvae**

Fig. 7.2.1 Larval viability in response to solar light

Larval viability scored as percentage of eclosed flies over larvae in CS (*Canton-S*), *Dfp53* (strain used to create congenic strains) and *p53[5A-1-4]* (*Dmp53 null mutant*). 25 third instar larvae (*N* replicates=5) were irradiated with 280.8KJ/m² solar simulated light (Suntest CPS+ machine) under a UV shielded fume hood. The larva were washed with distilled water and placed in a petri dish with 0.5ml PBS. The statistical difference was calculated using unpaired t-tests and standard deviation is shown. There appears to be *Dmp53* gene dose-dependent affect with CS being the most resistant and Dmp53 null mutant the least.
7.2.5 UVI map of Europe

![UVI map of Europe](image)

Fig. 7.2.2 UVI in Europe on 1\textsuperscript{st} August 2014

The levels of UVI in Europe on the 1\textsuperscript{st} August 2014 as measured by SCIAMACHY/GOME-2 instruments and data accessed using TEMIS (Tropospheric Emission Monitoring Internet Service (http://www.temis.nl/index.php) which is a part of ESA (European space agency).

7.2.6 Genes deleted/disrupted in Dfp53 strain Df(3R)BSC803

Table 7.2.3 Genes deleted/disrupted in Df(3R)BSC803

<table>
<thead>
<tr>
<th>Gene</th>
<th>CR44226</th>
<th>CG13833</th>
<th>CG6733</th>
</tr>
</thead>
<tbody>
<tr>
<td>lmd</td>
<td>Gr94a</td>
<td>CenB1A</td>
<td>p53</td>
</tr>
<tr>
<td>CG13830</td>
<td>cnc</td>
<td>CG31365</td>
<td>CG4467</td>
</tr>
<tr>
<td>CG31457</td>
<td>Rassf</td>
<td>lrk1</td>
<td>unk</td>
</tr>
<tr>
<td>CG13837</td>
<td>CG17111</td>
<td>sav</td>
<td>CG6726</td>
</tr>
<tr>
<td>fzo</td>
<td>Ublcp1</td>
<td>CG17121</td>
<td>CG6738</td>
</tr>
<tr>
<td>hh</td>
<td>CG13829</td>
<td>CG17119</td>
<td>CG17109</td>
</tr>
<tr>
<td>CG13838</td>
<td>CG17110</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7.2.7 Association analysis of adult viability and Dmp53 expression level with SNP

Fig. 7.2.3. Lack of association of adult viability and p53 expression level with SNP 4213.
Boxplots for UVR response of congenic strains carrying alleles of SNP 4123. A) Adult viability in males, B) Adult viability in females and C) Relative expression of Dmp53 in males. There is no significant effect of this SNP on these phenotypes.
7.2.7 Testing for selection

Table 7.2.4 SNPs in D. simulans p53 sequence

<table>
<thead>
<tr>
<th>Strain</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>w501</td>
<td>ATCTNNNNNTCTAGTTGGTCGCACTGCTCAGCGCTTGTCAGGCCCTCCATGGTTTGAAGCTCTT</td>
</tr>
<tr>
<td>sim.v2</td>
<td>GCCACGGAATATCCTAATGGCACCTCATAACCCACCTCACTACACAGCATACTCTGTGCTCGAGGGC</td>
</tr>
<tr>
<td>md106</td>
<td>ACTATGGCTTAAATAGCNNNNNNNNNTGCGGTGAGTTGTAANNNCGGTCTCAGGGN</td>
</tr>
<tr>
<td>md199</td>
<td>ACCATCANAANNNNNNNNNNTAGTTGTCCTTAAACGCCCTACACAGACCNNATTGCAACGCA</td>
</tr>
<tr>
<td>c1674</td>
<td>NNNNNGATATTCTACTATNNNNNNNNNNNNNNTACACCGCTTGGTCAGGNNNNNNNNNNNGC</td>
</tr>
<tr>
<td>Newc</td>
<td>NNNNNNTATTGGTTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTTTGAGC</td>
</tr>
</tbody>
</table>
7.3 Appendix for Chapter 6

7.3.1 Preliminary experiment

The survival curves for male Canton-S flies ($N_{\text{replicates}}=10$) under control (normal food), DMSO-control and 10µM drug X conditions. The overall comparison using the log-rank test reveals the curves are statistically different ($\chi^2(3, N=400)=54.15$, $p<0.0001$) after Bonferroni correction ($p<0.0001$). Flies under DMSO-control condition have reduced lifespan compared to control condition, the median survival reduced by ~13% ($\chi^2(1, N=200)=9.826$, $p=0.0017$). Flies under the 10µM drug X condition have extended lifespan compared to DMSO-control condition, median survival increased by ~35% ($\chi^2(1, N=200)=8.27$, $p=0.004$). Flies under the drug condition have extended lifespan compared to non-DMSO control condition, with an extension in median survival of ~19%; but is not significant ($\chi^2(1, N=200)=0.18$, $p=0.6714$). The 100µM drug X condition drug was stopped as it did not seem to be as effective as 10µM treatment, and would not have been an economical use of the drug to continue.
7.3.2 Weight of flies after drug administration

Table 7.3.1 Mann-Whitney tests with adjusted p-values

<table>
<thead>
<tr>
<th>Day</th>
<th>p-value</th>
<th>Adjusted p-value</th>
<th>Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0057</td>
<td>0.0076</td>
<td>**</td>
</tr>
<tr>
<td>8</td>
<td>0.0004</td>
<td>0.00064</td>
<td>***</td>
</tr>
<tr>
<td>11</td>
<td>0.0102</td>
<td>0.01165714</td>
<td>*</td>
</tr>
<tr>
<td>15</td>
<td>0.0001</td>
<td>0.0002</td>
<td>***</td>
</tr>
<tr>
<td>18</td>
<td>0.0001</td>
<td>0.0002</td>
<td>***</td>
</tr>
<tr>
<td>22</td>
<td>0.0001</td>
<td>0.0002</td>
<td>***</td>
</tr>
<tr>
<td>26</td>
<td>0.0001</td>
<td>0.0002</td>
<td>***</td>
</tr>
<tr>
<td>29</td>
<td>0.459</td>
<td>0.459</td>
<td>ns</td>
</tr>
</tbody>
</table>
Chapter 8: Bibliography


Hartl, D.L., 2000. *A primer of population genetics*. [e-book]. Sunderland, Mass: Sinauer Associates. Available from: http://le.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtiz0EUrEwxSDY0TLuYNDJTDMyTjMy5zBNNzEmUmpIsOeVgWcPEFcfihXwbk1MTKI5ogySbg4hzh66OanxFQGNERTJ9DR6IzIDcAfnKqBINCcJjJqoFJcpppRSKwyE1LTEwBttVTILIFdmOQU0DlvAM4-JOM.


Koning, T.M., Davies, R.J., Kaptein, R., 1990. The solution structure of the intramolecular photoprodut of d(TpA) derived with the use of NMR and a combination of distance geometry and molecular dynamics. Nucleic Acids Research. 18, 277-284.


