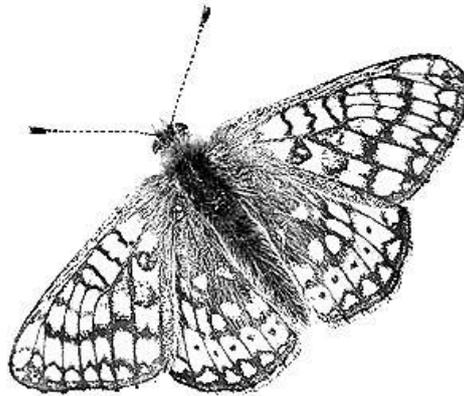


# Population Ecology and Genetics of the Marsh Fritillary Butterfly *Euphydryas aurinia*.



Submitted by Melanie Rose Smee to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Sciences in June 2011.

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: .....  .....

## ABSTRACT

The past two decades have witnessed an unprecedented decline in Lepidopteran species, with more than a third of the UK's butterflies now either considered threatened, or already lost from the country. The vulnerable marsh fritillary, *Euphydryas aurinia*, after a long term loss in the UK of 73% in abundance, has become an almost iconic species as the target of many well-funded conservation projects across the UK. Despite extensive ecological studies, populations of *E. aurinia* are shown in Chapter 2 to still be declining in south-west UK even after recommended management strategies have been implemented. This necessitates the need for prompt research beyond that of management requirements and butterfly habitat preferences.

In Chapter 3, microsatellite markers (EST-SSRs) were developed for *E. aurinia* and using these markers in Chapter 4, it is shown that *E. aurinia* populations in southern UK and Catalonia, Spain, are severely genetically differentiated at all geographical scales, and genetically depauperate, causing huge concerns for the conservation of this enigmatic and ecologically important species.

Dispersal is fundamental to metapopulation existence and survival. Phosphoglucose isomerase (PGI – an enzyme in the glycolysis pathway) is a well-endorsed candidate gene for dispersal, extensively studied in the Glanville fritillary (*Melitaea cinxia*) and Orange Sulphur (*Colias eurytheme*). In Chapter 5, an analysis across 27 sites in the UK discovered six non-synonymous SNPs (single nucleotide polymorphisms) within PGI. A single charge-changing SNP of interest showed no evidence of balancing selection, contrary to findings in *M. cinxia*, instead appearing to be neutral when analysed alongside microsatellite markers developed in Chapter 3. No link was found between genotype and flight, morphology or population trend. These findings challenge the emerging perspective that PGI could be used as an adaptive molecular marker for arthropods.

*Wolbachia* are endosymbiotic bacteria capable of dramatically altering the reproductive system of their host. In Chapter 6, a PCR-based diagnostic in conjunction with MLST (multi-locus sequence typing) identified 100% prevalence of a single strain of *Wolbachia* across all sampled *E. aurinia* populations in the UK. Total prevalence suggests that *Wolbachia* probably has little phenotypic impact on its host, but the

potential impacts of this endosymbiont on uninfected populations should be considered during any management plans for the conservation of *E. aurinia*.

Current management plans will need to incorporate all areas of research, from basic ecological requirements to molecular adaptation and unseen manipulators of host biology, to be able to fully and effectively conserve declining fragmented species.

## **ACKNOWLEDGEMENTS**

There are a small number of people I couldn't have done this without, and a large number of people I wouldn't have wanted to do it without.

I will be forever grateful to my supervisors, Dave Hodgson (a.k.a. 'Hodgy') and Richard French-Constant, for giving me the opportunity to do this PhD in the first place, fresh out of my Undergrad. The most humungous thanks go to Hodgy, for making me feel special, I quote "there's only one Mel Smee", but capable; for the door that was always open (and I had to creep past to avoid detection); and for letting me "fly away", knowing I would come back. Thanks also to Richard, for his constant humour and updates on my grandmother (!). And to Caroline Bulman, my supervisor at Butterfly Conservation, for always being encouraging and having a few spare minutes for me even when pregnant.

To Yannick Pauchet I owe so much – there isn't a big enough thank you for it. For always being there and supporting and encouraging me – picking me up when I was low, and bringing me back down to earth when my head was in the clouds. For nurturing my metamorphosis from an ecologist into a molecular ecologist, and so much more.

Two of the chapters of this thesis wouldn't have been possible without the amazing generosity and willingness of Mike Singer and Sasha Mikheyev – I am hugely indebted to both for the opportunity given to me to visit Okinawa, and the chance to meet some wonderful people; Tanya, Yutaka, Hitomi, Pascal, Roxy, Laurent, Émile, Yoko & more.

There is no way I would have kept my sanity through these few years if it wasn't for Julia Reger. Despite being a few hundred miles away, I am so thankful to Ju, just for always being there for me, and the constant emails preventing me from going completely crazy. It could never compare with sitting down with a cup of tea and biscuits and putting the world to rights, but it was damn close.

However, Kate Plummer and I did put the world to rights with endless cups of tea, giggles and chatting, and even tears. I am incredibly grateful for Kate's friendship over the past few years, even from the other side of the globe. I couldn't have asked to share a house for 3 years with better people than Kate and Xav Harrison, who provided much banter and makes the best chocolate cake in the world! A special thanks also goes to

Laura Bailey, for being so understanding over these years, and for just being you and always putting a smile on my face; and to Ceri Simmons, who also gave me many reasons to smile over these past few years, including her friendship and the chance to escape and be around horses whenever a spare minute allowed it. And to Pip and the Dorset crowd, my pseudo-family – I just love you loads!

There are endless lists of people who have made the past 3 or so years such brilliant fun. I would like to thank Andy – for my grey hairs. I attribute half to you and half to the PhD itself. Fieldwork would have been a much more tiresome task if it had not been for many helpers in the field: R. Hobson; M. Davey; M. Easter; M. Tunmore; the folk on the Lizard NNR Natural England team (for tea and giggles); Phil Harris for endless enthusiasm; MSc student Sian Rowland; and Nat Cumber, (Brown) Patch and Helen - for making life totally full to the brim of laughs (and Chelsea buns). Thanks also to NE, BC, CWT, DWT and all landowners allowing access to land, as well as all the voluntary collectors in Wales (CCW) and Scotland (SNH), especially Adrian Fowles and Tom Prescott, for all their help. To Blount and Hosken, for ‘corridor banter’ in the form of abusive shouts of ‘Smelly’ and ‘Smeeple’ down the hall way; to the ‘one month post-doc’ Sam – amazing how quickly you can form a firm friendship with someone; to Debbie Mason, the conservationist extraordinaire (marsh frits would be extinct now if you weren’t around to save them); to Casper Breuker, Chris Wheat, Paul Wilkinson, Michelle Hares, Nina Wedell, Zen Lewis, Anna Leonard and many more I’m sure – for help and advice freely given; and of course to all the past and present occupiers of the PhD office – it’s the people that make the experience, and mine has been a fantastic one at Tremough – Devi, Iain, Caro, Ruth, Cheryl, Damo, Lucy, Nic Chamberlain, Erika, Ross, Josie, Jan, Joe and Corrina, Sahran, Fran, Will, Stuart Hinchliffe, Andrea, Si Pickett, Dom & Dom, Ritika, Iva and many more.... A big thank you also goes to all the friends at home in Portsmouth, in Australia, and those made throughout my Undergrad who have stayed in touch and in some way or another helped me through the PhD.

And finally, huge thanks to my Mum, for knowing that ‘no phone call is a good sign’ and always understanding and supporting me; and to Dad and Alice – for constant encouragement and perspective. And finally to my brother Joe – for not leaving the planet entirely and coming back when I needed you most. I love you all very much.

# Table of Contents

<b>Abstract</b>	<b>2</b>
<b>Acknowledgements</b>	<b>4</b>
<b>Table of Contents</b>	<b>6</b>
<b>List of Figures and Tables</b>	<b>10</b>
<b>List of Plates</b>	<b>12</b>
<b>Author's Declaration</b>	<b>13</b>
<b>Definitions and Abbreviations</b>	<b>15</b>
<b>Chapter 1: Modern approaches to the conservation of lepidopteran species in fragmented landscapes. ....</b>	<b>18</b>
1.1 THE CURRENT CONSERVATION CRISIS .....	19
1.2 LEPIDOPTERA IN FRAGMENTED LANDSCAPES .....	19
1.3 CONSERVATION OF DECLINING SPECIES .....	21
1.3.1 ‘Visible’ factors to consider.....	21
1.3.2 ‘Invisible’ factors of importance.....	22
1.3.3 Captive breeding and reintroductions .....	24
1.4 CRITICAL FUTURE DIRECTIONS FOR CONSERVATION ECOLOGISTS .....	25
1.5 INTRODUCTION TO STUDY SPECIES .....	26
1.6 SCOPE OF THESIS .....	29
<b>Chapter 2: Butterflies on the brink: habitat requirements for declining populations of the marsh fritillary (<i>Euphydryas aurinia</i>) in SW England.....</b>	<b>31</b>
2.1 ABSTRACT .....	32
2.2 INTRODUCTION .....	33
2.3 MATERIALS AND METHODS .....	35
2.3.1 Study species.....	35
2.3.2 Study area .....	35
2.3.3 Field Techniques.....	36
2.3.4 Data Analysis .....	37
2.4 RESULTS .....	39
2.4.1 Can habitat variables explain the presence of <i>E. aurinia</i> ?.....	39
2.4.2 Has management achieved optimum levels of habitat variables favoured by <i>E. aurinia</i> ? .....	42
2.5 DISCUSSION .....	48

2.5.1	Using management tools to create suitable habitat.....	48
2.5.2	Response of the butterfly – are we doing enough?.....	49
2.5.3	Implications for the conservation of <i>E. aurinia</i> across the mid-Cornwall Moors, and beyond. ....	51
<b>Chapter 3: Development of microsatellites for the Vulnerable marsh fritillary butterfly by de novo transcriptome sequencing.....</b>		<b>52</b>
3.1	ABSTRACT .....	53
3.2	INTRODUCTION .....	54
3.3	MATERIALS AND METHODS .....	55
3.4	RESULTS .....	59
3.5	DISCUSSION .....	60
<b>Chapter 4: To Protect or Connect? Population genetic analyses inform conservation strategies for the marsh fritillary butterfly, <i>Euphydryas aurinia</i>.....</b>		<b>61</b>
4.1	ABSTRACT .....	62
4.2	INTRODUCTION .....	63
4.3	MATERIALS AND METHODS .....	66
4.3.1	Sample collection and preparation.....	66
4.3.2	Genotyping.....	68
4.3.3	Analysis of genetic diversity and variation.....	69
4.3.4	Population genetics – is sampling information needed?.....	69
4.4	RESULTS .....	71
4.4.1	International broad-scale analysis.....	73
4.4.2	Regional fine-scale analysis.....	76
4.5	DISCUSSION .....	82
4.5.1	Genetic variation.....	83
4.5.2	Isolation by distance .....	83
4.5.3	Population structure .....	84
4.5.4	Comments on methods and future work.....	85
4.5.5	Conclusions.....	87
<b>Chapter 5: <i>Pgi</i>'s Might Fly: A purported 'gene for' dispersal shows no signal of selection in the marsh fritillary butterfly. ....</b>		<b>89</b>
5.1	ABSTRACT .....	90
5.2	INTRODUCTION .....	91
5.3	MATERIALS AND METHODS .....	94
5.3.1	SNP genotyping .....	94
5.3.2	Flight assays in the field .....	97

5.3.3	Statistical analysis of SNP genotype against ecological data .....	98
5.4	RESULTS .....	99
5.4.1	SNP variation .....	99
5.4.2	Is PGI under selection? .....	102
5.4.3	Genotypic effects on ecological variables .....	103
5.5	DISCUSSION .....	105
5.5.1	Variation in <i>Pgi</i> genotype frequencies.....	105
5.5.2	Relating <i>Pgi</i> genotypes to flight and population trend .....	107
5.5.3	Further studies needed for <i>E. aurinia</i> .....	107
5.5.4	Conclusions.....	108
5.6	SUPPLEMENTARY MATERIAL .....	110
5.6.1	RNA isolation and cDNA synthesis .....	110
5.6.2	Sequencing of <i>Pgi</i> cDNA .....	110
5.6.3	Single nucleotide polymorphisms (SNPs) in <i>Pgi</i> across 27 sites .....	111
<b>Chapter 6: Friend or foe? <i>Wolbachia</i> infection in the threatened British butterfly</b>		
<b><i>Euphydryas aurinia</i>..... 114</b>		
6.1	ABSTRACT .....	115
6.2	INTRODUCTION .....	116
6.3	MATERIALS AND METHODS .....	119
6.3.1	Sample collection and preparation.....	119
6.3.2	Detecting <i>Wolbachia</i> .....	120
6.3.3	Multilocus sequence typing (MLST).....	121
6.3.4	Phylogenetic analysis.....	122
6.4	RESULTS .....	123
6.4.1	Screening for <i>Wolbachia</i> .....	123
6.4.2	Multilocus sequence typing (MLST).....	124
6.4.3	Phylogenetic analysis.....	125
6.5	DISCUSSION .....	129
<b>Chapter 7: General Discussion ..... 134</b>		
7.1	GENERAL DISCUSSION .....	135
7.1.1	Baseline research of species ecology is still needed.....	135
7.1.2	Neutral molecular markers: useful, but questionable?.....	136
7.1.3	Adaptive molecular markers in lepidopteran conservation .....	140
7.1.4	Out of sight, out of mind – unseen foes must not be forgotten.....	141
7.1.5	Management recommendations .....	142

7.1.6 Conclusions.....	144
<b>Appendices.....</b>	<b>146</b>
<b>Appendix 1: Genomic DNA extraction protocol .....</b>	<b>147</b>
<b>Appendix 2: Sequencing the glycolytic enzyme phosphoglucose isomerase (<i>Pgi</i>) from</b>	
<b><i>Euphydryas aurinia</i>.....</b>	<b>148</b>
WHAT IS <i>Pgi</i> ?.....	148
WHY IS <i>Pgi</i> OF INTEREST? .....	149
ATTEMPTS TO OBTAIN THE WHOLE GENOMIC LENGTH OF <i>Pgi</i> .....	150
<b>References.....</b>	<b>153</b>

# List of Figures and Tables

## CHAPTER 1

- FIGURE 1.1 – The annual number of publications including the word *metapopulation*\* in the abstract or title when ISI Web of Science was queried. .... 20
- FIGURE 1.2 – The change in the British distribution of *E. aurinia* from the period of 1690 – 1994 to more recent times of 2000 – 2004. .... 27

## CHAPTER 2

- FIGURE 2.1 - South-west England, showing the project area across the Mid-Cornwall Moors ..... 36
- TABLE 2.1 - Results of a linear mixed effects model using multiple regression to relate habitat variables with larval web densities and adult butterfly densities..... 40
- TABLE 2.2 - Results of linear mixed effects models using single regressions to determine the gross influence of each habitat variable on density of larval webs and adult butterflies. .... 41
- FIGURE 2.2 - The significant quadratic effect of stock grazing on density of larval webs ..... 42
- FIGURE 2.3 - The year when grazing was started has a significant effect on the occurrence of *S. pratensis* ..... 43
- FIGURE 2.4 - Temporal comparison of *S. pratensis* density for each site over the five years of the project..... 44
- FIGURE 2.5 - Autumn sward height is significantly affected by the year in which grazing started and the intensity of grazing ..... 46
- FIGURE 2.6 - Temporal comparison of log transformed larval web densities and adult densities across the five years of monitoring at all sites and Individual temporal trends of web densities at only those sites where *E. aurinia* has an extant colony 47

## CHAPTER 3

- TABLE 3.1 – Characteristics of 14 microsatellite loci in *Euphydryas aurinia* (Lepidoptera: Nymphalidae)..... 58
- FIGURE 3.1 – Hardy-Weinberg equilibrium (HWE) ‘exact test’ statistics for each of the 28 populations included in the study ..... 59

## CHAPTER 4

FIGURE 4.1 – Map of samples sites across the UK and Catalonia area of Europe. ....	67
TABLE 4.1 – Microsatellite summary statistics for nine loci used to genotype <i>E. aurinia</i> across the UK and Catalonia area of Europe. ....	68
FIGURE 4.2 – LOSITAN results for all loci across all samples .....	69
TABLE 4.2 - Sites across the UK and Catalonia area of Europe from which <i>E. aurinia</i> was sampled, showing genetic diversity statistics per ‘geographical’ site .....	72
TABLE 4.3 - Euclidean distances (km) against genetic distance for the five regions studied.....	73
FIGURE 4.3 – Isolation-by-distance across all samples.....	74
FIGURE 4.4 - Output of STRUCTURE analysis for genetic clusters across the entire region studied.....	75
TABLE 4.4 – Analysis of molecular variance (AMOVA) results at an international scale, comparing the use of geographical and genetic ‘populations’. ....	76
TABLE 4.5 - Pairwise $F_{ST}$ distances against Euclidean distances (km) for populations in southern England. ....	77
TABLE 4.6 - Pairwise $F_{ST}$ distances against Euclidean distances (km) for populations in Catalonia.....	78
FIGURE 4.5 – IBD analyses for southern England and catalonia .....	79
FIGURE 4.6 – Output from a Bayesian clustering analysis in STRUCTURE v.2.3.3 for populations in southern England, and populations in the Catalonia area of Europe .....	80
FIGURE 4.7 – Overall view of clustering analyses in STRUCTURE v.2.3.3 .....	81
TABLE 4.7 – Analysis of molecular variance (AMOVA) results at a regional scale, comparing the use of geographical and genetic ‘populations’. ....	82
TABLE 4.8 – A summary of studies investigating population genetics of <i>E. aurinia</i> .....	85

## CHAPTER 5

FIGURE 5.1 – Map of samples sites across the UK and Catalonia.....	94
TABLE 5.1 - The observed frequencies of each genotype at SNP 373 across 30 sampling sites .....	100
FIGURE 5.2 - Genotype frequencies of <i>Pgi</i> SNP 373 across the study area, including southern England, as well as Wales and Scotland and Catalonia.....	101
FIGURE 5.3 - LOSITAN results for all samples in the UK for which there was data for both microsatellites and <i>Pgi</i> .....	102

FIGURE 5.4 - Logged means $\pm$ standard errors of distance flown and duration of flight for butterflies of differing genotypes.....	103
FIGURE 5.5 – Population trend against the frequency of each genotype at a site .....	104
FIGURE 5.S1 – Twenty seven sites from which an individual larvae was sampled and sequenced for the coding region of <i>phosphoglucose isomerase (Pgi)</i> .....	112
FIGURE 5.S2 – Synonymous and non-synonymous single nucleotide polymorphisms (SNPs) across the twelve exons of <i>Pgi</i> . .....	113

## CHAPTER 6

FIGURE 6.1 – Sample sites in the south-west UK, Wales and Scotland .....	119
FIGURE 6.2 – An example agarose gel for scoring presence/absence of <i>Wolbachia</i> .....	121
TABLE 6.1 – List of primers used during screening and MLST .....	122
TABLE 6.2 – Population information from each site sampled and numbers of individuals from each site screened with <i>wsp</i> primers .....	123
TABLE 6.3 – Putative sex ratios at a sub-sample of populations .....	124
FIGURE 6.3 – Comparison of <i>wsp</i> sequences from strains of differing phenotype, via sequence alignment and phylogenetic analysis .....	126

## CHAPTER 7

FIGURE 7.1 – Interplay of population spatial structure, IBD and $F_{ST}$ .....	139
--	-----

## APPENDICES

FIGURE A1 – The metabolic pathway of glycolysis. ....	149
TABLE A1 – Primer sequences tested for genomic length of <i>Pgi</i> .....	150
FIGURE A2 – Progress made to gain the genomic length of <i>Pgi</i> .....	152

# List of Plates

## CHAPTER 1

PLATE 1.1 – Dorsal and ventral views of an adult <i>E. aurinia</i> at Lydlinch, 2009.....	28
PLATE 1.2 – A web of 4 <sup>th</sup> instar <i>E. aurinia</i> larvae at Stithians, 2008.....	29

## **AUTHOR'S DECLARATION**

This thesis was made possible by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) in conjunction with a CASE partnership from Butterfly Conservation (BC). All of the chapters presented in this thesis were written by M. R. Smee, with additional editing by D. J. Hodgson. Individual contributions to each chapter are listed below.

### **CHAPTER 2** – published in *Journal of Insect Conservation*.

An initial 5 year partnership project was undertaken by Natural England; Cornwall Wildlife Trust; Butterfly Conservation; The Environment Agency and The Highways Agency, across central Cornwall as part of the Natura 2000 series of sites - sites of European importance for nature conservation. W. Smyth (Natural England) and M. Tunmore (Cornwall Environmental Consultants) were involved in this initial project. M. R. Smee continued the project for another year, analysed the 6-year dataset and wrote the manuscript. D. J. Hodgson assisted with statistical issues and reviewed the manuscript. W. Smyth and M. Tunmore reviewed the manuscript before it was sent for publication.

### **CHAPTER 3**

M. C. Singer (University of Texas) was instrumental in making this study possible, and also provided samples of *E. aurinia* from European populations, as did B. Wee (National Ecological Observatory Network, USA). R. ffrench-Constant provided funding for the transcriptome of *E. aurinia* to be gained from 454-pyrosequencing and Y. Pauchet extracted RNA with M. R. Smee and prepared the normalized library. P. Wilkinson assembled the 454 data and ran the query script for perfect microsatellite repeats. A. S. Mikheyev (Okinawa Institute for Science and Technology) hosted M. R. Smee whilst the lab work was carried out and provided the funding for all lab consumables, travel and accommodation for the two month duration, as well as advice on microsatellite marker choice and analysis. Initial draft prepared by M. R. Smee and improved by Y. Pauchet and D. J. Hodgson.

### **CHAPTER 4**

Contributions as for Chapter 3. X. Harrison provided initial training in the use of STRUCTURE, and R. ffrench-Constant provided useful comments on the initial draft.

## **CHAPTER 5**

Initial idea to study PGI in *Euphydryas aurinia* suggested by R. ffrench-Constant, who also provided funding for lab consumables and sequencing throughout the project. Y. Pauchet provided M. R. Smee with training in molecular techniques, both in the laboratory and on computer programmes, from PCR and primer design to sequence alignment. P. Wilkinson did the sequencing and D. Smith carried out some final PCRs to complete the dataset for analysis. M. R. Smee carried out statistical analyses and wrote the initial draft, both of which were improved upon by D. J. Hodgson.

## **CHAPTER 6**

Initial concept by M. R. Smee and Y. Pauchet. Sequencing by P. Wilkinson, with all other lab work, analysis and writing of draft by M. R. Smee. D. J. Hodgson, R. ffrench-Constant, Y. Pauchet and Z. Lewis commented on and improved initial draft.

## DEFINITIONS & ABBREVIATIONS

454 pyrosequencing	One method of next-generation sequencing (NGS), originating from a company called 454 Life Sciences but now owned by Roche Diagnostics, which uses a "sequencing by synthesis" approach - relying on the detection of pyrophosphate release on nucleotide incorporation rather than chain termination as in Sanger sequencing. See NGS definition for more general information.
AFLP	<u>A</u> mplified <u>f</u> ragment <u>l</u> ength <u>p</u> olymorphism PCR uses primers for known regions of the genome to amplify genomic DNA, which is then digested by restriction enzymes and run on polyacrylamide gels to visualise differences in fragment lengths.
bp	<u>b</u> ase <u>p</u> airs of DNA.
cDNA	Complementary DNA synthesized from an mRNA template.
Co-dominant markers	Molecular markers in which both the alleles present at a locus contribute to the expressed phenotype and heterozygotes can be distinguished from homozygotes.
Contig	Often used to determine the original DNA sequence of the source material, contigs are sets of overlapping segments of DNA from a single genetic source.
Endosymbiont	An organism living in close contact – symbiosis – with another organism for the majority of its life.
EST	<u>E</u> xpressed <u>s</u> equences <u>t</u> ag: a short sub-sequence of a cDNA sequence.
HWE	<u>H</u> ardy- <u>W</u> einberg <u>e</u> quilibrium: allele and genotype frequencies in a population are in equilibrium, or remain constant, from generation to generation.
LIFE and Natura2000	LIFE (the Financial Instrument for the Environment), is one of the spearheads of the European Union's environmental policy and was introduced in 1992. It co-finances projects in three areas: LIFE-Nature, LIFE-Environment and LIFE-Third Countries. The LIFE-Nature fund supports the positive management and enhancement of sites of European importance for nature conservation, collectively known as the Natura 2000 series of sites - SPAs and SCIs (Site of Community Importance).

Metapopulation	Existing at the landscape level and consisting of several (>2) smaller geographically separated subpopulations. The metapopulation as a whole can persist if there is a balance between stochastic extinctions and recolonisations across the fragmented landscape.
Microsatellite, (EST-SSR)	Also known as simple sequence repeats (SSRs) or short tandem repeats (STRs). These are tandem repeating sequences of (perfect or compound) di-, tri- or tetra- (and so on) nucleotide motifs, for example, a tri-nucleotide repeat; CAACAACAACAACA. Microsatellites are co-dominant and typically neutral, hence their use as molecular markers in population genetics. If derived from a transcriptome, as is the case here, they are also referred to as EST-SSRs, which are SSRs derived from ESTs (see above).
MLST	<u>M</u> ulti <u>l</u> ocus <u>s</u> equencing <u>t</u> yping: isolates of bacterial species are characterised using DNA sequences of internal segments of several housekeeping genes (usually five or six). Each loci may have several alleles, and so the combination of alleles across these genes defines the unique allelic profile (or sequence type - ST) of the bacterial strain.
Neutral variation	Variation that has a small selection coefficient relative to the population size, such that $ s  < 1/2N_e$ , where $s$ is either the selective disadvantage of a detrimental genetic variant or the selective advantage of an adaptive genetic variant and $N_e$ is the effective population size.
NGS	<u>N</u> ext- <u>g</u> eneration <u>s</u> equencing is massively parallel DNA sequencing, where millions of reads (sequences) can be generated over the course of a few hours. Generally, the reads generated are smaller (300 - 500 bp) than when using Sanger sequencing technologies (800 - 1000 bp). Roche 454 FLX Titanium system (454 pyrosequencing) is one of the better known platforms for NGS, along with Illumina's (Solexa) Genome Analyser and ABI's SOLiD.
Null Alleles	allele(s) that fail to amplify within an individual because of mutations at primer annealing sites, and hence cause an individual to appear homozygote for the second allele present at that locus. Null alleles can be a significant problem in studies of population genetics as they skew results towards an excess of homozygotes, where there actually might not be an excess.
PCR	<u>P</u> olymerase <u>c</u> hain <u>r</u> eaction: a technique to amplify thousands to millions of copies of a particular DNA sequence from a single or just a few original copies.

PGI	<b><u>P</u>hospho<u>g</u>lucose <u>i</u>somerase</b> : an enzyme in the metabolic pathway of glycolysis. PGI catalyses the second step of the pathway in which glucose-6-phosphate is converted to fructose-6-phosphate. High energy molecules of ATP are released during glycolysis, which are used for flight and other essential cellular functions.
SNP	<b><u>S</u>ingle <u>n</u>ucleotide <u>p</u>olymorphism</b> : a difference of a single nucleotide (A, T, C or G) in the DNA sequence of different individuals.
Subpopulation	One unit of a metapopulation, containing a breeding population of the species.
Thelytoky	Parthenogenesis in which only female offspring are produced.
Univoltine	One generation a year.
UTR	<b><u>U</u>n<u>t</u>ranslated <u>r</u>egion</b> : the sections either side of the coding region of a gene which are not translated into amino acids.
'vulnerable'	IUCN Red Data List category: species which are classified as 'vulnerable' are likely to become 'endangered' unless trends are reversed. A decline over 10 years of 30 - 49% deems a species as 'vulnerable'.

# Chapter 1



## **Modern approaches to the conservation of lepidopteran species in fragmented landscapes.**

*'...the study of butterflies – creatures selected as the types of airiness and frivolity  
– instead of being despised, will someday be valued as one of the most  
important branches of Biological science.'*



*Henry Walter Bates 1863  
Naturalist on the River Amazon*

## 1.1 THE CURRENT CONSERVATION CRISIS

As sensitive ‘environmental indicators’ of habitat health, and with very short generations, the Lepidoptera are the first major taxon to visibly illustrate the effects of habitat deterioration or other ‘unseen foes’. The past few decades have witnessed an unprecedented decline in lepidopteran species, with more than a third of the UK’s regularly breeding butterflies now either considered threatened, or already lost from the country (37% - Fox *et al.* 2010) in comparison to 29% of birds (not including extinct species; Eaton *et al.* 2005) and 20% of vascular plants (Cheffings and Farrell 2005). Anthropogenic impacts during the current and previous century are the main causes of these rapid declines, via changes in the climate and landscape (for examples, see: Hill *et al.* 1999; Maes and Van Dyck 2001; Thomas and Hanski 2004; Van Dyck *et al.* 2009; Warren *et al.* 2001; Wilson *et al.* 2007). As almost 77% of the UK’s land surface is used to some degree in farming practices (Department of the Environment 1995), it is no wonder that much of this global ‘sixth extinction’ of biodiversity (Leakey and Lewin 1996) is attributable to agricultural improvements which result in habitat fragmentation, destruction, or degradation (Fahrig 2003; Pimm *et al.* 1995; Saunders *et al.* 1991).

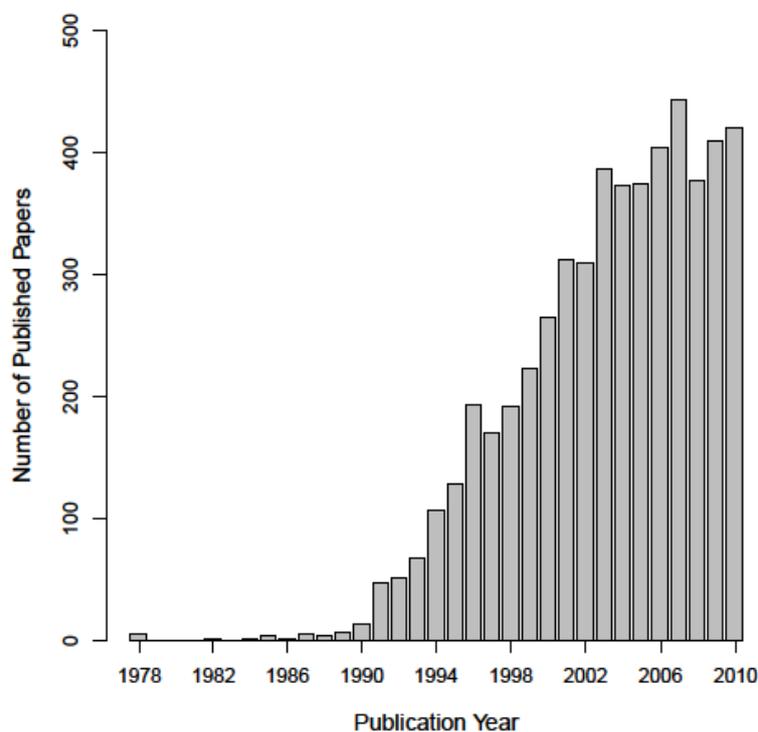
As human populations continue to grow rapidly (United Nations 2010) increased strain is placed upon biodiversity to not only survive, but also provide economic value and a benefit to society (i.e. food production; health benefits and potential cures for diseases; physical protection; a diverse pool of genetic material for improving crops and domestic species; and so on). The more traditional forms of land management have consequently become unsustainable and uneconomic (Department of Environment 1995), and species must therefore rapidly adapt to changing environments, or be lost to the ever-increasing numbers of extinctions.

## 1.2 LEPIDOPTERA IN FRAGMENTED LANDSCAPES

As the loss of suitable habitat invariably leads to fragmentation and the creation of small, isolated patches of habitat across the landscape, lepidopteran species existing in these landscapes become increasingly at risk of extinction from environmental and demographic stochasticity (Wilson and Roy 2009). The focus of research on these fragmented distributions of species over the past few decades introduced the concept of a ‘population’ of populations (Levins 1969), known commonly as ‘metapopulations’.

‘Metapopulations’ are hence groups of local populations existing in discrete patches of suitable habitat linked by migration, and exhibiting extinction-colonisation dynamics (Hanski 1999). Metapopulation theory states that persistence in a fragmented landscape requires a balance, or quasiequilibrium, between these stochastic extinctions and re-colonisations in the form of asynchronous fluctuations in constituent local populations (Hanski 1998; 1999; Hanski and Gaggiotti 2004). Processes of disturbance, succession, or habitat improvements, can cause these patch networks to be dynamic, whereby metapopulation dynamics will then follow (or fail to follow) the shifting mosaic of suitable environmental and habitat conditions (Thomas 1994). In an anthropogenically modified landscape, restoration or creation of new habitat can allow such habitat tracking to occur and metapopulations to persist.

The number of citations to the term *metapopulation* has increased almost linearly since 1990 (**Figure 1.1**), demonstrating the increased awareness that many species exist in this type of population and because of the detrimental effects of habitat fragmentation and loss, many are also in need of drastic, but sound, conservation approaches.



**Figure 1.1** – The annual number of publications including the word *metapopulation*\* in the abstract or title when ISI Web of Science was queried.

### 1.3 CONSERVATION OF DECLINING SPECIES

#### 1.3.1 ‘Visible’ factors to consider

The initial, and intuitive, approach to the conservation of declining lepidopteran species is to ascertain habitat preferences and requirements, and then manage habitats accordingly. Due to the complexity of management needs, and an inadequate understanding of the causes of decline, many early attempts to conserve declining butterfly species failed (New *et al.* 1995; Thomas 1980; Thomas 1995). An emphasis on sound ecological understanding encouraged more long term, detailed, research-based studies to be undertaken, taking into account both the preferences of the immature larval stages and the adult stage to conserve the species as a whole, whilst also analysing the effect of patch size and isolation on a landscape scale (Anthes *et al.* 2003; Eichel and Fartmann 2008; Liu *et al.* 2006; Severns *et al.* 2006; Turner *et al.* 2009). Such approaches have proven highly successful for a few species of concern, such as the return of the large blue butterfly, *Maculinea arion*, to the UK (Thomas *et al.* 2009), and recovery of both the silver-spotted skipper butterfly, *Hesperia comma*, (Davies *et al.* 2005) and the chalkhill blue butterfly, *Polyommatus coridon*, (Brereton *et al.* 2008) after appropriate habitat management.

As increasing levels of complexity were added to the requirements of a ‘thorough’ conservation approach, such as the agreement that both habitat quality and isolation of patches were of crucial importance (Bauerfeind *et al.* 2009; Moilanen and Hanski 1998; Thomas *et al.* 2001) along with the dispersal tendencies of a species (Clobert *et al.* 2004; Conradt *et al.* 2000; Conradt *et al.* 2001; Singer and Ilkka 2004; Wilson and Thomas 2002), the need for reliable metapopulation models grew. A wide variety of modelling tools have thus been developed to address an equally wide variety of aspects of metapopulations; from the original spatially implicit Levins model (Levins 1969), to the acclaimed ‘crystal ball’ of conservation: (meta)population viability analyses (PVA), such as individual-based models (IBM) and stochastic patch occupancy models (SPOMs; Ovaskainen and Hanski 2004b; 2004a). PVA has since been developed as a tool to ‘predict’ the future movements of species and aid management decisions (Schtickzelle and Baguette 2009; Schtickzelle *et al.* 2005). Bulman *et al.* (2007) fitted the incidence function model (IFM - the most popular and developed SPOM; Hanski 1994) to populations of the vulnerable marsh fritillary butterfly, *Euphydryas aurinia*, in the UK, which suggested long term survival in the largest networks of patches, but a

severe lack of available habitat in other patches and widespread extinction debt (Bulman 2001; Bulman *et al.* 2007), necessitating urgent conservation management to avoid local extinction in those areas.

### 1.3.2 'Invisible' factors of importance

The field of conservation genetics is a relatively new area of research, facilitated greatly by the recent advances in molecular markers and technologies (Behura 2006; Primmer 2009). The use of 'neutral' molecular markers, those that measure neutral variation and are not under the influence of natural selection, and which illustrate only the effects of demography and their own mutational rates, has been useful to inform patterns of relatedness, gene flow, population structure, past bottlenecks, past effective population size, inbreeding and other crucial parameters for aiding conservation of declining species (For a review of neutral variation and implications for conservation genetics, see: Kohn *et al.* 2006). More recent multilocus genotyping methods have been used in population genetic structure studies to allocate individuals to genetic 'populations' and estimate the different genetic contribution of source populations to admixed populations using either a maximum likelihood or Bayesian estimation procedure for likelihood-based approaches (Gaggiotti 2004; Pritchard *et al.* 2000). The most important factors determining population structure are the spatial distribution of habitat, and the ability of individuals to disperse across the intervening habitat matrix. Hence, coupled with the effects of landscape structure and resource availability, these population genetic approaches are providing powerful arguments for the need for landscape wide approaches in conservation (Nève 2009).

There is, however, controversy over the relative importance of genetics in the decline of species and the need to address it in conservation plans. For small populations, it is expected that demographic (and environmental) stochasticity will both override any underlying genetic influence and have more immediate effect on population dynamics (Lande 1988). However, it is undeniable that in small, isolated populations there is likely to be a decrease in local genetic variability and an increase in genetic differentiation between populations due to genetic drift, reduced gene flow and inbreeding, which can all heighten extinction risk (Saccheri *et al.* 1998), and at the very least reduce the adaptive capabilities of populations to environmental and demographic change (Frankham *et al.* 2002; Young *et al.* 1996). Very small populations may also experience the problem of the Allee effect, whereby population growth rate decreases

due to a deficit of mates, foraging efficiency, and protection against predators (e.g. Kuussaari *et al.* 1998). Coupled with the most dispersive individuals emigrating, this may result in the evolution of reduced mobility in these populations.

Dispersal is fundamental to metapopulation persistence (Clobert *et al.* 2004; Hanski 1999; Singer and Ilkka 2004), such that the inability of individuals to disperse across a growing matrix of unsuitable habitat to locate good quality breeding areas may exacerbate already severe declines. Understanding the functional genetic variation underlying such traits is critical for species of conservation concern. Molecular research on a candidate dispersal gene, and potential adaptive molecular marker for arthropods (a gene in which variation predicts fitness in populations) in the Glanville fritillary, *Melitaea cinxia*, reported the first evidence of single gene effects on population dynamics, challenging the perspective of Lande (1988) and others, that genetic influence is dismissible. Hanski and Saccheri (2006) demonstrated that variation in the well-studied glycolytic enzyme phosphoglucose isomerase (*Pgi*) had significant influence on population growth rates with certain alleles favoured in small, isolated populations versus large, well-connected patches. This complimented studies showing *Pgi* to be associated with flight metabolic performance, lifespan and fecundity (Haag *et al.* 2005; Orsini *et al.* 2009; Saastamoinen *et al.* 2009). Such modern approaches are extremely valuable for aiding the prediction of long-term species responses to environmental change and informing conservation strategies.

Complex relationships with other species, such as parasitoid wasps (Bulman 2001; Porter 1981; Shaw *et al.* 2009; van Nouhuys and Hanski 2004) and endosymbiotic bacteria such as the well-studied *Wolbachia* (Charlat *et al.* 2003; Hornett *et al.* 2009; Jiggins *et al.* 2000a; Nice *et al.* 2009; Rokas 2000; Werren *et al.* 2008), should not be ignored in respect to conservation strategies, as they may strongly influence population dynamics and the extinction likelihood of populations. Current intensive management strategies could actually be significantly aiding the spread of pathogens and parasites as corridors and increased connectivity in habitat networks could open up infection routes across the landscape. Both parasitoid wasps and *Wolbachia* have only recently been given the amount of attention in research that is indicative of their importance to lepidopteran ecology and conservation. *Cotesia* species of parasitoid wasp are among the most well-known, and have been the focus of many studies on Checkerspot butterflies (Kankare *et al.* 2005; van Nouhuys and Hanski 2004). Rates of larval

parasitism by *Cotesia* have been shown to decrease with cool but sunny weather conditions in spring, whereby the dark gregarious larvae develop faster than the pale parasitoid cocoons (Porter 1983; Van Nouhuys and Lei 2004), allowing populations to escape from higher levels of parasitism. *Wolbachia* are endosymbiotic bacteria capable of dramatically altering the reproductive system of their host and potentially causing speciation and dramatic population declines due to severe sex-ratio biases (Charlat *et al.* 2007; Dyson and Hurst 2004; Engelstadter 2010; Hornett *et al.* 2009). However, in some populations a suppressor evolves that either prevents transmission or reduces the action of the bacterium (Hornett *et al.* 2006), and which then spreads rapidly through populations (Charlat *et al.* 2007) allowing sex-ratios to balance once more.

The final, but perhaps most pressing of these ‘invisible’ factors of importance to conservation, is the issue of climate change. Not only will climate warming affect the distribution of suitable habitat for thermally sensitive butterflies and cause both latitudinal and altitudinal range shifts (Hill *et al.* 2002; Wilson *et al.* 2007), but it will impact on all of the aforementioned influences on the population dynamics of declining metapopulation species, from interactions with parasitoid wasps to genotype by environment interactions at *Pgi*, to basic habitat preferences of adults and larvae and host plant distributions. In some cases, where species are not able to track the suitable habitat as climate change progresses, intervention may be needed in the form of ‘assisted colonisation’ (Carroll *et al.* 2009) whereby a species is re-introduced into an area deemed suitable climatically for the foreseeable future.

### 1.3.3 Captive breeding and reintroductions

In extreme cases of local extinction, current re-introduction schemes seem hopeful and show steady increases in butterfly numbers and patch colonisations, such as the hugely successful re-introductions of *M. arion* (Thomas *et al.* 2009) and *E. aurinia* (Porter 2007; Porter and Ellis 2011). However, the long term success of these ‘fake’ populations is unknown. Will they eventually regulate themselves, or will they need monitoring and manipulating forever more? Is the genetic variation in the source populations for the re-introduction enough? Will it need supplementing in the future or is the available variation sufficient to allow adaptation to occur? Many factors need to be taken into consideration. If captive breeding stock is infected with endosymbiotic bacteria such as *Wolbachia* and used in reintroduction programmes, more harm than good may come of it.

#### 1.4 CRITICAL FUTURE DIRECTIONS FOR CONSERVATION ECOLOGISTS

Modern day conservation of declining species in fragmented landscapes is a multi-level venture. With model projections indicating a dire future for many species in the 21<sup>st</sup> Century (Pereira *et al.* 2010), it is well recognised now that not only is knowledge on behaviour and life history needed, but further underlying traits of populations need to be investigated to successfully conserve species. It would consequently be of great benefit to be able to predict differential success and fitness at the organismal level by studying potential adaptive markers such as *Pgi* at the molecular level, in parallel with neutral markers of population structure. If there are pleiotropic effects of one locus causing individuals of certain genotypes to have increased flight ability at higher temperatures, but to suffer lower fecundity, as found by Watt (1992), there may also be severe consequences for populations in the event of climate change. In terms of conservation, this can help predict how populations will respond to environmental change, and even identify source populations for potential translocations or re-introductions in combination with knowledge gained from neutral markers.

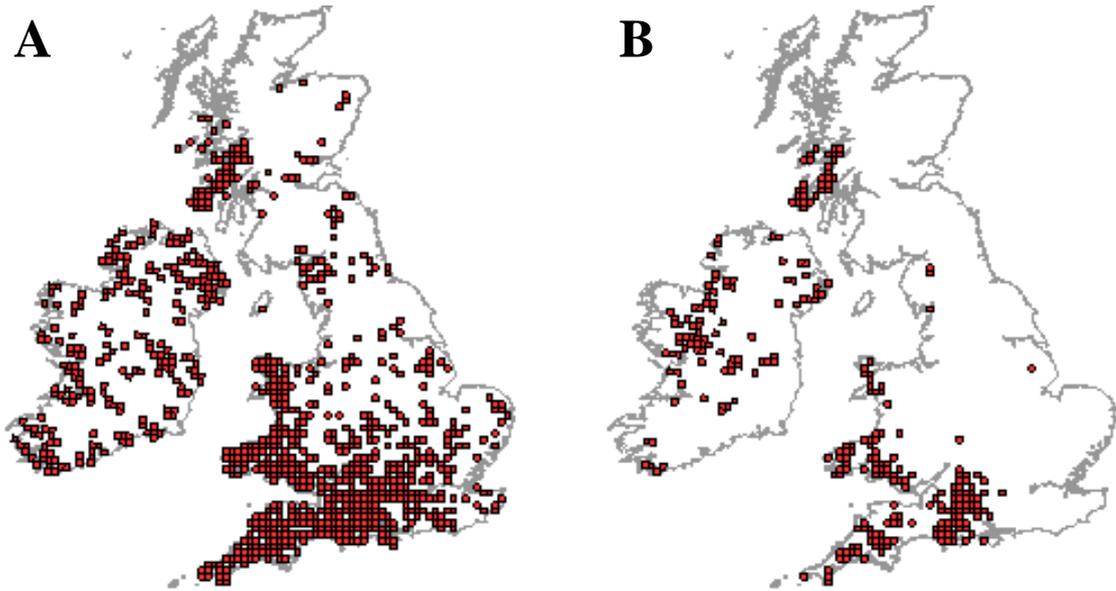
Until recently, genomic resources in Lepidoptera have been lacking, despite the immense biological, societal and economic value of butterflies in particular. Thanks to recent developments of next-generation sequencing technologies (Margulies *et al.* 2005) and the resulting increase in publicly available EST data, more neutral markers with highly variable loci, or extensive sequence data, should make population genetic approaches far more informative as well as allow the identification of adaptive markers (Hedrick 2004; Stapley *et al.* 2010) which may be directly involved in responses to processes such as environmental change (Kohn *et al.* 2006).

In conclusion, the principle conservation objectives for declining species must be, at minimum, the maintenance of existing distributions and genetic diversity, to enable the greatest chance of adaptability to the ensuing environmental changes, and hence, with optimal management, allow the future expansion of populations.

## 1.5 INTRODUCTION TO STUDY SPECIES

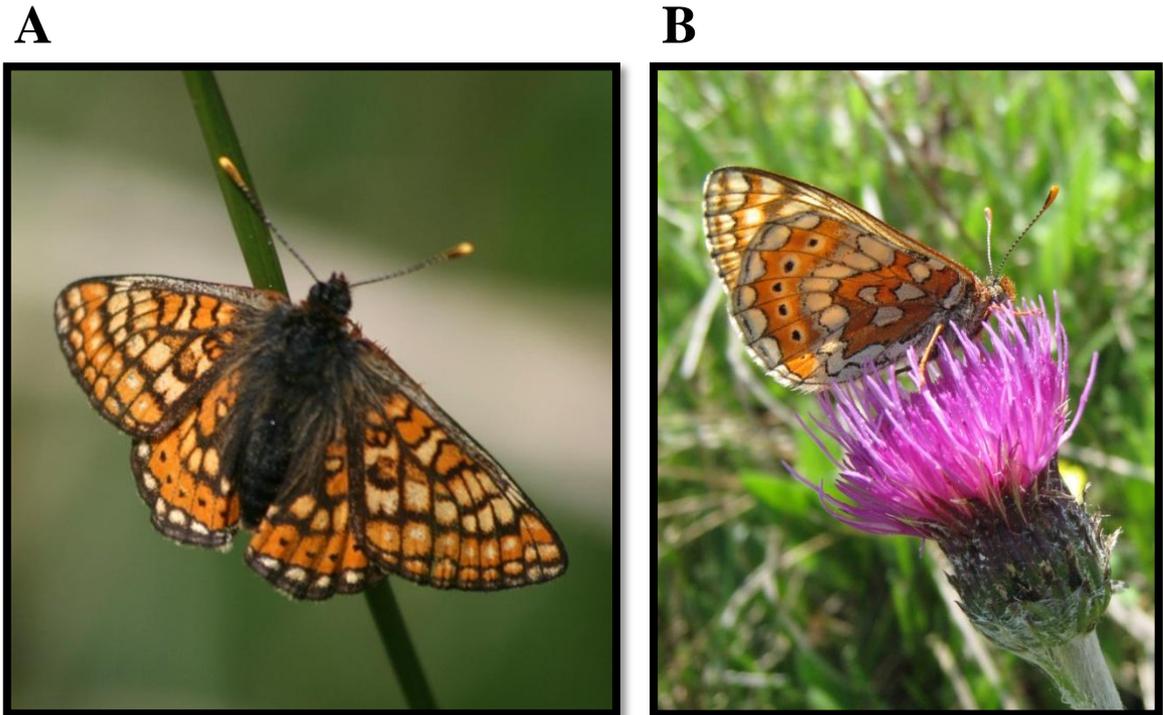
Historically known as the ‘greasy’ fritillary due to its shiny appearance once its bright colours fade, and belonging to the group of butterflies known collectively as the ‘Checkerspots’, the marsh fritillary butterfly, *Euphydryas aurinia*, (Rottemburg, 1775; Lepidoptera: Nymphalidae ) is a well known example of a British butterfly facing rapid declines and local extinction, and hence has become an almost iconic species for conservation in the UK (**Plate 1.1**). The range of *E. aurinia* spans the Palaearctic, from Ireland in western Europe through 38 European countries, Morocco, Algeria and across Russia and temperate Asia to Korea (Tolman and Lewington 2009; van Swaay and Warren 1999). Unfortunately, across much of this distribution *E. aurinia* is declining, with an overall distributional decline of 20-50% over 25 years across Europe and now extinct in one European country where it was once present (van Swaay and Warren 1999), causing it to be not only listed as a protected species in the UK (Wildlife and Countryside Act 1981 and UK Biodiversity Action Plan (BAP) priority species), but also the only British species to be listed under European legislation too (Annexe II of both the EC Habitats and Species Directive and the 1979 Bern Convention).

The UK is currently considered a stronghold for *E. aurinia*, supporting 5 – 15% of the European distribution (van Swaay and Warren 1999) but even here it has seen a long term loss of 73% in abundance over approximately 25 years (Fox *et al.* 2010) which is attributed to changing grazing patterns and habitat fragmentation (Asher *et al.* 2001). The two main biotypes that *E. aurinia* occupies in the UK are damp, neutral (dominated by *Deschampsia cespitosa*) or acidophilous (dominated by *Molinia caerulea*) pastures and dry, calcicolous grasslands (Barnett and Warren 1995). In Devon and Cornwall where *E. aurinia* has large strongholds, only 8% of their favoured acidic *Molinia-Juncus* pastures still remain of that which was present in 1900 (Department of Environment 1995). **Figure 1.2** shows the change in distribution of *E. aurinia* between the period of 1690 – 1994 (A) and more recently 2000- 2004 (B), illustrating the severe contraction of the species towards the south-west UK and western Scotland.



**Figure 1.2** – Plotted by 10km squares of the National Grid, the British distribution of *Euphydryas aurinia* has declined dramatically from the period of 1690 – 1994 (A) to more recent times of 2000 – 2004 (B). Maps courtesy of the National Biodiversity Network (NBN).

*E. aurinia* (**Plate 1.1**) is a univoltine species, only having one generation a year, with the flight period spanning from mid May to late June in southern England, and with males emerging first. Although the larvae will feed on other plants, the sole host plant in the UK is Devil's-bit scabious *Succisa pratensis* Moench (Dipsacaceae). Female *E. aurinia* will usually lay an initial large batch of up to 500 eggs, soon after emergence and successful mating, on the underside of the largest leaves of *S. pratensis* in the natal patch, and subsequent smaller batches on plants further away (Porter 1981). Studies on the habitat requirements of *E. aurinia* in Germany and the Czech Republic have showed that females prefer to lay egg clutches on large, prominent *S. pratensis* plants surrounded by low, cushion-forming vegetation (Anthes *et al.* 2003; Konvicka *et al.* 2003). The eggs are a cream colour when laid, and darken to orange-brown as they develop, until they hatch, within 30 to 40 days later (Porter 1981). The larvae then live gregariously in silken webs from late July, through winter diapause, until spring the following year (Porter 1982; see **Plate 1.2**). Larvae then become solitary in the penultimate instar to feed and eventually final 6<sup>th</sup> instar larvae pupate. Although some individuals have been known to move up to 15-20km (Warren 1994) the adult butterfly is generally sedentary (Junker and Schmitt 2010; Porter 1981).



**Plate 1.1** – Dorsal (A) and ventral (B) views of an adult *Euphydryas aurinia* at Lydlinch, Dorset, 2009. Photos taken by M. R. Smee.

Two common methods of habitat management for *E. aurinia*, which have now become adaptive tools used to create suitable, but usually temporary, habitats (Wahlberg *et al.* 2002) are extensive cattle grazing of local, hardy breeds; and swaling (annual controlled burning of gorse and scrub to thin out the old vegetation in order to allow new grass shoots to grow, thus providing grazing for the livestock). However, responses to grazing are equivocal (Pöyry *et al.* 2005), and burning can cause heavy larval mortality at certain times of the year (Warren 1994) emphasizing the need for more detailed study of the effect of these techniques on important habitat variables for *E. aurinia*.

Populations of *E. aurinia* are known to show large fluctuations in numbers from year to year (Ford and Ford 1930), and are severely affected by adverse weather conditions during the flight period (pers.obs.). Larval parasitoid wasps, *Cotesia bignelli* in southern England, are also known to play a major role in the fluctuations of *E. aurinia* populations and can complete three generations within one host generation (Bulman 2001; Klapwijk 2008; Porter 1983), decimating large proportions of *E. aurinia* populations.



**Plate 1.2** – A web of 4<sup>th</sup> instar *E. aurinia* larvae at Stithians, Cornwall, 2008. Photo taken by M. R. Smee.

## 1.6 SCOPE OF THESIS

The remaining six chapters of this thesis aim to address some of the conservation issues discussed in earlier sections here, in the hope of providing a modern conservation framework for *E. aurinia* and other similar species of concern.

Chapter two provides the context for the remainder of this study, as the severe decline of *E. aurinia* is analysed at a stronghold in south-west England, but has wider applicability to other populations also surviving in *Molinia-Juncus* habitat. Targets were to determine the most important habitat requirements of *E. aurinia*, and whether management could provide these ideal conditions and consequentially increase population numbers to secure a positive future for this species. The management undertaken included both cattle grazing and swaling, and aimed not only to improve the quality of the habitat, but also sought to increase the connectivity between suitable habitat patches.

Chapter three describes the process involved in the time- and cost-efficient development of EST-derived microsatellite markers using a 454 transcriptome of *E. aurinia*. Both the advantages and disadvantages of this method are considered with consideration to the protected status of *E. aurinia* and therefore lack of any previous EST data. These ‘neutral’ molecular markers are then used in chapter four to explore levels of genetic structuring and variation within *E. aurinia* populations across the UK and the Catalonia area in northern Spain and southern France. Methods both with and without the consideration of sampling location are analysed to determine whether similar patterns can be gained from studying genetic structure as an emergent property of the genetic samples alone. The aim, as in most conservation genetics studies, is to more accurately inform conservation strategies for *E. aurinia* both in the UK where it has contracted to the south-west, and in Catalonia – is the same story true for all regions of its distribution? If not – what are the implications for the conservation of this species in these regions? A secondary aim is to clarify, for all studies of population genetic structuring in fragmented populations, the implications of a variety of structural patterns. Should conservation management differ for metapopulations that are mixed, completely isolated, or isolated by distance? How are management decisions influenced by conservation goals: to maintain population size, genetic diversity, genetic integrity or genetic identity?

To compliment the analysis using neutral markers in chapter four, chapter five assesses the variation in the potential adaptive molecular marker for arthropods, phosphoglucose isomerase (*Pgi*). Variation in this gene is analysed across geographically distant samples to determine if balancing selection is evident, as suggested in other species (Wheat *et al.* 2010; Wheat *et al.* 2005) and, if so, whether heterozygote advantage may be the causal factor. Measures of flight ability and propensity, as well as population trend, are also analysed for any influence of genotype.

Chapter six explores the infection frequency of *Wolbachia* across UK samples of *E. aurinia*, and attempts to infer a phenotype from sequence homology with known strains. Implications of the rapid spread of *Wolbachia* across declining populations of butterflies are discussed in light of the potential devastating effects this endosymbiont could have.

A general discussion is provided in chapter seven, where the conservation implications of this work are considered and future studies are recommended.



## Chapter 2

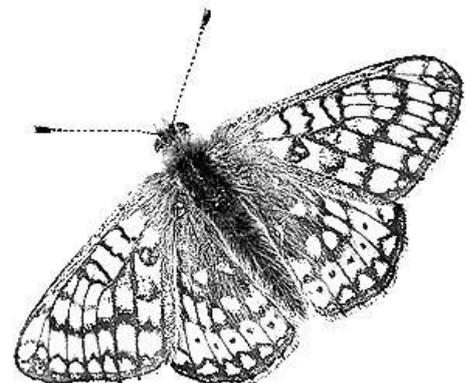
# **Butterflies on the brink: habitat requirements for declining populations of the marsh fritillary (*Euphydryas aurinia*) in SW England.**

PUBLISHED

*Journal of Insect Conservation* **15**: 153 - 163

*'Butterflies in profusion tell us that all is well with nature.'*

*Sir David Attenborough*



## 2.1 ABSTRACT

1. The marsh fritillary *Euphydryas aurinia* is one of our most endangered butterflies, and the only to be protected under European legislation as well as British. It persists in fragile subpopulations threatened by habitat fragmentation and degradation.
2. A combination of swaling and cattle grazing are accepted to be best practice for managing wet, unimproved grasslands – the favoured habitat for *E. aurinia* in Cornwall. These two well-endorsed methods of management were used to increase and improve the quality of habitat for *E. aurinia* over a five year period, 2004-2008, at a stronghold network of habitat patches in mid Cornwall, south-west England.
3. Analyses of adult and larval densities over five years in fifty-four transects across nine sites found *E. aurinia* to favour habitat patches with higher densities of the larval food plant (Devil's-bit scabious *Succisa pratensis*), higher sward height in autumn, and intermediate optimum levels of stock grazing.
4. Main findings indicated most sites experienced significant declines in numbers. Unfavourable weather in the last two years of monitoring was likely to have had a significant impact on the response of individual subpopulations to habitat management though poor recovery rates may also reflect a time-lag in colonisation events after habitat improvement has occurred.
5. Habitat management produced an improvement, albeit an inconsistent improvement in habitat variables across patches – *S. pratensis* shows a clear recovery at some sites. Autumn sward height increased significantly at one site, and a quadratic relationship between stock grazing and important habitat variables has been found which will aid further improvement over all sites for the long term persistence of *E. aurinia*.

## 2.2 INTRODUCTION

The marsh fritillary, *Euphydryas aurinia*, is vulnerable to local extinction from much of its current distribution across the UK and Europe. It is the only British butterfly species to be listed under the EC Habitats and Species Directive as well as the Bern Convention, fully protected under the Wildlife and Countryside Act 1981 and a priority species in the UK Biodiversity Action Plan (BAP). Agricultural improvement or abandonment, and changes in management practices have meant a loss of 92% of south-west England's damp pasture, the main habitat used by *E. aurinia* (Hobson *et al.* 2001). Having suffered serious declines across most of its range in Europe, *E. aurinia* has undergone a 55% loss over a period of 30 years (Asher *et al.* 2001) in the UK – currently considered a stronghold for the species, supporting 5 – 15% of the European distribution (van Swaay and Warren 1999). This dramatic decline is mirrored in Cornwall, with 62% of known populations going extinct between 1990 and 2000 (Hobson and Budd 2001). These trends have caused the butterfly to contract to a handful of core geographical areas (Asher *et al.* 2001), one of these being a set of subpopulations (temporary populations as part of a metapopulation) found across the mid-Cornwall Moors.

Metapopulation theory states that persistence in a fragmented landscape requires a balance between stochastic extinctions and re-colonisations (Hanski 1994; 1998; 1999; Hanski and Gaggiotti 2004; Levins 1969). Processes of disturbance and succession cause natural patch networks to be dynamic, with regional extinction being prevented through the process of habitat tracking, whereby a species is able to follow the shifting mosaic of suitable environmental conditions (Thomas 1994). In an anthropogenically modified landscape, restoration or creation of new habitat can allow such habitat tracking to occur. As a species widely accepted to persist as metapopulations in fragmented landscapes (Hula *et al.* 2004; Wahlberg *et al.* 2002; Warren 1994) *E. aurinia* thereby requires restoration of habitat through appropriate management at a landscape-scale, to reduce the probability of subpopulation extinction, and to promote connectivity among suitable habitat patches.

*E. aurinia* occupies two main biotypes in the UK: damp, neutral or acidophilous grasslands, and dry, calcicolous grasslands (Barnett and Warren 1995). Consequently, 'appropriate' management is not the same for both and has to be assessed on a 'per site' basis. Traditional farming methods benefitted the species by creating heterogeneous

landscapes, providing open, low vegetation for basking larvae (Porter 1982), as well as tall nectar-rich flowering herbs for adults on the wing – a combination some rural areas can still provide (Liu *et al.* 2006). However, the advent of modern farming techniques and intensive grazing of cattle in western Europe have caused habitat to become too homogeneous and unsuitable for *E. aurinia* (van Swaay and Warren 1999). Two common methods, which have now become adaptive tools used to create suitable, but usually temporary, habitats (Wahlberg *et al.* 2002) are extensive cattle grazing and swaling (controlled burning). However, responses to grazing are equivocal (Pöyry *et al.* 2005), and burning can cause heavy mortality to *E. aurinia* larvae at certain times of the year (Warren 1994) emphasizing the need for more detailed study of the effect of these techniques on important habitat variables for *E. aurinia*. Studies on the habitat requirements of *E. aurinia* in Germany and the Czech Republic have shown that females prefer to lay egg clutches on large, prominent *S. pratensis* plants surrounded by low, cushion-forming vegetation (Anthes *et al.* 2003; Konvicka *et al.* 2003).

The aim of this study was to determine the most important habitat requirements of *E. aurinia* on the mid-Cornwall Moors, and whether management can achieve this and consequentially increase population numbers to secure a positive future for this species. Our *a priori* expectations include a significant positive effect of the density of the larval host plant *S. pratensis* and an effect of sward height on occurrence of *E. aurinia*, as these have been shown to be important in previous studies (Betzholtz *et al.* 2007; Bulman 2001; Hula *et al.* 2004; Konvicka *et al.* 2003). A medium, optimum level of sward height is likely to be important for *E. aurinia*. Too short and *S. pratensis* may become too small or isolated and cause higher mortality of larval groups through starvation or predation, but too tall and rank and *S. pratensis* may become too difficult for adult butterflies to find or be outcompeted and lost altogether (Konvicka *et al.* 2003). Either extreme of sward height could also create unfavourable microclimatic conditions for both larval and adult stage. We also hypothesised that an important quadratic relationship may exist between intensity of management and grass sward height, as over-grazing can result in very short, unsuitable host plants and areas of bare ground, whereas under-grazing and abandonment can lead to rank or scrubby, unsuitable habitat. The management undertaken aimed not only to improve the quality of the habitat, but also sought to increase the connectivity between suitable habitat patches. Here we report on the successes and failures of these management techniques, and the consequent trajectory of the mid-Cornwall Moors metapopulation. We predicted that habitat

management would improve key predictors of *E. aurinia* density with the aim of a subsequent network-wide increase in *E. aurinia* densities. However this prediction should be considered against the cyclic nature of *E. aurinia* subpopulations (Ford and Ford 1930) whereby a peak in numbers was likely attained in 2006, followed by lower numbers in more recent summers of 2007 and 2008, alongside substantial declines in many other UK butterfly species in the same time period (Botham *et al.* 2008).

## 2.3 MATERIALS AND METHODS

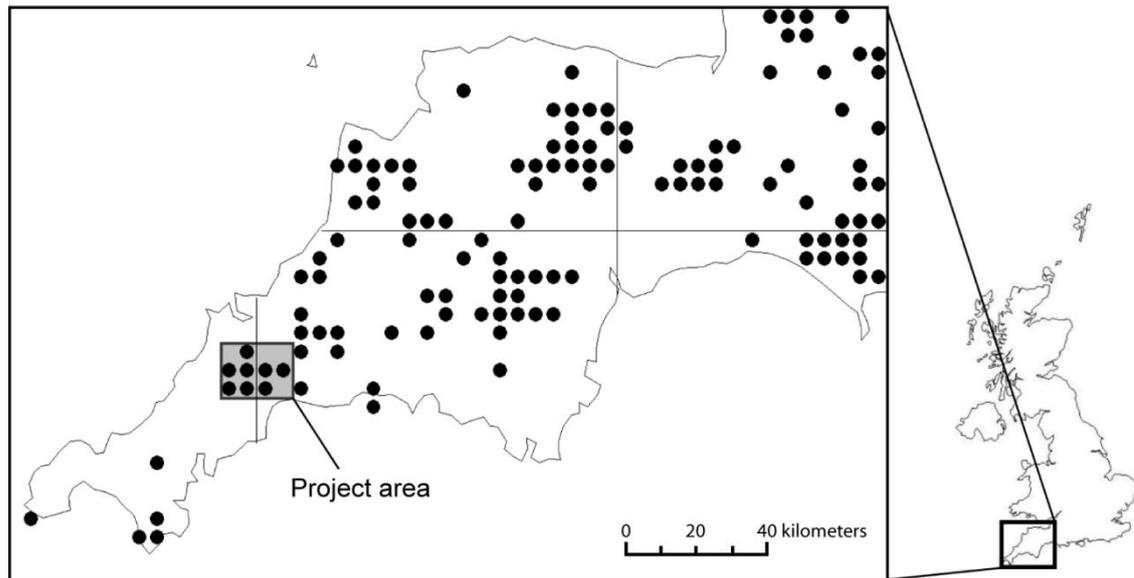
### 2.3.1 Study species

*E. aurinia* (Lepidoptera; Nymphalidae) is a univoltine species with its flight period from late May until early July in the study area. Although some individuals have been known to move up to 15-20km (Warren 1994) the adult butterfly is generally sedentary (Junker and Schmitt 2010; Porter 1981). Females usually lay an initial large batch of c. 300 eggs soon after emergence in the natal patch, and smaller subsequent batches on plants further away (Porter 1981). Eggs are laid on the underside of leaves of the sole host plant used in this area, Devil's-bit scabious *S. pratensis* Moench (Dipsacaceae). The larvae live gregariously in silken webs from late July, through winter diapause, until spring the following year (Porter 1982). Larvae then become solitary in the penultimate instar to feed and eventually final 6<sup>th</sup> instar larvae pupate.

### 2.3.2 Study area

The study area was located in south-west England, encompassing an area known as the mid-Cornwall Moors (**Figure 2.1**). Previous surveys showed an extant set of subpopulations of *E. aurinia* to be breeding across areas of Goss Moor National Nature Reserve (NNR) (SW 9560) and Breney Common Cornwall Wildlife Trust Reserve (SX 0561). These areas are a wet mosaic of unimproved purple moor grass (*Molinia caerulea*) pasture, heath and scrub, and are a Designated Special Area of Conservation (SAC), partly for the *E. aurinia* subpopulations they support. The project was a partnership with several organisations (Butterfly Conservation, Natural England, Cornwall Wildlife Trust, Highways Agency and Environment Agency) supported by the EU LIFE Natura fund. Prior to the establishment of the LIFE Project a survey was carried out to find other suitable habitat in the surrounding areas. Transect areas defined as 'suitable' for the study were open, unimproved grassland with the host plant present (Wahlberg *et al.* 2002), or areas that showed the potential to become suitable with

management. In addition to the sites at Goss Moor and Breney Common a further seven sites were selected for inclusion in the LIFE Project area, making a total of nine sites encompassing fifty-four individual transect areas over a total area of 1050 ha, for inclusion in a five-year monitoring programme.



**Figure 2.1** - South-west England, showing the project area across the Mid-Cornwall Moors. It is the most westerly stronghold for the species, as they have already declined from more coastal areas in Cornwall and only a couple of diminishing small colonies remain otherwise.

Distribution map courtesy of Butterfly Conservation.

### 2.3.3 Field Techniques

Two field visits were made each year; the first to survey adult butterflies in spring (late May through June); and the second to survey pre-hibernating larval webs in autumn (late August through September).

#### 2.3.3.1 Adult surveys

Adult monitoring surveys were carried out using the standard Butterfly Conservation methodology (Tunmore *et al.* 2008). Across each transect area a zig-zag path was walked, covering both suitable and unsuitable areas of habitat, recording numbers of adult *E. aurinia* and the time taken to complete the survey (minutes). Two field visits were made to each transect area ensuring the peak flight period was covered. Using the flight area of each transect (Ha), relative adult density was calculated for each visit to each transect area, and an average of the two relative densities was taken and used in the analyses. Because of poor weather, some transects were not repeated in the final year of surveys, so just the single value was used where this occurred. Sites were visited

between 10:30hrs and 16:00 hrs on days of suitable weather (minimum of 13°C with at least 60% sun or a minimum of 17°C with at least 40% sun, and little or no wind) (Tunmore *et al.* 2008).

### 2.3.3.2 Larval web surveys

Most larval web counts were carried out using the standard Butterfly Conservation transect methodology; which involves walking a zig-zag transect across the patch of habitat recording numbers of occupied webs seen 1 metre either side of the transect line (Tunmore *et al.* 2008). The length (metres) of the transect was also measured using a pedometer and the time taken to walk the transect (minutes) was recorded. As part of another continuing monitoring programme, more larval web counts were undertaken by a second field surveyor using area searches of fixed plots. During two years of the current study, a number of comparative surveys (both surveyors conducting larval web counts of the same transect area but using different methodologies) were carried out in suitable areas (where larval webs had been found) in order to calculate a 'conversion factor' to convert the number of webs found during area searches of fixed plots to a transect equivalent. Regression analysis gave the best conversion factor (Transect based search = 3.4236\*Area search of fixed plot) to allow statistical analysis using both sets of data.

### 2.3.3.3 Habitat condition monitoring

Several habitat variables were measured on larval web transects to assess the habitat quality in which larval webs were found. Sward height (cm) was measured using a drop disk at thirty random points along each transect (Stewart *et al.* 2001). We also recorded the presence/absence of *S. pratensis*, *M. caerulea* tussocks, invading scrub and evidence of stock grazing at each of these thirty points (Brereton *et al.* 2005). The percentage of points where each habitat variable was present, out of the thirty measured, gave a proxy of density for *S. pratensis*, *M. caerulea* tussocks and invading scrub and a measure of intensity of stock grazing over each transect area. Three measures of habitat management were used in analyses: the years in which grazing was commenced on each site, the number of swaling events over the five year management periods, and the number of years since the most recent swaling event.

## 2.3.4 Data Analysis

To investigate the relationship between adult or larval densities and habitat variables, we used a restricted dataset of only those transect areas with at least one occurrence of either adults or webs over the five year period (thirty-three out of fifty-four transects) as

we were interested in effects on density not presence/absence. The response variables in our analyses were adult densities and larval web densities. The explanatory variables were intensity of stock grazing (% of transect points that showed evidence of grazing), density of host plant (% of transect points where *S. pratensis* was observed), scrub density (% of points), *M. caerulea* density (% of points), and spring and autumn sward height (in cm). We included quadratic terms for all fixed effect explanatory variables due to our *a priori* expectations of some humped relationships. Modelling the impacts of habitat variables on *E. aurinia* densities was performed in two phases. First, we used model simplification of multiple regressions, to account for covariance among explanatory variables and to clarify the habitat variables that best predicted density. Second, we analysed each habitat variable in separate regressions, to highlight individual links between habitat variables and *E. aurinia*.

In analyses of the effects of management on habitat variables we used the entire dataset of all fifty-four transect areas, as we were interested in whether management could improve the habitat variables of importance across all sites, to allow potential re-colonisation events. To determine if there had been a positive response to the management, temporal trends of habitat variables were analysed as response variables, with year as a fixed effect explanatory variable, at an individual site level, as often this is more informative for individual landowners, and an overall average can be misleading. Temporal trends of adult and web densities were also analysed at both individual sites and across all nine sites to give an overall trend.

All data analysis was carried out using R version 2.9.0 (R development core team 2010). Linear mixed effects models were used in all analyses and we accounted for spatial structuring and repeated measures by modelling survey area, nested within site, as random effects on intercepts. Year was modelled as a random slope. Response variables (adult or web density) were log-transformed to reduce skew and improve homogeneity of variance in residuals, and models were checked for homoscedasticity of residuals and Gaussian errors. We used likelihood ratio tests and simplified maximum likelihood versions of the mixed effects models to test for significance of fixed effect explanatory variables at the  $p < 0.05$  level (Crawley 2007).

## 2.4 RESULTS

Of the nine sites selected for inclusion in the study throughout the five year period, only five had extant populations of *E. aurinia*, and only thirty-three of the fifty-four designated transects had at least one occurrence of either adults or webs over the period.

### 2.4.1 Can habitat variables explain the presence of *E. aurinia*?

As expected, a higher density of the host plant *S. pratensis* significantly increased the number of webs found, as did higher sward height in autumn (**Table 2.1**). Single term regressions (**Table 2.2**) revealed a humped relationship between web density and level of stock grazing: above an optimum level, grazing intensity became detrimental (**Figure 2.2**). This relationship became non-significant when considered in the same model as host plant density and sward height, implying that grazing effects are indirect: it seems likely that optimal grazing can maximise sward height and host plant density, rather than affecting webs directly. In single term regressions, larval web densities were still positively influenced by the height of autumn sward. However, without accounting for the effect of the other habitat variables, the influence of *S. pratensis* was lost ( $\chi^2_1 = 2.74$ ,  $p = 0.098$ ). There were no significant habitat variables associated with adult butterfly densities.

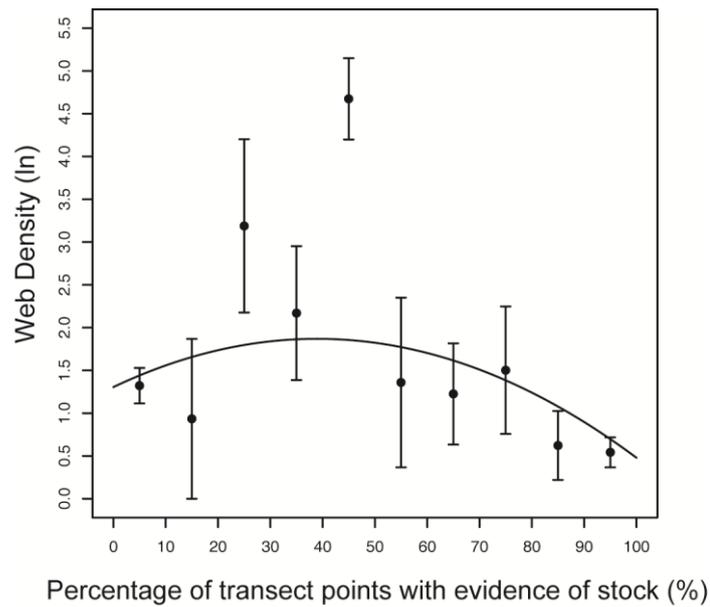
Although none of the measured habitat variables had a significant effect on the densities of adult butterflies seen, the number of larval webs the previous year was found to be a strong predictor of the densities of adult butterflies seen in the current year (linear coefficient 0.01,  $\chi^2_1 = 12.59$ ,  $p < 0.001$ ). Adult butterfly densities in the current year also predict numbers of larval webs in the current year (linear coefficient 0.033,  $\chi^2_1 = 23.08$ ,  $p < 0.001$ ), so that management for one life stage also benefits the other.

**Table 2.1** - Results of a linear mixed effects model using **multiple** regression to relate habitat variables with larval web densities and adult butterfly densities. Models used data only from occupied transect areas.

	Coefficient	$X^2_1$	$p$
<b>Webs</b>			
<b>Significant terms</b>			
Autumn sward height (cm)	0.041	10.82	<0.001
<i>S. pratensis</i> density (%)	0.015	4.58	0.032
<b>Non-significant terms</b>			
Quadratic autumn sward height (cm)		1.18	0.28
Spring sward height (cm)		0.19	0.66
Quadratic spring sward height (cm)		0.21	0.65
Quadratic <i>S. pratensis</i> density (%)		0.16	0.69
<i>M. caerulea</i> density (%)		3.10	0.08
Quadratic <i>M. caerulea</i> density (%)		1.18	0.28
Scrub density (%)		0.30	0.58
Quadratic scrub density (%)		1.84	0.18
Intensity of stock grazing (%)		1.60	0.21
Quadratic intensity of stock grazing (%)		2.98	0.08
<b>Adults</b>			
Autumn sward height (cm)		0.06	0.80
Quadratic autumn sward height (cm)		0.06	0.81
Spring sward height (cm)		2.45	0.12
Quadratic spring sward height (cm)		0.31	0.58
<i>S. pratensis</i> density (%)		1.79	0.18
Quadratic <i>S. pratensis</i> density (%)		0.01	0.91
<i>M. caerulea</i> density (%)		2.62	0.11
Quadratic <i>M. caerulea</i> density (%)		2.56	0.11
Scrub density (%)		0.61	0.43
Quadratic scrub density (%)		0.75	0.39
Intensity of stock grazing (%)		1.99	0.16
Quadratic intensity of stock grazing (%)		0.11	0.74

**Table 2.2** - Results of linear mixed effects models using **single** regressions to determine the gross influence of each habitat variable on density of larval webs and adult butterflies. Models used data only from occupied transect areas.

	Coefficient	$X^2_1$	p
<b>Webs</b>			
<b>Significant terms</b>			
Autumn sward height (cm)	0.0380	8.97	0.003
Intensity of stock grazing (%)	0.0290	3.91	0.048
Quadratic intensity of stock grazing (%)	-0.0004	6.02	0.014
<b>Non-significant terms</b>			
Quadratic autumn sward height (cm)		3.49	0.06
Spring sward height (cm)		2.42	0.12
Quadratic spring sward height (cm)		1.73	0.19
<i>S. pratensis</i> density (%)		2.74	0.10
Quadratic <i>S. pratensis</i> density (%)		0.65	0.42
<i>M. caerulea</i> density (%)		0.14	0.71
Quadratic <i>M. caerulea</i> density (%)		0.67	0.41
Scrub density (%)		0.01	0.94
Quadratic scrub density (%)		2.90	0.09
<b>Adults</b>			
Autumn sward height (cm)		0.59	0.44
Quadratic autumn sward height (cm)		0.00	0.99
Spring sward height (cm)		2.65	0.10
Quadratic spring sward height (cm)		0.42	0.52
<i>S. pratensis</i> density (%)		0.52	0.47
Quadratic <i>S. pratensis</i> density (%)		0.11	0.75
<i>M. caerulea</i> density (%)		1.80	0.18
Quadratic <i>M. caerulea</i> density (%)		1.60	0.21
Scrub density (%)		0.01	0.91
Quadratic scrub density (%)		0.89	0.35
Intensity of stock grazing (%)		2.30	0.13
Quadratic intensity of stock grazing (%)		0.14	0.71

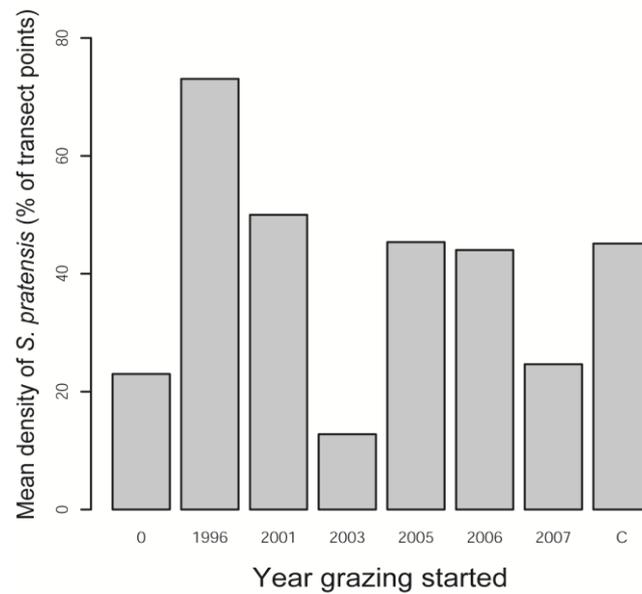


**Figure 2.2** - The significant quadratic effect of stock grazing on density of larval webs. Larval web density (ln) is averaged across all transects, +/- standard error, for each ten percent 'bin' of stock grazing intensity.

#### 2.4.2 Has management achieved optimum levels of habitat variables favoured by *E. aurinia*?

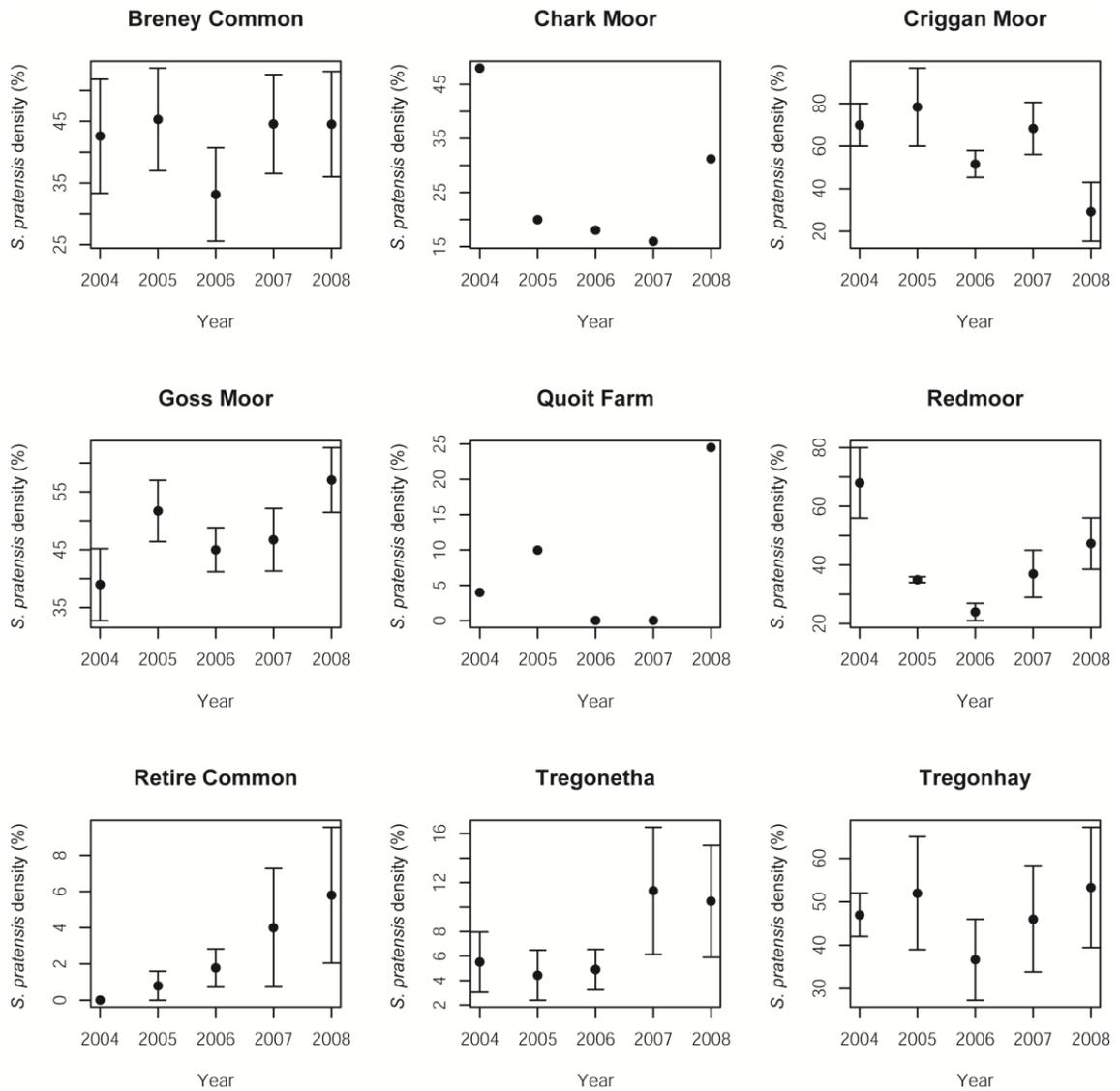
Three measures of management were analysed: the year grazing started; the number of times a transect area was burnt during the five year project; and how many years since the most recent burning of a transect area. The evidence of stock grazing, measured over each transect area, was also included in analyses as an indicator of the intensity of grazing over each area.

Over all five years of the project and over all nine sites whether occupied by *E. aurinia* or not, the year in which grazing started was the only management measure to affect the density of *S. pratensis* ( $\chi^2_1 = 15.29$ ,  $p = 0.033$ ) with less host plant available the more recently grazing had been started and if the area hadn't been grazed at all. In areas where grazing had occurred for longer (continuous grazing and grazing since 1996 and 2001) the density of the host plant was highest (Figure 2.3).



**Figure 2.3** - The year when grazing was started has a significant effect on the occurrence of *S. pratensis* over the transects ( $\chi^2_1 = 15.29$ ,  $p = 0.033$ ) with less host plant being available for *E. aurinia* the more recently grazing had started or if it was not started at all. 0 = never grazed, C = continuous grazing.

Unfortunately the host plant *S. pratensis* did not increase significantly over the five year period at all nine sites despite management ( $\chi^2_1 = 2.83$ ,  $p = 0.09$ ). However, when considering sites individually, which is often more informative, there are opposing significant results. As **Figure 2.4** illustrates, at three of the sites there was a considerable increase in *S. pratensis* over the five years (Goss Moor: linear coefficient 2.51,  $\chi^2_1 = 7.71$ ,  $p = 0.006$ ; Retire Common: linear coefficient 1.395,  $\chi^2_1 = 4.65$ ,  $p = 0.031$ ; and Tregonetha: linear coefficient 1.681,  $\chi^2_1 = 3.86$ ,  $p = 0.049$ ). At one site, Criggan Moor, there was a vast decline in the density of *S. pratensis* found (linear coefficient -8.263,  $\chi^2_1 = 5.21$ ,  $p = 0.022$ ) possibly due to under grazing as the areas are very wet and difficult for cattle to access.

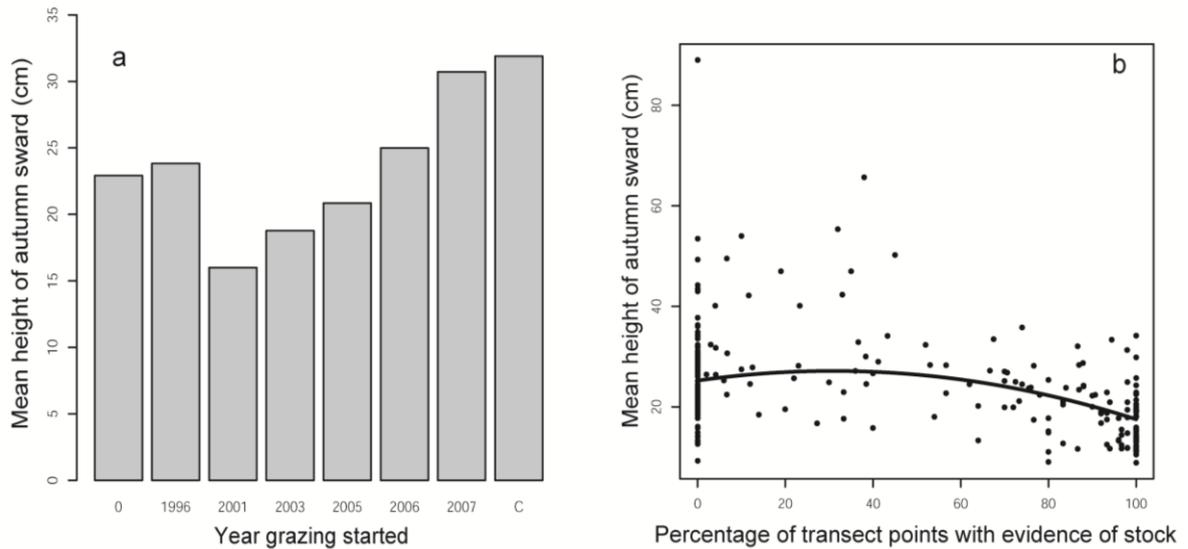


**Figure 2.4** - Temporal comparison of *S. pratensis* density for each site (% of transect points where *S. pratensis* was present) over the five years of the project. +/- standard errors of mean *S. pratensis* density calculated across transect areas within sites. Analysis across all nine sites showed no significant change in *S. pratensis* density ( $\chi^2_{1} = 2.83, p = 0.09$ ).

There was no evidence for either increasing or decreasing trends in the intensity of stock grazing across all nine sites over all five years ( $\chi^2_1 = 1.55$ ,  $p = 0.21$ ). Eight out of the nine sites appeared to be diverging from the optimal level of stock grazing (determined from quadratic relationship = stock grazing at 42% of stop points along the transect). The one site that had moved towards the optimum was Goss Moor (linear coefficient = 7.99  $\chi^2_1 = 8.41$ ,  $p = 0.004$ ), with 50.8% of stop points having evidence of stock grazing where originally in 2004 it had 14.3%. Other sites significantly decreased (Brenay Common: linear coefficient = -21.03  $\chi^2_1 = 26.61$ ,  $p < 0.001$ ; Redmoor: linear coefficient = -28.42  $\chi^2_1 = 17.41$ ,  $p < 0.001$ ) towards 0% and one (Quoit Farm: linear coefficient = 27.14  $\chi^2_1 = 14.74$ ,  $p < 0.001$ ) increased, almost reaching 100%.

In an analysis of habitat management variables (when grazing started, years since the last burn and evidence of stock grazing), two of the management variables had a significant impact on the height of autumn sward over the fifty-four transect areas. The more recently grazing had been started, the higher the average sward height at sites ( $\chi^2_1 = 14.95$ ,  $p = 0.037$ , **Figure 2.5(a)**), although if started in 1996 the average was slightly higher and similar to at sites where grazing had never occurred. Highest autumn sward height is found at sites with continuous grazing occurring, which could be a result of low-level extensive grazing rather than intensive grazing in other areas. This agrees with the relationship between the evidence of stock found along transect walks and autumn sward height (linear coefficient = 0.006,  $\chi^2_1 = 8.43$ ,  $p = 0.004$ ; quadratic coefficient = -0.0001,  $\chi^2_1 = 17.64$ ,  $p < 0.001$ , **Figure 2.5(b)**) whereby higher levels of stock grazing decrease the average autumn sward height, but at a medium level, an optimum sward height was achieved.

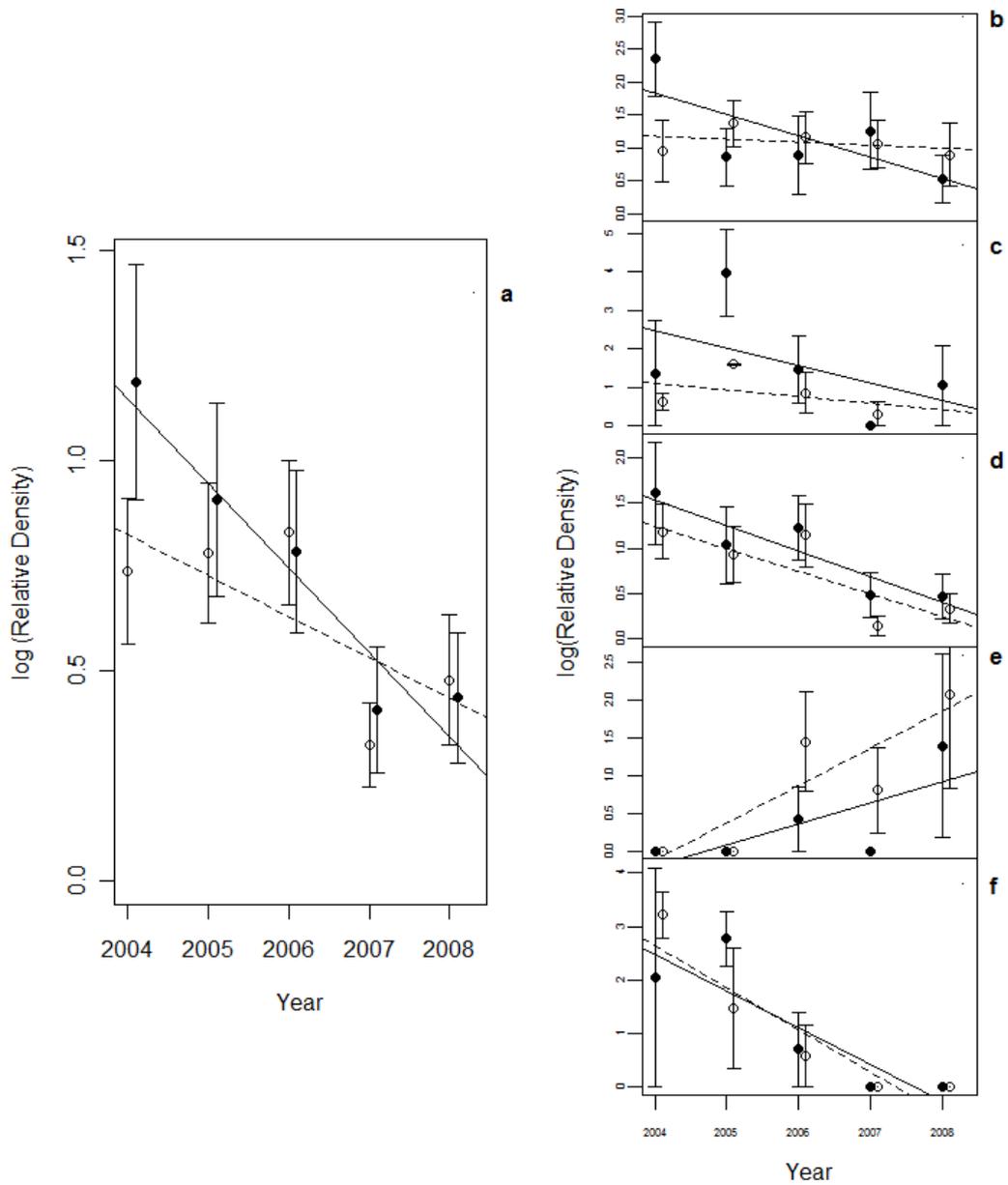
As described in both multiple regression models (**Table 2.1**) and single regression models (**Table 2.2**), numbers of webs increase with higher sward height in autumn. Over all nine sites there was a significant decrease in autumn sward height (linear coefficient -1.49,  $\chi^2_1 = 19.20$ ,  $p < 0.001$ ), although at one site, Redmoor, there was a positive increase (linear coefficient 1.87,  $\chi^2_1 = 7.50$ ,  $p = 0.006$ ).



**Figure 2.5** - In an analysis using data from all fifty-four transect areas, autumn sward height is (a) significantly affected by the year in which grazing started (C = continuous grazing; 0 = never grazed), and (b) the intensity of grazing it is subjected to.

#### 2.4.2.1 Has *E. aurinia* responded to management across the mid-Cornwall Moors?

**Figure 2.6(a)** clearly shows the dramatic decline in numbers of larval webs found (linear coefficient  $-0.32$ ,  $\chi^2_1 = 15.45$ ,  $p < 0.001$ ) and adult butterflies seen (linear coefficient  $-0.19$ ,  $\chi^2_1 = 7.83$ ,  $p = 0.005$ ) over the entire study site despite management, at those transects occupied at least once during the study. However, during the project there was a colonisation event at Redmoor (**Figure 2.6(e)**) and consequently there was an increase in web density (linear coefficient  $0.321$ ,  $\chi^2_1 = 1.94$ ,  $p = 0.164$ ) and adult density (linear coefficient  $0.776$ ,  $\chi^2_1 = 6.49$ ,  $p = 0.011$ ), but only the latter significant.



**Figure 2.6** - (a) Temporal comparison of log transformed larval web densities (solid circles and solid line) and adult densities (open circles and dashed line) across the five years of monitoring at all sites. (b-f) Individual temporal trends of web densities at only those sites where *E. aurinia* has an extant colony: b = Breney Common; c = Criggan Moor; d = Goss Moor; e = Redmoor; f = Tregonhay.

## 2.5 DISCUSSION

Across the thirty-three transects occupied at least once during the study period, larval webs preferred areas of habitat with higher density of their host plant *S. pratensis*, in combination with a suitably tall sward height in autumn and optimum levels of stock grazing. However, across the entire nine sites there could be a possible conflict of interest between achieving the desired levels of *S. pratensis* and autumn sward, as more recent grazing seems to increase autumn sward height but lower the density of *S. pratensis*. Being an inferior competitor (Buhler and Schmid 2001), *S. pratensis* may be getting outcompeted by taller vegetation such as *M. caerulea* by reducing access to light and nutrients, and reducing opportunities for germination. Caution should be taken when interpreting results of sward height, as it is intuitively unlikely that an ever-increasing height of sward would continue to have positive effects on numbers of larval webs. Previous studies have also shown an optimum level of sward over which the likelihood of *E. aurinia* presence decreases (Konvicka *et al.* 2003). A lack of data, or the short period over which the study was done, may mean that a quadratic relationship between larval webs and autumn sward height was not found. It should be noted that ‘continuous’ levels of grazing produce both desired outcomes, and without knowing the particular stocking densities could be inferred as low-level extensive grazing over prolonged periods. This compliments the quadratic relationship found between evidence of stock grazing and both numbers of larval webs and autumn sward height – both responding favourably to optimum medium levels of stock grazing. Although densities of adult *E. aurinia* were not affected by the habitat variables measured, they were a direct result of the density of webs the previous year, and so by managing for one the other will also benefit. There will undoubtedly be other habitat variables that were not measured in the current study, but which will be of huge importance for the presence of adult *E. aurinia*, such as availability or density of nectar sources.

### 2.5.1 Using management tools to create suitable habitat.

Due to differences between the needs of adults and larvae, it is often difficult to determine the ‘optimal’ level of habitat variables needed for the species as a whole. Previously, much work to decipher what constitutes as ‘suitable habitat’ and ‘favourable condition’ has been done on the basis of the adult butterfly’s preferences (Betzholtz *et al.* 2007), but this is often not the most useful measure. The density of the host plant is irrelevant for adult butterflies’ survival as they feed opportunistically upon the nectar of

other flowering plants (Anthes *et al.* 2003), as *S. pratensis* does not flower during the flight period. Here we presented both larval web counts and adult surveys, as it is still crucial to maintain a favourable habitat for the adult stage. However, it appears the adults rely on none of the habitat variables measured, but densities are directly related to the number of webs the previous year. Consequently, if the habitat suits the requirements of the larvae, it will also have a positive, if indirect, effect on adult butterfly numbers.

Intermediate levels of grazing were seen to maximise autumn sward height, perhaps to an optimum level, and density of *S. pratensis* – both important for larval webs, but neither showing significant improvements over all sites. Although Lewis and Hurford (1997) found no significant effect of management on population size, Peet *et al.* (1999) found that plant species richness was higher in managed plots, which were also more structurally heterogeneous. Peet *et al.* (1999) therefore suggested that patches could be left unmanaged on a 2-year rotation without significantly affecting the composition of the plant community. Cutting, or mowing, is also used as a management tool on some *E. aurinia* sites and is reported to be similarly suitable (Dolek and Geyer 1997) yet intrinsically difficult on damp grasslands due to uneven terrain. Swaling can open up vegetation that was previously too dense, but eventually leads to a reduction in *S. pratensis* abundance by encouraging growth of *M. caerulea*. With the lack of swaling influencing any habitat variables on the mid-Cornwall Moor sites, it seems the key for management of these wet grasslands is grazing at an intermediate level so as not to allow over-grazing but not leave sites abandoned. Achieving this optimum is difficult as it is not easy to arrange suitable grazing regimes on damp grassland sites of low agricultural value.

### **2.5.2 Response of the butterfly – are we doing enough?**

Bulman *et al.* (2007) demonstrated that to achieve population persistence in the long term more than 70 ha of suitable breeding habitat is needed to be in favourable condition. 1050 ha of land was encompassed by the nine mid-Cornwall Moors project sites, but not all was in the right condition. In 2004 it was estimated from digitised National Vegetation Classification (NVC) maps that there was 315.8 ha of potential habitat across the nine sites. At the start of the project 15 ha of this was thought to be in favourable condition. By the end of the project, this figure had increased by 115 ha to 130 ha in 2008. Although there seems to be enough suitable habitat to support the

continued functioning of the metapopulation, numbers of *E. aurinia* cast doubt on the ability of the metapopulations to survive over the longer term.

It is possible that, regardless of the management and potentially improved habitat, *E. aurinia* simply cannot respond because of other limiting factors. Ford & Ford (1930) reported huge fluctuations in numbers of *E. aurinia* across the years, from swarms covering hedgerows to dwindling numbers close to extinction. Such large fluctuations may be attributed to unmanageable climatic fluctuations, or to top-down effects of natural enemies, especially parasitoids (Porter 1983).

It is recommended to use larval web counts as the most reliable indicator of population size (Lewis and Hurford 1997), but our results show a steep significant decline in both adult and larval densities across all nine sites of the mid-Cornwall Moors, except at Redmoor where there was an apparent colonisation event during the lifetime of the project. The increase in webs and adults at this site was not found to be significant, but if the population continues to grow over the next generation or two, it will be a huge success for the project.

It is possible that the overall trend presented here is not wholly correct, as the last two years of monitoring experienced unusually high levels of rainfall and poor weather during the flight period. Recent densities may have been low for this reason alone, and good weather for the next couple of generations could see a massive improvement due to better condition of the habitat.

The significant decline could also be a result of severe inbreeding (Saccheri *et al.* 1998). While there are ad hoc records of wandering adults outside of the known breeding subpopulations it is not known whether the mid-Cornwall Moors populations interperse with each other, and hence whether the metapopulation is still functioning and fulfilling the four necessary conditions for persistence (Hanski 1999; Hanski *et al.* 1995). If it is not, there is likely to be a significant increase in homozygosity, or loss of genetic variation. Hanski *et al.* (1994) estimated that approximately 80% of *Melitaea cinxia* butterflies spend their entire lifetime in the natal patch – a pattern that is likely to be mirrored in *E. aurinia* and will undoubtedly increase the probability of inbreeding depression. Recent studies (Joyce and Pullin 2003; Sigaard *et al.* 2008) have indicated restrictions of gene flow within *E. aurinia* metapopulations and emphasize the need for

networks of suitable habitat. Increased levels of emigration as a mechanism of the ‘Allee effect’ (Kuussaari *et al.* 1998) may also be exacerbating the problems of small populations by reducing numbers further, or we may be witnessing the payback of the mid-Cornwall Moors’ ‘extinction debt’ (Bulman *et al.* 2007; Tilman *et al.* 1994).

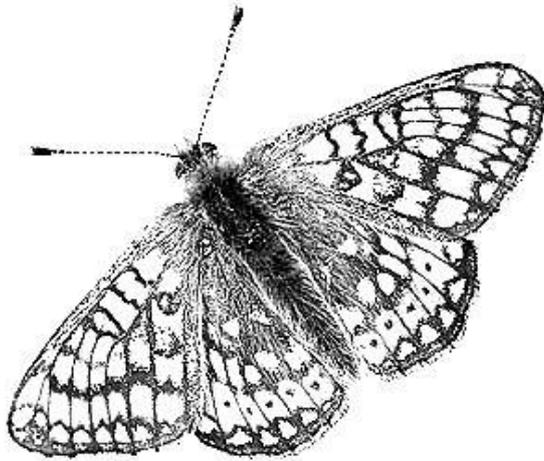
### **2.5.3 Implications for the conservation of *E. aurinia* across the mid-Cornwall Moors, and beyond.**

Much is now known about the ecology and conservation of *E. aurinia* (e.g. Bulman 2001; Porter 1981; Warren 1994) and the microhabitat preferences of both adults and larvae (e.g. Anthes *et al.* 2003; Betzholtz *et al.* 2007; Konvicka *et al.* 2003) as well as the viability of some populations across the species’ range (e.g. Bulman 2001; Schtickzelle *et al.* 2005; Wahlberg *et al.* 2002) using the incidence function model (IFM: Hanski 1994) to predict occurrences of *E. aurinia* over a network of habitat patches. A next step for the mid-Cornwall Moors metapopulation would be to conduct a population viability analysis and determine what possible future *E. aurinia* has there. Combined with the usage of habitat distribution models (Early *et al.* 2008) and studying the gene flow within and between populations, a solid management plan for the area could be envisaged.

It is sensible to conclude that short term studies are not enough for the responses of species to be measured on, or decisions based on for correct management. We may either be underestimating the re-colonisation of restoratively grazed patches (Pöyry *et al.* 2005) or alternatively we could be underestimating levels of population turnover and consequently, underestimating extinction threats (Thomas *et al.* 2002). Most assessments of metapopulation dynamics have based inferences on one or two year abundances or dynamics (e.g. Anthes *et al.* 2003; Hanski 1994; Hanski and Thomas 1994; Wahlberg *et al.* 1996). We show that even after five years of monitoring, more time is needed to allow restoration of habitats and responses of populations. However, our analyses of the mid-Cornwall population time series demonstrate that key, manageable habitat variables can be used to predict population densities, and manipulated to increase the suitability of habitats for population growth and establishment (whether by natural dispersal processes or via reintroduction schemes).

## Chapter 3

### **Development of microsatellites for the vulnerable marsh fritillary butterfly by *de novo* transcriptome sequencing.**



*'The one process now going on that will take millions of years to correct is the loss of genetic and species diversity by the destruction of natural habitats. This is the folly our descendants are least likely to forgive us.'*

*E. O. Wilson 1984  
The Biophilia Hypothesis*

### 3.1 ABSTRACT

Novel microsatellite markers were developed for the threatened marsh fritillary butterfly, *Euphydryas aurinia*, using both a time and cost-reducing methodology in comparison to current methods such as enrichment procedures. Using a transcriptome achieved through 454-pyrosequencing, EST-derived microsatellites were isolated and sequenced using an economic method of fluorescently labelling primers. Despite these markers being inherently difficult to isolate in Lepidoptera, nine polymorphic microsatellite loci were successfully achieved based on a combination of novel loci from this study and previously published loci. Although significant deviations from Hardy-Weinberg equilibrium are evident, we show potential for these markers in studies of population structure and gene flow desperately needed for declining species across fragmented landscapes.

### 3.2 INTRODUCTION

For many species the usual route to isolating and characterizing microsatellite markers involves an intensive, time-consuming process of enriching a genomic DNA library for a variety of di-, tri- or tetra-nucleotide repeat motifs, followed by a process of cloning and sequencing, to finally design species-specific primers (Glenn and Schable 2005; Kandpal *et al.* 1994; Ostrander *et al.* 1992). The subsequent sequencing process itself, involving the design and thorough testing of optimal multiplexes, then requires the purchase of an expensive primer carrying a fluorescent dye label, per locus being genotyped - causing costs to escalate rapidly depending on the study. However, microsatellites remain one of the most popular molecular markers for studies of population structure as they are highly polymorphic, abundant in relation to other popular markers such as amplified fragment length polymorphisms (AFLPs) and allozymes; and are easily genotyped by high throughput automated sequencing machines, meaning final scoring of alleles can be done with relative ease.

For the majority of species, microsatellites need to be isolated *de novo* due to the lack of having a genome already sequenced. Despite a wide variety of approaches for isolating microsatellites (Zane *et al.* 2002) it still proves difficult when dealing with lepidopteran genomes (Ji and Zhang 2004; Meglecz *et al.* 2004; Nève and Meglecz 2000; Zhang 2004). This inherent difficulty is associated with a lack of polymorphisms, similarity in flanking regions of different microsatellite loci, presence of null alleles and possible associations with mobile elements (Ji and Zhang 2004; Meglecz *et al.* 2004; Prasad *et al.* 2005; Van't Hof *et al.* 2007; Zhang 2004). However, there are still a large number of studies detailing the isolation of microsatellite markers from Lepidoptera. Inevitably it seems, these studies usually show a low number of loci, low levels of polymorphism, high occurrence of null alleles, or strong deviations from Hardy-Weinberg equilibrium (HWE) (Ji *et al.* 2003; Keyghobadi *et al.* 1999; Meglecz and Solignac 1998; Petenian *et al.* 2005; Prasad *et al.* 2005).

The recent advent of next generation 454-pyrosequencing (Margulies *et al.* 2005) and the resulting increase in publicly available EST data has encouraged quick and easy isolation of microsatellite markers (Kong and Li 2008; Metzger *et al.* 2011; Mikheyev *et al.* 2010; Sinama *et al.* 2011). With the future cost of 454-pyrosequencing decreasing, this is potentially a relatively cheap and easy method for rapid dissemination of results. However, microsatellites associated with expressed regions of the genome may be

under selection, even if they represent untranslated regions (UTRs) of genes, and hence almost always show strong deviations from HWE. Fortunately, the studies that have compared the performance of EST-derived microsatellites with other genotyping methods have generally found similar results (Mikheyev et al, in prep.; Mikheyev *et al.* 2010; Varshney *et al.* 2007; Woodhead *et al.* 2005).

The marsh fritillary, *Euphydryas aurinia*, is a severely declining species in both the UK and Europe (Asher *et al.* 2001; Fox *et al.* 2010). We report here the characterization of seven novel microsatellite loci using next-generation 454-pyrosequencing methods and confirmation of two previously published loci for *E. aurinia*, for use in a conservation genetics study with the aims of confirming conservation priorities and targets by determining population structure and gene flow across a fragmented landscape.

### 3.3 MATERIALS AND METHODS

RNA was extracted from a pool of 64 *E. aurinia* larvae collected from 16 sites in south-west UK (under Natural England licence number 20081071), using TRIzol (Invitrogen). Genomic DNA contamination was removed by DNase treatment (TURBO DNase, Ambion) for 30 min at 37 °C. RNA was further purified by using the RNeasy MinElute Clean up Kit (Qiagen) following the manufacturer's protocol. Full-length, enriched, cDNAs were generated from 2 µg of total RNA using the SMART PCR cDNA synthesis kit (BD Clontech) following the manufacturer's protocol. Reverse transcription was performed with the PrimeScript reverse transcription enzyme (Takara) for 60 min at 42 °C and 90 min at 50 °C. In order to prevent over-representation of the most common transcripts, the resulting double-stranded cDNAs were normalized using the Kamchatka crab duplex-specific nuclease method (Trimmer cDNA normalization kit, Evrogen) (Zhulidov *et al.* 2004). The resulting normalized cDNA library was sent to the Advanced Genomics facility at the University of Liverpool (<http://www.liv.ac.uk/agf>) for sequencing on the Roche 454 GS-FLX pyrosequencing platform. The assembly of the obtained reads was achieved using MIRA v2.9.26x3 (assembly performed 29-01-2009).

Both a custom Perl script and a custom python script were used to query the assembled data for  $\geq 5$  perfect repeats of di-, tri- and tetra-nucleotide repeats. Primers were then designed using Primer3 (Rozen and Skaletsky 2000) for those contigs containing microsatellites. Following the methods of Schuelke (2000), forward primers were designed with an addition of a M13(-21) tail at the 5' end (M13(-21): 5'-TGTAACGACGGCCAGT-3'). This is to allow the use of fewer expensive fluorescently labelled primers by only needing one 'universal' fluorescently labelled M13(-21) primer per multiplex. PCR conditions are such that the forward primer with M13(-21) tail is incorporated first, in partnership with a sequence specific reverse primer (no fluorescent label and no M13(-21) tail), and once this is used up the annealing temperature is lowered so that the 'universal' fluorescently labelled M13(-21) forward primer anneals instead (its annealing temperature is a few degrees lower). The fluorescent dye is thus incorporated into the PCR product (for a more thorough description, see: Schuelke 2000). If conducting several multiplexes, or groups of markers in one PCR reaction, differently labelled M13(-21) primers can be used for each group/multiplex and then all multiplexes can be combined when sequencing. Primers of previously published *E. aurinia* microsatellite loci (Peteani *et al.* 2005; Sigaard *et al.* 2008) were also ordered with a M13(-21) tail for testing across our samples.

All designed primers were initially tested for amplification on four samples. Each 10  $\mu$ l PCR mix contained approximately 1.5 ng genomic DNA, 1  $\mu$ l 10x EX Buffer (Takara), 2  $\mu$ M dNTPs, 0.25 U EX Taq HS (Takara), 1  $\mu$ l 10 mg/ml BSA, equimolar amounts of reverse primer and fluorescently labelled M13(-21) forward primer (10 pmol) and 2.5 pmol of the sequence specific forward primer (the sequence specific forward and reverse primers should be diluted in a 1:4 ratio, Schuelke 2000). Using a PTC-200 Thermo Cycler (MJ Research) the PCR conditions were as follows: 95°C for 5 min; followed by 25 cycles of 95°C for 5 s, 60°C for 30 s, 68°C for 1 min; followed by 8 cycles of 95°C for 5 s, 53°C for 30 s, 68°C for 1 min; and ended with a final extension at 72°C for 30 min. A maximum of 2  $\mu$ l of PCR product was then added to 10  $\mu$ l HiDi formamide (Applied Biosystems) and 0.1  $\mu$ l GeneScan -500 ROX standard (Applied Biosystems) and run on an ABI 3130xl Genetic Analyzer (Applied Biosystems). Allele presence and sizes were assigned using GeneMarker 1.7 (SoftGenetics, LLC). Potentially polymorphic loci were then tested on a further four samples from different populations. Multiplexes were designed (with loci that amplified across all eight

samples and were seen to be polymorphic over all eight) in groups with products of differing sizes. Multiplexes using differently labelled M13(-21) ‘universal’ forward primers were then mixed together and sequenced (on average 0.5 µl of each multiplex was added to the ROX and HiDi formamide mix, but this was adapted according to the strength of peaks resulting). GENEPOP v4.1 (Rousset 2008) was used to assess deviations from HWE. All microsatellite sequences were submitted to GenBank under accessions numbers JN116271 and JN116273 to JN116283.

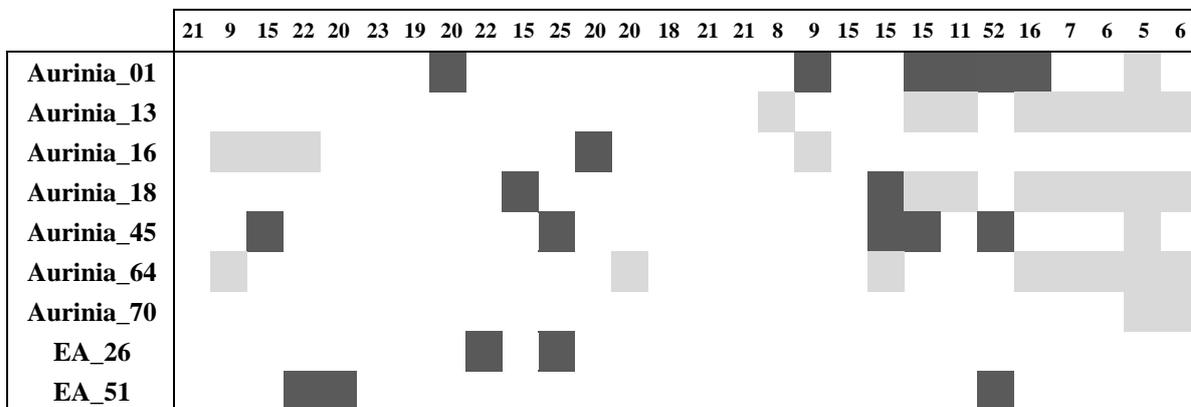
**Table 3.1** – Characteristics of 14 microsatellite loci in *Euphydryas aurinia* (Lepidoptera: Nymphalidae).

PCR #	Locus	Primer sequences (5' to 3')	M13 Label	Repeat motif	Range (bp)	N <sub>A</sub>	H <sub>O</sub>	H <sub>E</sub>	% missing	GenBank Acc. no.
1	Aurinia_01	F:CGAGCGTATTTGTCAAAGAAAG R:AGCGAATTAGGGTTGTCACATT	HEX	(CAT) <sub>6</sub>	252 - 270	6	0.396	0.503	8.1%	JN116271
2	Aurinia_13	F:AACGTTAACACTAGGGGTCTCA R:TATGATATAGTGTACGCGTTTTT	TET	(ATT) <sub>6</sub>	228 - 237	4	0.332	0.294	0.85%	JN116273
2	Aurinia_16	F:CCCGCTATGATCCATGTTTTA R:AAATTCATTTGTACTTTCGGTACAT	TET	(TTA) <sub>6</sub>	173 - 204	6	0.343	0.367	0.2%	JN116275
4	Aurinia_18	F:AAAAGCGCTGAAAGAAGAAAAA R:CAGTCTCAAAGATTTTCGCATAAAA	TET	(TAT) <sub>5</sub>	189 - 193	4	0.327	0.384	1.1%	JN116276
1	Aurinia_45	F:GGGTGAAATTGCGAATGAGT R:TCCCGCTACAGATGAAATC	HEX	(GTT) <sub>6</sub>	193 - 213	8	0.452	0.566	3.4%	JN116280
3	Aurinia_64	F:CAACCTGTAGCCGAAAAGA R:GCTTTTCTGTTGCCATCGTT	TAM	(TAC) <sub>8</sub>	201 - 210	4	0.188	0.203	4.0%	JN116282
1	Aurinia_70	F:CAACTTCAGTATGATCTCATTGCTTT R:TCACAATTTGCAGTGGCTCTAT	HEX	(GA) <sub>10</sub>	130 - 144	8	0.431	0.370	1.5%	JN116283
1	EA26*	F:CCGAGATACTCACCTACAAG R:CAGTGTATTTTCGGAACACAG	HEX	(TG) <sub>5</sub> TA(TG) <sub>11</sub> (AGCG) <sub>4</sub> (TG) <sub>3</sub>	163 - 180	13	0.705	0.730	5.7%	AY491806
5	EA51*	F:TGACGACAGATGGGTGTTC R:TGTAAGCGACTCAGTCTCATTTT	HEX	(TGTA) <sub>9</sub> (TGAT) <sub>2</sub>	129 - 142	4	0.483	0.622	5.1%	AY491828
-	Aurinia_14	F:TTTGTATGGGGAGAATTTATTGTTT R:TTTCTTTTAATCACAGATAACCTTTTT	-	(AT) <sub>8</sub>	226 - 232	4	-	-	-	JN116274
3	Aurinia_23	F:TACAAGCGCTTACCGAAGAAAC R:TCTGTCTGTTTCATCGCTCTCA	TAM	(GAT) <sub>7</sub>	147 - 162	5	-	-	-	JN116277
2	Aurinia_31	F:CAATTTAGGCGGCAAATTAAGA R:CCCGAGCTAAGCGACTACTACTA	TET	(AAC) <sub>5</sub>	252 - 258	2	-	-	-	JN116278
2	Aurinia_35	F:CAAGAACATGAATTTAGGTAAGCA R:CATAAGATTATGCCGGTATATAAAGTT	TET	(TAT) <sub>5</sub>	122 - 130	4	-	-	-	JN116279
1	Aurinia_62	F:TGTAGGCGACGTTCTTGCTA R:AATGCATTTTCGCATTCTAGG	HEX	(AAT) <sub>5</sub>	240 - 243	2	-	-	-	JN116281

Summary statistics based on a survey of 468 individuals from 28 populations. N<sub>A</sub>, number of observed alleles; H<sub>O</sub> observed and H<sub>E</sub> expected heterozygosity per locus. All Forward primers had M13 tail (TGTAACGACGCCAGT) added to 5' end, as described in text. Shaded loci were lost due to a lack of amplification across all samples, or because of null alleles, but may prove of use in other populations, or closely-related species. PCR reactions 1, 2 and 3 were pooled and analyzed together in the same sequencer run, as were reactions 4 and 5 – see text for description. \*Primers were originally published by Petician *et al.* (2005).

### 3.4 RESULTS

The 454 run generated 186,835 reads, and a total of 40,865,774 bases, after quality filtering. These were assembled into 22,032 contigs with a minimum overlap of 40% and 80% identity, and an average coverage of 3.15 X. Primers were designed for 74 microsatellite loci out of the 97 found in the 454 transcriptome, as some were unusable due to them being close to the end or beginning of a contig. Across an initial sample of four individuals, 36 of the novel loci did not amplify at all, 24 appeared monomorphic and 14 were potentially polymorphic. Out of the eleven previously published loci, only two appeared polymorphic in our samples. Testing across another four samples meant two novel primers were lost due to a lack of amplification across all samples. The remaining loci (see **Table 3.1**) were developed into multiplexes to enable all samples to be genotyped in just two groups. The locus Aurinia\_14 was lost during the multiplex designing period as it did not successfully amplify alongside other primers and was generally very weak and unreliable. Loci Aurinia\_23B, Aurinia\_31B, Aurinia\_62 and Aurinia\_35 either did not amplify well or gave large amounts of missing data and thus were excluded from further analyses, leaving nine loci for large-scale genotyping. In many of the sampled populations there was significant deviation from HWE, as expected (**Figure 3.1**).



**Figure 3.1** – Hardy-Weinberg equilibrium (HWE) ‘exact test’ statistics for each of the 28 populations included in the study (each column represents a population with sample size in the top row, with the order from left to right being the same order as populations in **Table 4.2, Chapter 4**) for each locus, after a sequential Bonferroni correction for multiple tests (Holm 1979; Rice 1989). Dark shading indicates significant deviation from HWE (chi-squared test,  $p < 0.05$ ); light shading indicates monomorphic loci, or only a single copy of a second allele if present. For loci with less than five alleles the complete enumeration method was used (Louis and Dempster 1987) but for all others a Markov chain algorithm was implemented (Guo and Thompson 1992) in GENEPOP v.4.1 (Rousset 2008).

### 3.5 DISCUSSION

This study has successfully isolated and characterized microsatellites for a threatened Lepidopteran species using a quick method of data mining for perfect repeats combined with inexpensive primer labelling. Schuelke's (2000) M13(-21) primer protocol is very advantageous to small research groups performing low-throughput genetic analyses, avoiding exponential costs for fluorescently labelled primers used on a minimal number of samples possibly only once.

Effective and reliable DNA markers are very desirable for a taxon with important roles in agricultural, ecological and horticultural scenes – the Lepidoptera. The markers developed in this study originated from a 454-pyrosequencing run not obtained from the most recent technology (i.e. not with titanium chemistry), and hence may not represent the entire breadth of markers to be found. There were also time constraints which made the study more of a quick venture: taking a full time researcher only two months from receiving primers to analysis.

This study also confirmed the successful use of two microsatellite loci previously developed (Petenian *et al.* 2005), although they both show considerable deviations from HWE across the sampled populations. However, loci developed here from a 454 transcriptome, despite some populations showing deviations from HWE (**Figure 3.1**), have perfect repeat motifs (neither compound microsatellites nor with any interruptions), are PCR 'multiplexable', and seem to be transferable between geographically distant populations (the 28 populations included in the study span from Cornwall in south-west UK, northwards to Scotland and across the English Channel to the Catalonia area in northern Spain – see **Figure 4.1** in **Chapter 4**), which fulfils a large number of desired criteria for developing 'robust' microsatellite markers (Sinama *et al.* 2011).

The extent to which deviations from HWE may affect analyses using EST-derived markers is unclear, and so verification of their performance through comparisons with other neutral markers such as AFLPs is useful. Mikheyev *et al.* (in prep.) show similar results for AFLP markers on a subset of the populations analysed here, and so confirm that these loci are applicable to studies of population structure in declining, fragmented populations of *E. aurinia*.

## Chapter 4



# To Protect or Connect? Population genetic analyses inform conservation strategies for the marsh fritillary butterfly, *Euphydryas aurinia*.



*'What a tragic irony, that the more we understand of biology, the less we have of it to learn from and enjoy..... to cherish and protect Nature, or to see butterflies and zebras and much more vanish into legend.'*

*Sean B. Carroll 2005  
Endless Forms Most Beautiful*



## 4.1 ABSTRACT

Habitat fragmentation, degradation and destruction have caused alarming declines in many species' distributions globally. For species existing in fragile metapopulations, this decline is exacerbated by the inability of individuals to disperse across a growing matrix of unsuitable habitat to locate good quality breeding areas. Subsequent loss of genetic diversity through genetic drift, reduced gene flow and inbreeding, is a fundamental concern for species persistence as their potential to adapt to changing environments diminishes and their genetic integrity is lost. This leads to a critical strategic decision for conservation management: should efforts be focused on the subpopulations themselves, to promote population growth, or should efforts be made to re-connect fragmented patches? Key information can be derived from population genetic analyses: populations may be diverse or genetically depauperate, and isolated or well-connected to other populations. The vulnerable marsh fritillary butterfly, *Euphydryas aurinia*, is the only British butterfly species to be protected in both the UK and Europe and is declining dramatically across both regions. Using microsatellite markers developed from a 454 transcriptome, we show signs of restricted gene flow at various spatial scales across the UK and Catalonia area of Europe. This conclusion is robust to analytical methodology: clear evidence of genetic structuring emerges from analyses of isolation by distance using sampling location information, as well as using the programme STRUCTURE to decipher population structure without the use of a priori information on sampling location. Genetic structure occurs at the international and regional scales of analysis, and persists among groups of populations whose separation distances are much closer to the predicted dispersal capabilities of the butterfly. With low allelic diversity (ranging from 2.25 to 5.78), relatively low levels of observed heterozygosity (ranging from 0.286 to 0.550), and some populations showing clear signs of inbreeding, it is emphasized that current management strategies and targets need to be adaptive to not only conserve genetic diversity but also genetic integrity of populations.

## 4.2 INTRODUCTION

Alarming declines in species distributions and densities have been witnessed over the past few decades in what is referred to as the ‘sixth extinction’ (Leakey and Lewin 1996), with much of the loss attributed to direct and indirect anthropogenic causes such as habitat fragmentation, degradation and destruction alongside climate change (Abbitt *et al.* 2000; Fahrig 2003; Lee and Jetz 2008; Saunders *et al.* 1991; Travis 2003). The result of these processes is often the creation of small, isolated populations. Consequently there is likely to be a decrease in local genetic variability and an increase in genetic differentiation between populations due to genetic drift, reduced gene flow and inbreeding which can all heighten extinction risk, and at the very least reduce the adaptive capabilities of populations to environmental and demographic change (Frankham *et al.* 2002; Young *et al.* 1996).

The marsh fritillary, *Euphydryas aurinia*, has been the subject of many studies over the past few decades, owing to its rapid distributional decline. Belonging to the group of butterflies known collectively as the ‘Checkerspots’ and historically known as the ‘greasy fritillary’, its distribution spans the Palaearctic, from Ireland in western Europe through 38 European countries, Morocco, Algeria and across Russia and temperate Asia to Korea (Tolman and Lewington 2009; van Swaay and Warren 1999). However, across much of this distribution it is considered to be vulnerable to local extinction, with an overall distributional decline of 20-50% over 25 years across Europe and already extinct in one European country where it was once present (van Swaay and Warren 1999). Listed under the EC Habitats and Species Directive and the Bern Convention, it is the only British butterfly species to be listed under European legislation as well as British (Wildlife and Countryside Act 1981 and UK Biodiversity Action Plan (BAP) priority species).

*E. aurinia* is one of many species considered to survive in what is termed ‘metapopulations’, whereby a single large metapopulation is the result of several smaller subpopulations linked by processes of stochastic local extinction and re-colonisation (Levins 1969). Metapopulation theory states that if levels of dispersal, and hence colonisation, are equal to or greater than levels of extinction across subpopulations, then the metapopulation as a whole should survive (Hanski 1994; Hanski 1999; Hanski and Gaggiotti 2004; Hanski *et al.* 1995; Levins 1969). However,

to enable levels of dispersal to remain relatively high and reduce the likelihood of detrimental inbreeding effects and loss of genetic diversity (Nieminen *et al.* 2001; Saccheri *et al.* 1998), conservation requires landscape-scale management to achieve well-connected networks of breeding habitat large enough to support metapopulations for generations into the future (Bulman *et al.* 2007). The genetic downside of these management targets is not always fully appreciated however. If subpopulations are genetically diverse and also highly differentiated from neighbouring subpopulations, connecting such patches could mean the loss of rare alleles, or a general lowering of genetic diversity across the landscape. Conversely, if isolated populations are genetically depauperate and neighbouring populations are not genetically differentiated at all; connecting patches is of no benefit to the overall genetic diversity and translocation from more diverse populations may be more appropriate.

Lepidoptera can be highly mobile (e.g. migratory painted lady, *Vanessa cardui*), or sedentary (e.g. black hairstreak, *Satyrrium pruni*), which has consequences for species-specific genetic characteristics and population structure (Habel and Schmitt 2009; Wang *et al.* 2003). *E. aurinia* has been known to disperse up to 15 km (Warren 1994) but is usually a sedentary species, estimated to move approximately 130 m on average per day (Wang *et al.* 2004). Such low levels of movement could exacerbate already restricted gene flow due to fragmented landscapes, and cause geographically isolated populations to also become highly genetically isolated. Isolation by distance (IBD: Wright 1943), whereby increasing geographic distance between populations results in further genetic differentiation between populations due to reduced gene flow, has been shown to be important in many species when explaining patterns of population structure across landscapes (Broquet *et al.* 2006; Peterson and Denno 1998; Vianna *et al.* 2003) and is likely to have a large influence on the genetic structure of *E. aurinia*.

There is a vast array of DNA marker techniques available currently for applications in an equally vast field of studies, from mapping agriculturally important genes in pest species to population genetics in species of conservation concern. Due to the ease of their isolation in comparison to microsatellites, genetic variation and population structure in Lepidopteran species has previously been investigated using either allozymes (Habel and Schmitt 2009; Joyce and Pullin 2001; 2003; Nève *et al.* 2000; Nève *et al.* 2008), or mitochondrial DNA sequences (mtDNA: Fauvelot *et al.* 2006; Jeong *et al.* 2009; Zhang *et al.* 2011; Zimmermann *et al.* 2000), with occasional studies

employing amplified fragment length polymorphism markers (AFLP: Wee 2004), random amplified polymorphic DNA markers (RAPD: Zhou *et al.* 2000) or direct amplification of length polymorphism markers (DALP: Wang *et al.* 2003). Although microsatellites are one of the most popular markers for most taxa due to their abundance, high levels of polymorphism and ease of scoring, there are significant problems still associated with their isolation in species of Lepidoptera. Despite studies using microsatellites in Lepidoptera being published, there is usually an inevitable lack of either numbers of loci used or levels of polymorphism, strong deviations from Hardy-Weinberg equilibrium (HWE) and suspected occurrence of null alleles (for example: Endersby *et al.* 2007; Ji *et al.* 2003; Keyghobadi *et al.* 1999; Petenian *et al.* 2005). These problems are associated with inherent characteristics of microsatellites in Lepidoptera such as a general lack of polymorphisms, similarity in flanking regions of different microsatellite loci, and possible associations with transposable elements (Ji and Zhang 2004; Meglecz *et al.* 2004; Nève and Meglecz 2000; Van't Hof *et al.* 2007; Zhang 2004).

Previous studies into the population genetics of *E. aurinia* in either the UK or the Catalonia area of northern Spain / southern France have used allozymes (Joyce and Pullin 2003) or AFLP markers (Wee 2004), respectively. The former study found significant genetic differentiation between populations in the UK at a National level, but very little differentiation at a regional level and distances closer to the predicted dispersal capabilities of the butterfly, and no evidence of IBD, inferring good levels of dispersal and population mixing at this scale. Wee (2004), on the other hand, found evidence of IBD and highly significant differentiation between populations in Catalonia, at a spatial scale not so different from the regional UK scale in Joyce and Pullin's (2003) study (see **Table 4.8** for a summary of *E. aurinia* studies). In a previous study by Joyce and Pullin (2001) using allozymes to assess the phylogeography of *E. aurinia* across the UK, France and Portugal, there had also been no evidence of IBD, which the authors attributed to a possible single post-glaciation colonization event, about 10,000 BP, and suggested consequently that populations should therefore show no significant structuring. These ambiguities between studies could be attributed to the difference in resolution afforded by the different genetic markers, as allozymes are less likely to show high levels of variation (Saccheri *et al.* 1999) which may lead to an underestimation of genetic differentiation.

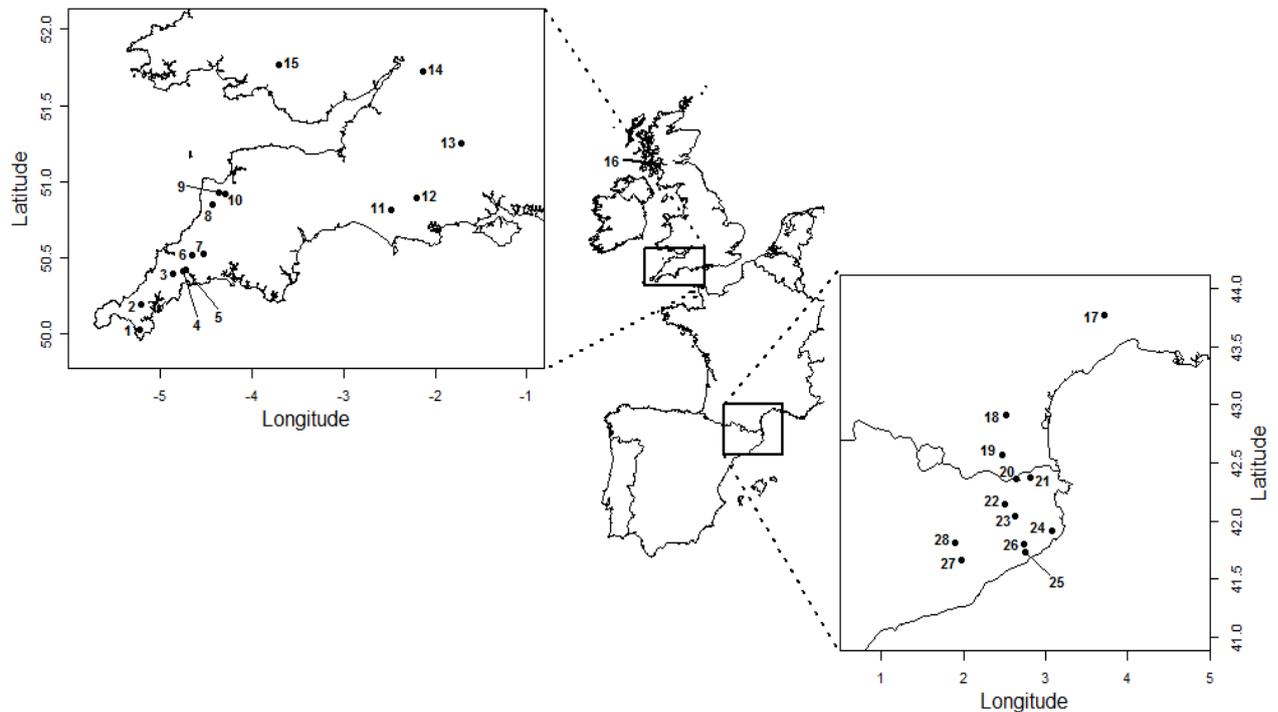
Here, another look is taken at a broader scale across both regions with new microsatellite markers developed from a 454 transcriptome (Smee et al, in prep.: see **Chapter 3**), comparing populations spanning from Scotland, through Wales, southern England and across to the Catalonia area of Europe (northern Spain / southern France). But there is also a focus on two key areas for *E. aurinia* in two very different ecological areas, in regards to climate and habitat variables – southern England and Catalonia. It is crucial to gain information on population structure and hence levels of gene flow at these different spatial scales across fragmented landscapes in order to determine suitable conservation goals and management practices for *E. aurinia*. Data gained from genetic markers will also be used to identify populations where long-term survival may be affected by genetic issues, such as levels of inbreeding and susceptibility to environmental and demographic stochasticity as the genetic potential to adapt is diminished. Conversely, this will also indicate those populations most suited as potential source populations for future reintroduction efforts, if needed. Hence, this study aimed to determine: (i) population structure across the landscape and the spatial levels at which population differentiation was significant, (ii) the genetic variation within regions and individual populations and (iii) whether current management strategies are appropriate in light of these results to preserve not only genetic integrity and potential for adaptation, but ensure the long-term survival of *E. aurinia* across its distribution.

### 4.3 MATERIALS AND METHODS

#### 4.3.1 Sample collection and preparation

Larvae of the univoltine *E. aurinia* (Lepidoptera; Nymphalidae) live gregariously in silken webs from the date they hatch in early to mid-July, through winter diapause, until the penultimate instar when they separate off to feed individually and then pupate in late Spring (Porter 1981). Across south-west UK, a single larva was collected from each web (to avoid siblings being included) found over a period of two generations (under Natural England licence numbers 20082745 and 20091722) to gain samples from populations across the entire English distribution (N = 14). In autumn 2008, samples were also collected by volunteers in Wales and Scotland (under Countryside Council for Wales (CCW) licence OTH:SPCA:09:2008 and Scottish Natural Heritage (SNH) licence number 9154, respectively) to facilitate comparisons across the entire UK mainland distribution of *E. aurinia*. European samples (from the Catalonia area of Spain

and southern France) of *E. aurinia* were collected by Wee (2004) and Singer, M. (pers. comm.) which allowed further comparisons with populations across the European distribution of *E. aurinia* (N = 12) (**Figure 4.1**).



**Figure 4.1** – Map of samples sites across the UK and Catalonia area of Europe. Population numbers refer to those in **Table 4.2**. In further ‘international’ scale analyses, these populations are grouped into the following: populations 1 to 10 are ‘south-west England’; populations 11 to 14 are ‘mid-south England’; population 15 is ‘Wales’ as the sample includes individuals from across the whole distribution in Wales; population 16 is ‘Scotland’ as the sample includes individuals from across the whole distribution in Scotland; and populations 17 to 28 are ‘Catalonia’ in Europe.

Each larva collected from south-west UK, Scotland and Wales was starved for 24 hours to remove plant material from the digestive tract and then dissected to remove the midgut and potentially any parasitoids, before then being snap frozen. Following the methods of Martinez-Torres *et al.* (1998) (with some modifications – see **Appendix 1**), genomic DNA was extracted, diluted to a working concentration (~15 ng/μl), and stored at -20°C until required. DNA from samples collected in Europe was extracted from the heads of larvae using DNeasy Tissue kits (Qiagen) (Wee 2004).

### 4.3.2 Genotyping

As described by Smee *et al.* (in prep. – see **Chapter 3** of this thesis) nine microsatellite loci (**Table 4.1**), seven of which were novel and developed from a 454 transcriptome of *E. aurinia* and two which were successfully amplified from previously published loci, were used to genotype a total of 468 individuals from 28 populations across the UK and Catalonia area of Europe (**Figure 4.1**). Using a method described by Schuelke (2000), that requires fewer expensive fluorescently labelled primers than usual methods, loci were multiplexed and genotyped using an ABI 3130xl Genetic Analyzer (Applied Biosystems), and then alleles scored using GeneMarker 1.7 (SoftGenetics, LLC). The data were formatted using GenAlEx (Peakall and Smouse 2006) for input into various other population genetics software programmes. Summary statistics for each locus were then determined using GenePop v.4.1 (Rousset 2008) (**Table 4.1**).

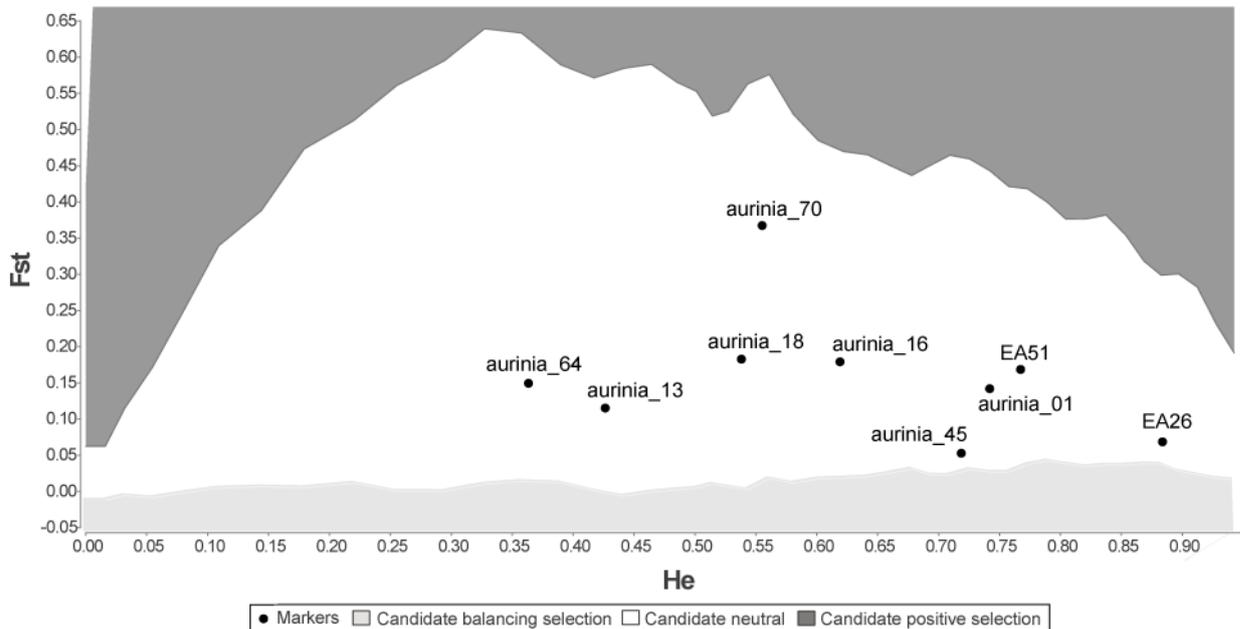
**Table 4.1** – Microsatellite summary statistics for nine loci used to genotype *E. aurinia* across the UK and Catalonia area of Europe.

Locus	GenBank	N <sub>A</sub>	N	H <sub>O</sub>	H <sub>E</sub>
Aurinia_01	JN116271	6	431	0.396	0.503
Aurinia_13	JN116273	4	464	0.332	0.294
Aurinia_16	JN116275	6	467	0.343	0.367
Aurinia_18	JN116276	4	463	0.327	0.384
Aurinia_45	JN116280	8	452	0.452	0.566
Aurinia_64	JN116282	4	449	0.188	0.203
Aurinia_70	JN116283	8	461	0.431	0.370
EA26*	AY491806	13	442	0.705	0.730
EA51*	AY491828	4	444	0.483	0.622

\*Microsatellite loci were originally published by Petenian *et al.* (2005). N<sub>A</sub>, number of alleles; N, number of samples; H<sub>O</sub> (Observed) and H<sub>E</sub> (Expected) Heterozygosity.

For some of the developed loci across a few populations, deviations from HWE were evident even after sequential Bonferroni correction for multiple tests (Holm 1979; Rice 1989) (see **Chapter 3** in this thesis). However, after using the programme LOSITAN (Antao *et al.* 2008; Beaumont and Nichols 1996) to identify putative loci under selection across all samples, using a method of comparing  $F_{ST}$  values against expected heterozygosity ( $H_E$ ) to detect outliers, it was shown that all loci fell into the candidate neutral category and so were applicable for use in the current study (**Figure 4.2**). There also appeared to be some linkage disequilibrium (LD) amongst three of the loci, although this pattern was not seen across all populations and so could be due to real

genetic structure across the sampled locations, rather than the loci being physically linked.



**Figure 4.2** – LOSITAN results for all loci across all samples (geographically grouped into south-west England, mid-south England, Wales, Scotland and Catalonia). This workbench uses an  $F_{ST}$  outlier method to detect potential loci under selection and therefore unsuitable for use in studies reliant on neutral markers, such as population genetics. Under a stepwise mutation model (SMM), all loci were found to be candidate neutral markers – located in the white region of the figure. The dark grey region above indicates where candidate positive selection markers would be found, and the lighter region below indicates where candidate balancing selection markers would be found. Under an infinite alleles model (IAM) of mutation, the result is the same.

### 4.3.3 Analysis of genetic diversity and variation

General diversity statistics per population (mean number of alleles and heterozygosities) were calculated in GenePop v.4.1 (Rousset 2008), and the number of private alleles (alleles found only in one population) were calculated in GenAlEx (Peakall and Smouse 2006).  $F$ -statistics (Weir and Cockerham 1984) were calculated at an international, regional and individual population scale using Arlequin v.3.5 (Excoffier and Lischer 2010) and their significance tested with 16000 permutations.

### 4.3.4 Population genetics – is sampling information needed?

Population structure and genetic differentiation were analysed here using approaches with and without a priori information on sampling locations. The former analysis

promotes the more traditional study of isolation-by-distance (IBD), while the latter follows that the genetic structure should be studied as an emergent property of the genetic samples alone, and promotes post-hoc mapping of genetic structure back onto the landscape. Interpretation of the two protocols can be very different, and can be frustrating if their results suggest divergent hypotheses of population structure. However we did not prefer one approach over the other *a priori*, and in this instance their results lead to similar conclusions. Both approaches were analysed at different spatial scales using progressive partitioning from international to regional: 1. internationally over all sites (using the five groupings as previously stated, see **Figure 4.1**); 2. across southern England (populations 1 to 14 in **Figure 4.1**); and 3. across the Catalonia sites in Europe (populations 17 to 28 in **Figure 4.1**).

#### 4.3.4.1 Isolation-by-distance (IBD) analyses using geographical sampling sites

As IBD is thought to be the most obvious source of intraspecific variation in gene flow (Wright 1943) and is common across phytophagous insects (Peterson and Denno 1998), it was investigated using the online-based programme IBD (Jensen *et al.* 2005). Pairwise  $F_{ST}$  distances (Slatkin's linearized  $F_{ST}$ ; see Slatkin (1995) or Rousset (1997)) were calculated using Arlequin v.3.5 (Excoffier and Lischer 2010) alongside pairwise geographical distances (km) between each population. Mantel tests were an appropriate method to determine significant correlations between the two matrices of distances as they consider the unit of replication to be a population and not a pairwise contrast. When conducting each Mantel test and plotting the two distances against each other, only the geographic distance was log-transformed as recommended by Rousset (1997) for a two-dimensional stepping stone model. For the analysis on the international scale, a third 'indicator' matrix was added which coded for the presence (1) or absence (0) of sea between the populations, as the English Channel (an area of sea separating the UK from France by up to 240km) was expected to be a significant barrier to dispersal for *E. aurinia*. For this analysis, a partial Mantel test was conducted using methods described by Legendre and Legendre (1998). Reduced major axis (RMA) regression was used to calculate slopes and intercepts for all tests.

#### 4.3.4.2 Assessing population structure without prior sampling information

Using the programme STRUCTURE v.2.3.3 (Pritchard *et al.* 2000), population structure was assessed across the three regions previously stated, without the use of a priori sampling location information. Bayesian methods are used by the programme to assign individuals probabilistically to 'populations' characterized by allele frequencies at the available loci, whilst maximising HWE and minimising LD. The number of potential

‘populations’, or distinct genetic clusters ( $K$ ), was investigated for the different spatial scales using values of  $K$  appropriate to that test, i.e. from 1 to  $N$  ( $N$  being the number of geographical sampling locations) for that subset of data. For each value of  $K$ , five independent runs were executed with a Markov chain Monte Carlo (MCMC) of 200,000 iterations following a burn-in of 100,000 iterations, assuming an admixture model with correlated allele frequencies (Falush *et al.* 2003). Using an online programme, Structure Harvester v.0.6.5 (Earl 2011), the number of genetic clusters,  $K$ , in the data was estimated using the ‘Evanno method’ (Evanno *et al.* 2005), whereby an ad hoc statistic based on the rate of change in the log-likelihood between successive values of  $K$  is used to identify the ‘true’ number of clusters.

Analyses of molecular variance (AMOVAs; Excoffier *et al.* 1992) were also carried out using Arlequin v.3.5 (Excoffier and Lischer 2010) at several spatial levels, with and without sampling location information, to determine if the genetic ‘populations’ determined from the STRUCTURE analyses could be validated, and to assess where the majority of genetic variation was found.

#### 4.4 RESULTS

General statistics revealed low levels of genetic diversity across *E. aurinia* populations, with the number of alleles at each locus ranging from 4 to 13 (**Table 4.1**) and an overall total of 57 alleles across all nine loci. **Table 4.2** shows diversity statistics per individual population, where the mean number of alleles ranged from 2.25 to 5.78 (overall mean and standard deviation, s.d., of  $3.47 \pm 1.66$ ) and a total of 13 private alleles were found. The frequency of private alleles was generally low ( $< 0.1$ ) apart from within the populations Stowford (0.233) and Darnius (0.115). Levels of expected heterozygosity,  $H_E$ , ranged from 0.423 to 0.617 with a mean of  $0.466 \pm 0.248$  s.d. across all populations and similar levels for the two geographical regions separately (southern England,  $0.501 \pm 0.224$ ; Catalonia,  $0.416 \pm 0.277$ ). There was a significant ( $p < 0.001$ ) heterozygosity deficit across the entire sample with a global  $F_{IS}$  value of 0.127 ( $p < 0.001$ ), likely due to the spatial population structure of *E. aurinia* or the presence of null alleles. Individual values of  $F_{IS}$  per population are given in **Table 4.2** and demonstrate a mixture of high levels of inbreeding (e.g. a high  $F_{IS}$  value of 0.45 for Sils) and low (e.g. a low  $F_{IS}$  value of -0.30 for Col Estelales).

**Table 4.2** – Sites across the UK and Catalonia area of Europe from which *E. aurinia* was sampled, showing genetic diversity statistics per ‘geographical’ site.

Site #	Area/County	Population Name	Coordinates		Sample Size	A	H <sub>O</sub>	H <sub>E</sub>	F <sub>IS</sub>	P <sub>A</sub>
			Lat.	Long.						
1	Cornwall	Lizard	50.0285	-5.2113	21	3.44	0.519	0.507	-0.02	-
2	Cornwall	Stithians	50.1917	-5.1985	9	3.14	0.524	0.541	0.03	1
3	Cornwall	Goss Moor	50.3957	-4.8590	15	3.56	0.396	0.497	0.21	-
4	Cornwall	Breney Common	50.4139	-4.7404	22	3.63	0.502	0.564	0.11	1
5	Cornwall	Redmoor	50.4215	-4.7225	20	3.22	0.426	0.510	0.17	-
6	Cornwall	Colvannick	50.5123	-4.6473	23	3.56	0.452	0.499	0.09	1
7	Cornwall	Carkeet	50.5251	-4.5168	19	3.89	0.387	0.479	0.19	-
8	Devon	Dunsdon	50.8446	-4.4204	20	4.22	0.495	0.609	0.19	-
9	Devon	Volehouse	50.9277	-4.3592	22	3.78	0.510	0.536	0.05	-
10	Devon	Stowford	50.9219	-4.2906	15	4.00	0.464	0.576	0.20	1
11	Dorset	Giants Hill	50.8175	-2.4755	25	4.11	0.406	0.508	0.20	-
12	Dorset	Hod Hill	50.8956	-2.2075	20	3.89	0.386	0.445	0.14	-
13	Wiltshire	Salisbury	51.2479	-1.7134	20	4.11	0.475	0.491	0.04	1
14	Gloucestershire	Strawberry Bank	51.7274	-2.1332	18	3.11	0.420	0.435	0.04	-
15	Wales	Wales – varied	51.7655	-3.7041	21	4.00	0.429	0.553	0.23	-
16	Scotland	Scotland - varied	55.9790	-5.6674	21	4.89	0.385	0.498	0.23	2
17	S. France	Pic St Loup	43.7715	3.7156	8	3.78	0.316	0.476	0.36	1
18	S. France	Col de la Redoulade	42.9122	2.5145	9	3.13	0.490	0.519	0.06	1
19	S. France	Col del Forn	42.5681	2.4686	15	3.00	0.394	0.423	0.06	-
20	Catalonia	Coustouges	42.3584	2.6473	15	4.38	0.506	0.606	0.17	-
21	Catalonia	Darnius	42.3704	2.8178	15	3.29	0.336	0.534	0.35	1
22	Catalonia	Can Jorda	42.1443	2.5038	11	3.38	0.537	0.513	-0.05	-
23	Catalonia	LaBaraca	42.0457	2.6254	52	5.78	0.332	0.474	0.28	2
24	Catalonia	Mas Calc	41.9112	3.0723	16	3.88	0.409	0.481	0.14	-
25	Catalonia	Tordera	41.7308	2.7475	7	3.00	0.409	0.617	0.33	-
26	Catalonia	Sils	41.8004	2.7310	6	2.86	0.286	0.525	0.45	-
27	Catalonia	Col Estelales	41.6635	1.9791	5	2.25	0.550	0.434	-0.30	-
28	Catalonia	El Guix	41.8159	1.9044	6	2.83	0.433	0.462	0.05	1

A, mean number of alleles per locus; H<sub>O</sub> observed and H<sub>E</sub> expected heterozygosity; mean estimates of the inbreeding coefficient F<sub>IS</sub> for each population; and P<sub>A</sub>, number of private alleles in each population.

#### 4.4.1 International broad-scale analysis

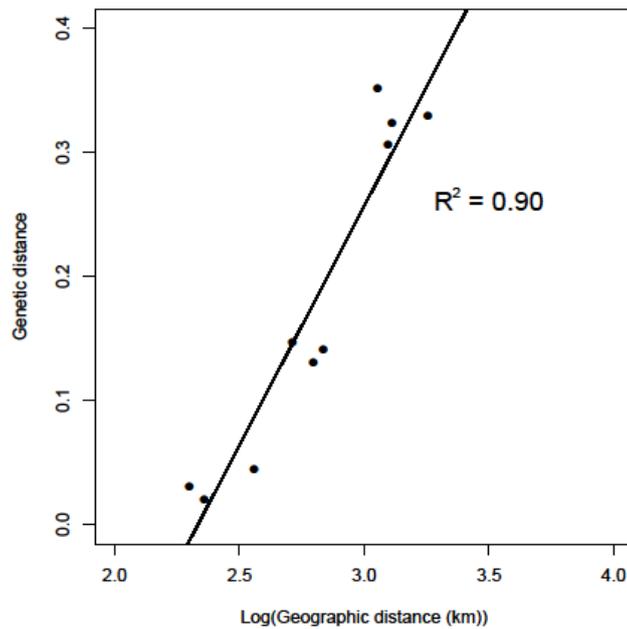
As would be expected over a large geographical scale in a moderately sedentary species, the overall international  $F_{ST}$  value for *E. aurinia* was 0.22 ( $p < 0.001$ ), which indicates a strong restriction of gene flow. As suggested by Wright (1978),  $F_{ST}$  values in the range of 0.15 to 0.25 indicate high genetic differentiation, whereas 0.05 to 0.15 values indicate moderate genetic differentiation.

##### 4.4.1.1 IBD using geographical locations

On an international scale, when grouping all individuals into the five main geographical regions, there is significant IBD ( $Z = 5.50$ ,  $r = 0.95$ ,  $p = 0.028$ ) with an intercept and slope of -0.96 and 0.41, respectively (**Table 4.3**, **Figure 4.3**). However, once an ‘indicator’ matrix had been added to the analysis to take into account the potential barrier to dispersal of the English Channel, this significance was lost (partial Mantel test controlling for the ‘indicator’ matrix:  $r = 0.85$ ,  $p = 0.094$ ). Consequently it appears the English Channel is a greater cause of differentiation between the European and UK populations than distance alone.

**Table 4.3** – Upper half: Euclidean geographical distance (km). Lower half: genetic distance - Slatkin’s linearized  $F_{ST}$  ( $F_{ST} / (1 - F_{ST})$ ; Slatkin 1995). All pairwise comparisons remained significant after a sequential Bonferroni correction for multiple tests (Holm 1979; Rice 1989).

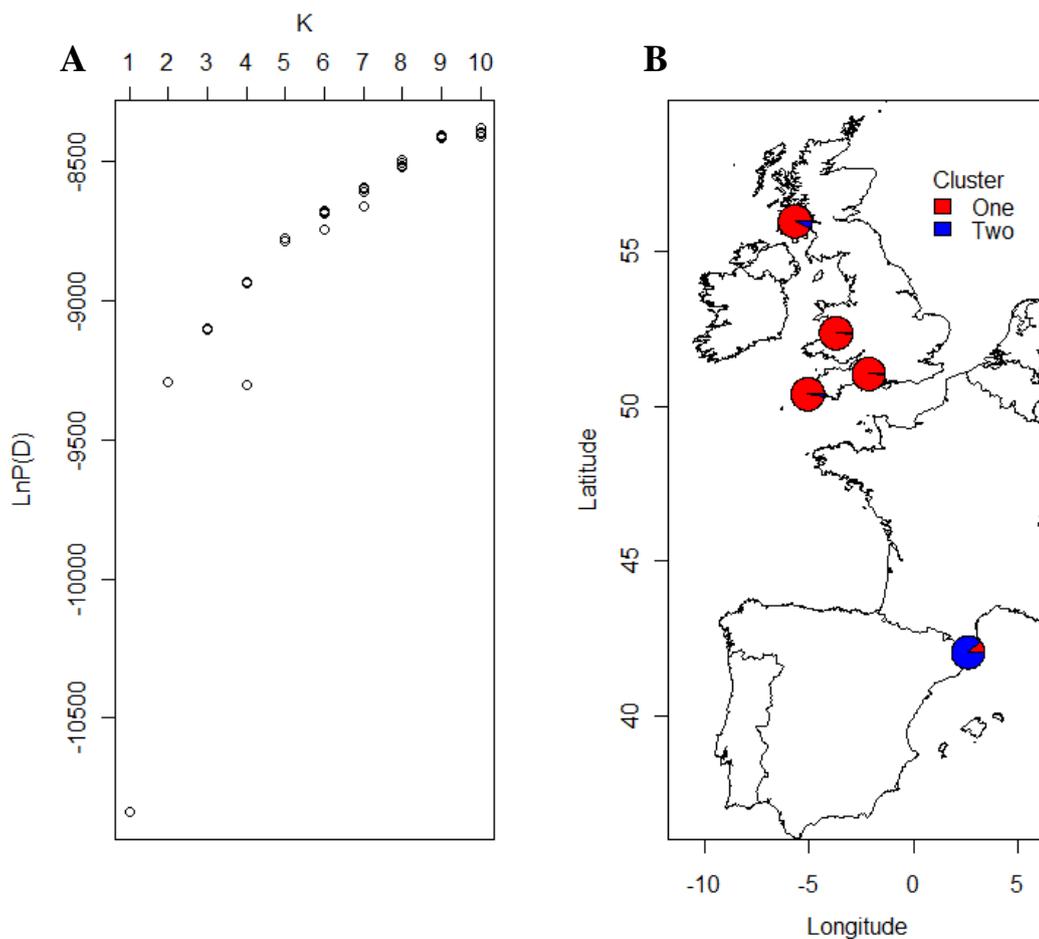
Region	Southwest	Mid-South	Wales	Scotland	Catalonia
Southwest	0	362.3398	198.8998	624.7107	1246.241
Mid-South	0.04431	0	228.6988	684.6907	1131.021
Wales	0.031194	0.020689	0	515.4438	1289.25
Scotland	0.130927	0.141045	0.146934	0	1801.324
Catalonia	0.306336	0.351863	0.323627	0.329469	0



**Figure 4.3** – Significant ( $p = 0.028$ ) isolation by distance (IBD) across all samples grouped into five localities as described in the text. Genetic distance is Slatkin’s linearized  $F_{ST} / (1 - F_{ST})$  (Slatkin 1995).

#### 4.4.1.2 Population structure without prior sampling information

Based on the genetic data alone, there is also strong evidence of genetic differentiation from the analysis of all 468 individuals across an international scale using the programme STRUCTURE v.2.3.3 (Pritchard *et al.* 2000). Although the mean log-likelihood from the five independent runs was highest at a value of  $K = 10$  (**Figure 4.4A**), the rate of change in the log-likelihood between successive values decreases substantially after  $K = 2$ , and so this was deemed the mostly likely ‘real’ number of genetic clusters. Using biological rationale this also appears likely, as the clustering mainly separates the UK samples from the European (**Figure 4.4B**), with Scotland having the largest similarity with Europe, and the two most northern populations from Europe (17 and 18 in **Figure 4.1**) causing the similarity with the UK (Pic St Loup and Col de la Redoulade having 82.8% and 90.5% identity with the UK group, respectively).



**Figure 4.4** – (A) Log-likelihood of each value of  $K$  (the number of distinct genetic clusters) for an analysis of all 28 populations in STRUCTURE v.2.3.3 (Pritchard *et al.* 2000), using an admixture model and correlated allele frequencies. (B) Using the online programme Structure Harvester v.0.6.5 (Earl 2011) to estimate an ad hoc statistic based on the rate of change in log-likelihood  $\text{LnP}(D)$ , there is evidence for two main clusters across the region.

In analyses of molecular variance (AMOVAs), a significant amount of variation ( $p < 0.001$ , see **Table 4.4**) is evident at all partitions of the data. As the geographic populations are grouped by Country, this absorbs a large amount of variation but for both analyses, the variation among individuals within populations accounts for the least, after the variation among populations, which does support the level of clustering used for the genetic populations ( $K = 2$ ). For both geographic populations and genetic populations, the majority of variance is within individuals (see **Table 4.4**).

**Table 4.4** – Analysis of molecular variance (AMOVA) results at an international scale, comparing the use of geographical and genetic ‘populations’.

	Source of variation	Variance components	% of variation
<i>Geographical populations</i>	Among groups (UK / Europe)	0.553	20.08***
	Among populations/within groups	0.277	10.06***
	Among individuals/within populations	0.288	10.47***
	Within individuals	1.634	59.39***
<i>Genetic populations</i>	Among populations	0.812	24.88***
	Among individuals/within populations	0.601	18.44***
	Within individuals	1.849	56.68***

\*\*\*p&lt;0.001

#### 4.4.2 Regional fine-scale analysis

After progressively partitioning the populations to a regional level, two key areas for *E. aurinia* in southern England (populations 1 to 14) and Catalonia (populations 17 to 28) both show significant indications of restricted gene flow with  $F_{ST}$  values of 0.10 and 0.18, respectively ( $p < 0.001$ ).

##### 4.4.2.1 IBD using geographical locations

For populations of *E. aurinia* in southern England, there is no significant evidence of IBD ( $Z = 23.78$ ,  $r = 0.24$ ,  $p = 0.063$ , **Table 4.5**, **Figure 4.5A**). However, population 6 (Colvannick) shows a high level of differentiation from nearby sites and appears quite distinct in the clustering analysis based on genetic data only (see next section and **Figure 4.7**). When this population is removed from the IBD analysis a significant correlation between geographic and genetic distance appears ( $Z = 19.59$ ,  $r = 0.29$ ,  $p = 0.017$ ) with an intercept and slope of -0.22 and 0.16, respectively.

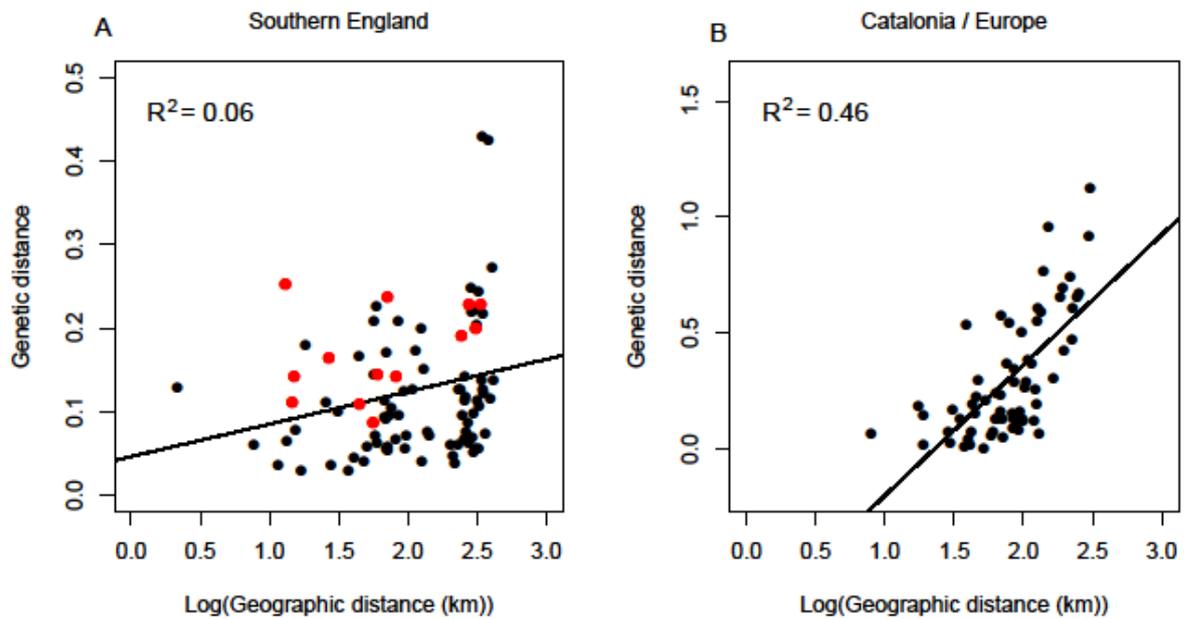
**Table 4.5 - Lower half:** Pairwise  $F_{ST}$  distances using Slatkin's linearized  $F_{ST}$  ( $F_{ST} / (1 - F_{ST})$ ) (Slatkin 1995) between fourteen populations in southern England. All pairwise comparisons remained significant after a sequential Bonferroni correction for multiple tests (Holm 1979; Rice 1989). **Upper half:** Euclidean distances (km). Population numbers correspond to those in **Figure 4.1**.

POPULATION	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	18.1281	56.4703	67.5605	69.6381	82.4964	94.8203	126.138	137.495	142.414	316.548	347.586	411.817	390.742
2	0.17867	0	43.9967	56.5738	58.7324	70.8380	84.3230	112.780	123.938	129.379	310.640	341.630	404.863	381.063
3	0.14444	0.16629	0	13.3410	15.4449	26.8528	40.6635	69.6688	81.0270	86.0020	269.133	300.007	362.339	337.206
4	0.09316	0.20842	0.06471	0	2.16123	15.0361	27.7501	59.5406	70.9944	75.3162	255.796	286.674	349.086	324.496
5	0.17126	0.2263	0.0768	0.12881	0	13.0835	25.5931	57.6865	69.1395	73.3710	253.689	284.562	346.944	322.339
6	0.1428	0.23623	0.16363	0.14229	0.2525	0	14.5801	44.6467	56.0950	60.2906	243.856	274.608	336.299	310.386
7	0.05541	0.20833	0.04394	0.03522	0.11024	0.11111	0	37.0052	47.9498	50.6746	229.290	260.054	321.886	296.746
8	0.03976	0.17391	0.05644	0.06273	0.07074	0.10937	0.02872	0	11.4539	16.7859	216.283	246.128	304.316	272.551
9	0.07476	0.1985	0.06585	0.05408	0.09157	0.0862	0.03994	0.03635	0	7.65503	209.814	239.284	296.339	262.958
10	0.07208	0.15083	0.09554	0.10401	0.09572	0.14461	0.05777	0.02969	0.06058	0	202.162	231.648	288.848	256.016
11	0.11376	0.20324	0.08618	0.11421	0.06589	0.19104	0.06022	0.03895	0.04667	0.05967	0	31.0378	97.3060	108.022
12	0.21684	0.43015	0.22163	0.21978	0.24914	0.22746	0.11668	0.09618	0.12625	0.12629	0.09958	0	67.4637	92.7941
13	0.1385	0.27153	0.07255	0.12125	0.12552	0.22862	0.05584	0.05512	0.09816	0.06834	0.07125	0.11299	0	70.8419
14	0.11627	0.42522	0.13855	0.10624	0.24289	0.20026	0.05116	0.06174	0.07574	0.1421	0.12654	0.12359	0.05703	0

**Table 4.6 - Lower half:** Pairwise  $F_{ST}$  distances using Slatkin's linearized  $F_{ST}$  ( $F_{ST} / (1 - F_{ST})$ ) (Slatkin 1995) between twelve populations in Catalonia, Europe. Non-significant differences after sequential Bonferroni correction for multiple tests (Holm 1979; Rice 1989) are in bold ( $p < 0.05$ ). **Upper half:** Euclidean distances (km). Population numbers correspond to those in **Figure 4.1**.

POPULATION	17	18	19	20	21	22	23	24	25	26	27	28
17	0	164.1333	192.5602	196.7838	184.8216	225.3816	226.7319	218.5308	250.8263	244.6867	303.458	296.1901
18	0.30459	0	38.56522	63.26476	68.98588	85.31301	97.04019	127.305	133.7602	125.8215	150.9754	139.4288
19	0.69536	0.53092	0	30.61803	44.60881	47.24247	60.59347	99.14193	98.04111	90.12988	114.3024	104.5054
20	0.42158	0.23593	0.16795	0	19.00553	28.63965	34.81911	68.55545	70.5934	62.67334	107.1509	102.2593
21	0.65703	0.57565	0.15364	0.14437	0	43.00914	41.93244	58.31937	71.46522	64.03833	121.9216	118.7883
22	0.47376	0.28764	0.29818	0.07253	0.19526	0	17.40087	68.31015	53.3287	45.79831	79.10669	75.98499
23	0.60682	0.50128	0.07407	0.12863	0.06846	0.18677	0	51.88945	37.51944	29.66885	83.4735	84.13907
24	0.73888	0.60772	0.12189	0.15937	<b>0.05457</b>	0.22971	<b>0.00101</b>	0	41.30038	39.89572	124.6336	130.2954
25	0.6688	0.59164	0.12606	<b>0.125</b>	<b>0.04729</b>	0.21046	<b>0.00576</b>	<b>0.01226</b>	0	7.945067	85.76871	94.22401
26	0.65563	0.55348	<b>0.11746</b>	0.12825	<b>0.12818</b>	0.2256	<b>0.02162</b>	<b>0.03958</b>	<b>0.06475</b>	0	84.97965	91.92986
27	1.12241	0.95463	0.36918	0.37897	0.25227	0.5455	0.15195	0.19339	<b>0.34305</b>	<b>0.12989</b>	0	18.86358
28	0.91891	0.76236	0.28612	0.26253	<b>0.11927</b>	0.37041	<b>0.08387</b>	<b>0.06712</b>	<b>0.16071</b>	<b>0.07969</b>	<b>0.01331</b>	0

In Catalonia however, there is significant evidence of strong IBD ( $Z = 41.58$ ,  $r = 0.68$ ,  $p < 0.001$ ) with an intercept of  $-1.27$  and slope of  $0.83$  (Table 4.6, Figure 4.5B).

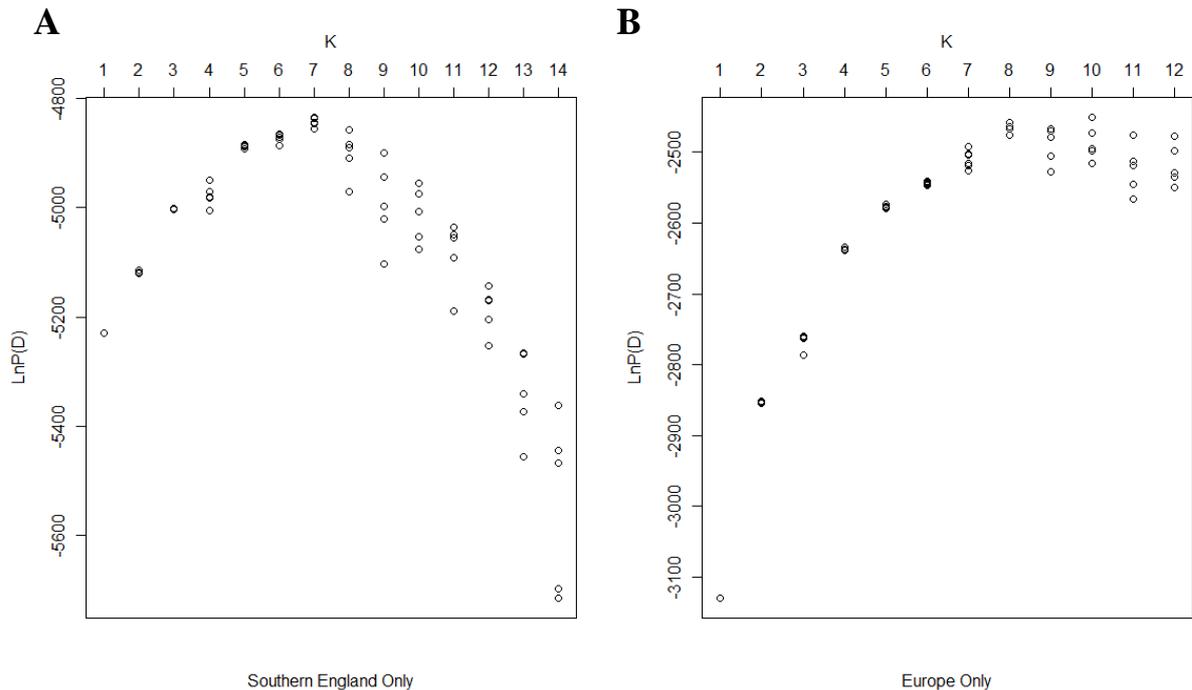


**Figure 4.5** – IBD analyses for (A) southern England (non-significant,  $p = 0.063$ ) until the population at Colvannick, highlighted in red, is removed ( $p = 0.017$ ) and (B) Catalonia (highly significant,  $p < 0.001$ ). Genetic distance is Slatkin’s linearized  $F_{ST} / (1 - F_{ST})$  (Slatkin 1995).

#### 4.4.2.2 Population structure without prior sampling information

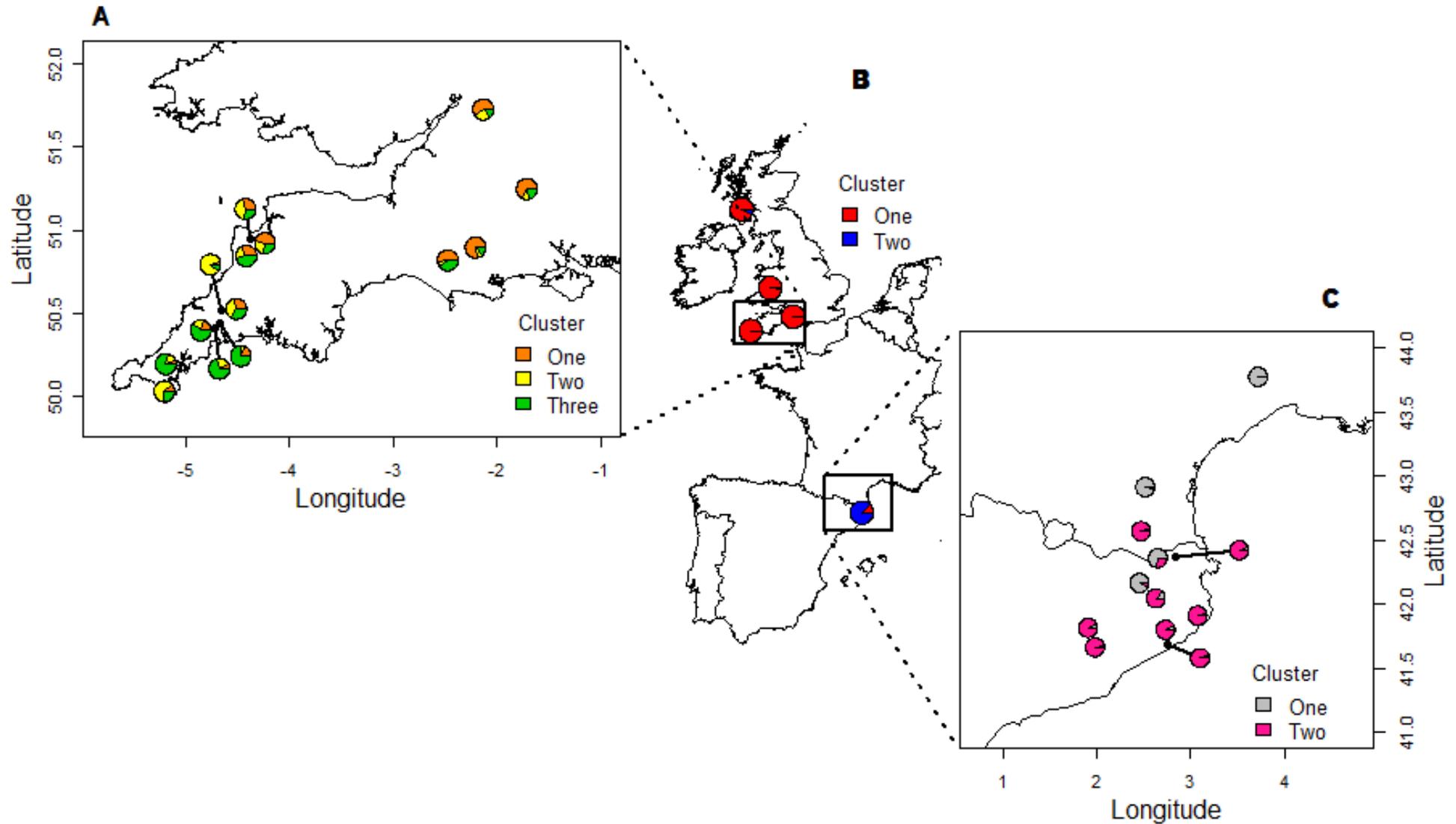
Using only genetic data for both regions individually, there is evidence for genetic structuring within the 264 individuals genotyped from southern England, and the 162 individuals genotyped from Catalonia, using STRUCTURE v.2.3.3 (Pritchard *et al.* 2000). Structure Harvester (Earl 2011) was again used to determine appropriate values of  $K$  according to the rate of change in log-likelihood shown for each region, resulting in  $K = 3$  for southern England, and  $K = 2$  for Catalonia (Figure 4.6). It is harder to justify these levels of clustering using biological rationale than for the international analysis as there is no clear barrier to gene flow within these regions, apart from much finer-scale habitat fragmentation and landscape features. However, Figure 4.7 shows a representation of how these genetic clusters are distributed across the regions of southern England and Catalonia, which gives support for the presence of IBD previously shown as within each population the dominant cluster membership changes on a West to East cline for southern England, and a North to South cline for Catalonia. When allocating individuals to higher numbers of clusters ( $K$ ) in either region, such

patterns are not as clear and a biological rationale is harder to provide for the patterns that are shown.



**Figure 4.6** – Output from a Bayesian clustering analysis in STRUCTURE v.2.3.3 (Pritchard *et al.* 2000) using no a priori information on sampling location. Five independent runs were carried out for each value of K for (A) populations in southern England, and (B) populations in the Catalonia area of Europe only, to gain a log-likelihood for each value of K and estimate the ‘real’ number of genetic populations in the sample based on the rate of change in log-likelihood between consecutive values of K.

In AMOVAs for each of the two regions using both geographic and genetically defined ‘populations’, there is a significant amount of the total variation explained at all partitions of the data ( $p < 0.001$ , **Table 4.7**). For southern England, when grouped according to geographical population, there is a roughly equal division of variance between individuals within populations and among populations themselves, whereas in Catalonia the latter is higher. For both regions, when grouped genetically, there is much more variance among individuals within populations than among populations themselves which does not provide strong support for the clustering of individuals into such genetic ‘populations’. The majority of the variance is still within individuals for both geographical and genetic populations in both regions.



**Figure 4.7** – Overall view of clustering analyses in STRUCTURE v.2.3.3 (Pritchard *et al.* 2000) at different spatial scales and across different regions, using no a priori information on sampling locations. In an international analysis, all individuals clustered into two groups (B). After progressively partitioning into regional areas, individuals in southern England clustered into three groups (A) and individuals in Catalonia clustered into two (C).

**Table 4.7** – Analysis of molecular variance (AMOVA) results at a regional scale, comparing the use of geographical and genetic ‘populations’.

	Source of variation	Variance components	% of variation
<b>Southern England</b>			
<i>Geographical populations</i>	Among populations	0.256	10.13***
	Among individuals/within populations	0.273	10.80***
	Within individuals	2.001	79.07***
<i>Genetic populations</i>	Among populations	0.086	3.39***
	Among individuals/within populations	0.455	17.88***
	Within individuals	2.001	78.73***
<b>Catalonia</b>			
<i>Geographical populations</i>	Among populations	0.358	17.23***
	Among individuals/within populations	0.267	12.84***
	Within individuals	1.455	69.93***
<i>Genetic populations</i>	Among populations	0.212	8.50***
	Among individuals/within populations	0.695	27.89***
	Within individuals	1.585	63.61***

\*\*\*p&lt;0.001

## 4.5 DISCUSSION

In contrast to findings in previous studies of UK *E. aurinia* populations (Joyce and Pullin 2001; 2003) this study has uncovered highly significant population differentiation at all spatial scales and significant isolation by distance (IBD), inferred also from patterns of population structure without the use of a priori sampling information. In Catalonia a similar result has been found, although a previous study using AFLP markers gave similar patterns (Wee 2004). It is clear that there are population and region-specific factors that need to be taken into account, and generalisations across species' ranges cannot be made. Conservation measures will need to address the problem of conserving genetic diversity and integrity within and between these regions, whilst also maintaining gene flow to avoid increased extinction risk.

### 4.5.1 Genetic variation

Over the entire study area from Scotland, through Wales and southern England in the UK to Catalonia in northern Spain and southern France, levels of genetic variation were generally low, with a significant heterozygosity deficiency and a mean  $H_E$  overall of  $0.466 \pm 0.248$ . Although Petenian *et al.* (2005) found a mean  $H_E$  of 0.865 across two European populations of *E. aurinia* (which is considerably higher genetic variation) using only five microsatellite markers, Sigaard *et al.* (2008) found a similar level to this study across Denmark using six microsatellite markers (0.491) and Sinama *et al.* (2011) found only a slighter higher mean across a much larger geographical distance including France, Estonia and Portugal (0.628) using 12 markers. The latter two studies did report higher levels of variability with up to 21 alleles per locus and 4 to 34 alleles per locus, respectively, which is much higher than the 4 to 13 found in this study. Sigaard *et al.* (2008) suggest that post-glacial colonization of northern areas from southern refugia could help explain lower variation in northern areas, but in this current study levels of expected heterozygosity in southern England ( $0.501 \pm 0.224$ ) are actually higher than those found in Catalonia ( $0.416 \pm 0.277$ ), although Sinama *et al.* (2011) did find higher levels in France (0.713). The low level of heterozygosity could be the result of several other situations: firstly, high levels of inbreeding within the populations sampled, which would not be unexpected; second, the microsatellite markers themselves may be exhibiting allele size homoplasy (i.e. alleles sharing identical size when amplified, but are not identical by descent – differing in indel and SNP composition); or it may be the result of null alleles, which is a common problem in lepidopteran microsatellite markers (Keyghobadi *et al.* 1999; Van't Hof *et al.* 2005; Zhang 2004).

### 4.5.2 Isolation by distance

Wright (1943) first used the term ‘isolation by distance’ when describing the effect of spatially limited gene flow on population genetic variation. IBD is now accepted to play a major role in the patterns of spatial genetic variation seen across landscapes, especially for moderately mobile species that are more likely to mate with closer neighbours. This study demonstrates variable levels of IBD for *E. aurinia* across different landscapes. Across the entire study region, a large barrier to dispersal in the form of the English Channel describes most of the differences between the two areas, rather than distance alone. Separately in Catalonia, even though many European countries still have extensive traditional land use, there is a highly significant

correlation between genetic distance and geographic distance, whereas across southern England which is extremely fragmented and degraded, the correlation is far weaker until the exclusion of one site. It is worth noting that the exclusion of this site does not markedly change the structuring of the populations, with only a very small decrease in  $F_{ST}$ . Such a difference in IBD even across roughly similar distances (see **Figure 4.7**) could be the result of adaptive features of each region – in Catalonia there is a switch as the species moves southwards onto other host plants than *Succisa* such as *Lonicera* and this has been seen to cause significant differentiation between populations (Descimon *et al.* 2001; Nève 2009; Wee 2004). The dispersal ability of a species will also impact on its adaptation to different habitats – in southern England, although there is a high level of habitat fragmentation the land is generally low-lying which may allow more efficient dispersal events when they do occur (Nève *et al.* 2008).

### 4.5.3 Population structure

In the case of *E. aurinia* in this study where there is highly significant levels of population differentiation and a clear indication of structure in the first instance, it is recommended to use the standard models in the Bayesian programme STRUCTURE v.2.3.3 (Pritchard *et al.* 2000) without a priori sampling information (Hubisz *et al.* 2009; Pritchard *et al.* 2010). The underlying model used by Structure is potentially unsuited to IBD data, but this study reveals the same pattern from the clustering method as the IBD analysis which only adds further confidence to the overall result.

At the international scale of study, there is an  $F_{ST}$  value of 0.22 which indicates severely restricted gene flow. In previous studies in Lepidoptera, the range of recorded  $F_{ST}$  values is 0.004 to 0.291, with a median of 0.053 (for a summary of lepidopteran studies see: Nève 2009). **Table 4.8** summarizes the results of studies on population structure in *E. aurinia* across its distribution, and at varying scales – most showing high levels of differentiation. It is interesting to note that studies in Catalonia using different markers have produced a very similar result, but studies in southern England have not, which could be a result of the low number of alleles used by Joyce and Pullin (2003). Other studies using different markers support their usual agreement (Dhuyvetter *et al.* 2004; Lehmann *et al.* 1996; Meglecz *et al.* 1998).

**Table 4.8** – Studies investigating population genetics of *E. aurinia* at different spatial scales, across different countries, and using different molecular markers. Although studies cannot be directly compared due to the many differences in their approach, it is valuable to acknowledge the similarities (or dissimilarities) in their conclusions.

Country	Scale of study	Marker used	$F_{ST}$	Reference
UK & Europe	International	Microsatellites	0.22	This study
UK	Regional (S. England)	Microsatellites	0.10	This study
Europe	Regional (Catalonia)	Microsatellites	0.17	This study
Europe	Regional (Catalonia)	AFLPs	0.18	Wee (2004)
Denmark	National	Microsatellites	0.16	Sigaard <i>et al.</i> (2008)
China	Regional	DALPs	0.07	Wang <i>et al.</i> (2003)
UK	National	Allozymes	0.16	Joyce & Pullin (2003)
UK	Regional (S. England)	Allozymes	0.05	Joyce & Pullin (2003)
S France	Regional	Allozymes	0.11	Descimon <i>et al.</i> (2001)

AMOVA results suggest that although the majority of genetic variance lies within individuals in all cases, there is significant variance among groups in all cases too, and therefore emphasizes the lack of gene flow and the isolation between populations. Southern England shows a lower level of differentiation than that across Catalonia, which agrees with the isolation by distance found in the latter, and supports the hypothesis that although highly fragmented, there is opportunity for successful dispersal events across the low-lying land in southern England.

#### 4.5.4 Comments on methods and future work

The advent of high throughput sequencing technologies has rapidly increased the amount of publicly available EST data and has encouraged the quick and easy isolation of microsatellite markers at a reasonable cost (Kong and Li 2008; Metzger *et al.* 2011; Mikheyev *et al.* 2010; Sinama *et al.* 2011). This approach is already proving successful for other taxa and even more so when combined with a biotin-enrichment protocol (Dubut *et al.* 2010). Recently this method has been applied to *E. aurinia* and successfully isolated 12 microsatellite loci, although only three showed evidence of transferability among populations and sub-species (Sinama *et al.* 2011). Studies such as this and the current study are slowly forming a valuable set of markers for studying the evolutionary dynamics and conservation genetics of *E. aurinia*.

There are apparent drawbacks of most approaches to studying population genetics, from the molecular marker chosen, the method of isolation, through to the analysis. Although the 454 transcriptome used in this study was constructed from pool of 64 larvae collected from southern England, and hence the developed markers are more monomorphic and show greater deviations from HWE in European populations (see **Chapter 3** for a population by population assessment for each locus), it has been shown that not only are these markers suitably ‘neutral’ (**Figure 4.2**), but they are transferable across geographically distant populations and show the same pattern of population structuring as when using AFLP markers on the same subset of samples (Mikheyev et al, in prep.). Therefore, it can be confidently concluded that these microsatellite markers are suitable for this study, and the approach of allowing the genetic markers alone to identify patterns of population structure, rather than using a priori sampling locations, is validated.

Occasionally, studies using microsatellite markers include a complimentary analysis using mitochondrial DNA to enable slightly different patterns of gene flow to be resolved. However, mtDNA should not be used where the host arthropod harbours maternally inherited microorganisms such as *Wolbachia* (Charlat *et al.* 2009; Shoemaker *et al.* 2000). As *Wolbachia* sweeps through the distribution of a species, the frequency of mitochondria from infected individuals also increases in each population due to the similar mode of transmission used by both. Consequently, mtDNA from infected individuals reaches high prevalence alongside the *Wolbachia* (Hurst and Jiggins 2005; Nice *et al.* 2009). Using mtDNA to infer population structure and evolutionary history would thus be misleading for *E. aurinia* as it is known to harbour a single strain of *Wolbachia* across the UK mainland (See **Chapter 6** of this thesis; Smees *et al.*, in prep.). Joyce and Pullin (2001) demonstrate this in a study of *E. aurinia* across the UK and France, whereby the low diversity and pattern of mtDNA haplotypes showed a similar distribution to allozyme allele frequencies, and no isolation by distance was detected at all.

Another potential marker suggested as a replacement for using microsatellites in difficult genomes, is exon-primed intron-crossing (EPIC) markers (Tay *et al.* 2008) although the slightly lengthy process of their isolation seems to have detracted from their popularity for use in non-model organisms so far. Arguably the best marker

currently available for population genetics studies is SNPs, which either requires the availability of genomic sequence data for the species in question, or large sums of funding. The recent advent of RAD-tagging, employing the use of SNPs and fast throughput sequencing, could be the answer for non-model organisms otherwise lacking in sequence data (Baird *et al.* 2008; Hohenlohe *et al.* 2010).

#### 4.5.5 Conclusions

In general, *E. aurinia* appears to be highly genetically differentiated across all areas of its distribution. Geographical populations show mixed membership to distinct genetic clusters, inferring that there is mixture between these populations, or at least has been in the past, although significant pairwise  $F_{ST}$  values emphasize that current gene flow is very low. Conservation measures for *E. aurinia* will have to address the lack of variation throughout the populations, whilst minimizing the loss of genetic integrity (such as private alleles). For stable, ‘healthy’, genetically diverse populations, whether or not there is evidence of IBD, as long as population size remains stable then general maintenance of habitat and moderate connectivity should allow genetic variation to remain high and some genetic integrity to be kept. For declining, genetically depauperate populations showing evidence of IBD, current management strategies of improving connectivity and habitat quality are no bad thing, but to increase genetic diversity there may be a need for translocations from more diverse populations further away. However, for those declining and depauperate populations that show no evidence of IBD, the likelihood is not that there are many strong dispersers keeping gene flow high (as demonstrated by significant population differentiation at all spatial scales), but that most populations of sedentary individuals have diverged significantly. This results in a worst-case scenario, whereby effective population size may also be low and therefore dramatic conservation efforts are required in the form of re-introductions, habitat improvement and increased connectivity, to restore genetic variation and encourage beneficial genetic integrity.

Often it is suggested that as conservationists, we need to move from site-specific actions to implementing conservation plans at multiple scales across broad landscapes. However, as ideal as this is in the context of mitigating climate change and ensuing changes in distribution, it is apparent that we still need detailed knowledge on individual sites, or regional collections of sites, to ensure strategies are appropriate – as what is best for one area may not be best at all for another (i.e. southern England vs.

Catalonia *E. aurinia* populations). Particular life-history traits, or local adaptations, may confer different vulnerability to habitat fragmentation effects, and hence require different approaches in conservation efforts. Inbreeding and loss of genetic diversity are inevitable in small, fragmented populations of threatened species, so it is crucial to understand both the status of individual populations and the patterns across different regions, to be able to fully and effectively conserve the species at all levels and maintain genetic integrity whilst also encouraging gene flow to heighten genetic diversity.

The difficulty with identifying patterns and causes of species declines is often the lack of power gained in analyses using small sample sizes, yet this is not a situation likely to change as the species we need information about to conserve effectively will always be the species where it is hard to collect a sufficient sample size due to regulations, protected status and low population sizes. Advances in non-invasive techniques (e.g. wing tip tissue samples: Lushai *et al.* 2000) combined with next generation high throughput sequencing methods provide huge potential for future studies of endangered Lepidoptera.

## Chapter 5

### ***Pgi*'s Might Fly: A purported 'gene for' dispersal shows no signal of selection in the marsh fritillary butterfly.**



*'Stay near me--do not take thy flight!  
A little longer stay in sight!  
Much converse do I find in thee,  
Historian of my infancy!'*

*William Wordsworth 1802  
'To A Butterfly'*



## 5.1 ABSTRACT

Species often persist as metapopulations, with fragmented distributions shaped by the combined impacts of environmental change and habitat degradation. Persistence of metapopulations depends on the re-colonisation of habitat fragments following local extinction events. Consequently, it is vital to understand the functionally important genetic variation underlying dispersal. Phosphoglucose isomerase (PGI – an enzyme in the glycolysis pathway) is a well-endorsed candidate gene for dispersal, extensively studied in the Glanville fritillary (*Melitaea cinxia*) and Orange Sulphur butterfly (*Colias eurytheme*), and potentially an adaptive molecular marker for arthropods, whereby variation in *Pgi* would predictably affect fitness across taxa. An initial attempt to sequence the entirety of the gene (estimated to be ~11 Kb) from the threatened marsh fritillary butterfly (*Euphydryas aurinia*) was unsuccessful, but did result in the entire coding region (~1.6 Kb) being obtained. Across 27 sites in the UK, six non-synonymous SNPs (single nucleotide polymorphisms) were discovered within the coding region of this gene, out of 48 SNPs altogether. One of these six non-synonymous SNPs, at amino acid 373, was in very close proximity to two SNPs of importance in *M. cinxia* at amino acids 372 and 375. *C. eurytheme* is also polymorphic at the latter SNP. Field collected larvae from 26 populations across the UK and Catalonia area of northern Spain were genotyped at SNP 373 and showed no evidence of balancing selection occurring, which was found in previous studies on *M. cinxia*. Flight tests in the field were performed and individuals were genotyped to determine whether heterozygotes displayed superior flight performance as also previously indicated. *Pgi* genotype was not found to influence, or correlate with, flight distance or duration, propensity to fly, or the trend of the population sampled from. These findings challenge the emerging perspective that PGI could be used as a molecular marker of selection for arthropods.

## 5.2 INTRODUCTION

Understanding the genetic variation that underlies adaptation to changing environments is an important and fundamental topic (Stapley *et al.* 2010), and critical for species of conservation concern. Dispersal is a key life history trait for Lepidopteran species persisting in metapopulations across landscapes fragmented by agricultural improvements, habitat destruction and degradation (Hanski 1999). Increasing habitat fragmentation is expected to cause quantitative variations in dispersal behaviour, whereby individuals may develop a reluctance to cross habitat patch boundaries (Schtickzelle *et al.* 2006). But, the persistence of metapopulations in fragmented landscapes relies on levels of dispersal that raise rates of colonisation above background rates of stochastic local extinctions (Levins 1969). Dispersal is a complex trait, and is affected by a vast array of behavioural, morphological and physiological traits (Ehrlich 1961; Hanski *et al.* 2002; Hovestadt and Nieminen 2009; Marden *et al.* 2008; Singer and Ilkka 2004), each of which is influenced by environmental variables and may have a genetic basis. Although it may prove difficult to determine the underlying genetic variation that causes variation in dispersal qualities of an organism, it is nevertheless essential in an era of climate and habitat change to be able to predict a species' ability to adapt and move.

Until recently, efforts in the area of conservation genetics focused solely on the utility of neutral molecular markers; those that are not under the influence of natural selection and illustrate only the effects of demography and their own mutational rates. Next-generation sequencing (e.g. Margulies *et al.* 2005) now allows the rapid identification of genetic loci responsible for adaptive evolution (Stapley *et al.* 2010) and opens up possibilities not formerly available to studies on non-model organisms. In the past, authors have generally argued that, aside from the effects of inbreeding, strong demographic and environmental stochastic effects override any underlying genetic influences on population dynamics (Lande 1988). Studies have since challenged this perception by reporting single (candidate) gene effects in the behavioural ecology of natural populations (Fitzpatrick *et al.* 2005). Hanski and Saccheri (2006) provided the first demonstration that molecular variation in a candidate gene could affect population dynamics, by analysing variation in the well-studied glycolytic enzyme phosphoglucose isomerase (*Pgi*) in metapopulations of the Glanville fritillary butterfly (*Melitaea cinxia*) and finding significant effects on population growth.

The enzyme phosphoglucose isomerase (PGI – references to enzyme level variation are denoted by the use of uppercase letters, whereas references to DNA level variation is denoted by lowercase, italicized letters) is just one of many enzymes involved in the metabolic pathway of glycolysis (See **Appendix 2, Figure A1**), which converts glucose and fructose (such as from flower nectar) to pyruvate, whilst also releasing the high-energy compounds ATP and NADH which are required for flight, as well as other essential cellular functions. Glycolysis is consequently a very important metabolic pathway across a wide variety of taxa, including Lepidoptera.

A series of influential papers by Watt and co-authors through the 1970s and 80s on *Colias* butterflies demonstrated evidence for direct natural selection on polymorphisms found at the *Pgi* locus. Electrophoretic variants, or different haplotypes identified by allozymes, were shown to differ in metabolic capacity (Watt 1977) leading to heterosis whereby heterozygotes demonstrated the ability to fly under a broader range of environmental conditions (Watt 1983). Further studies revealed differences in flight ability, survivorship and mating success between *Pgi* variants (Watt 1992; Watt *et al.* 1985; Watt *et al.* 1983). More recent studies by Wheat and co-authors have further proposed the utility of *Pgi* as an adaptive molecular marker for arthropods (Wheat 2010), demonstrating that in the Glanville fritillary butterfly (*Melitaea cinxia*), haplotypes (determined by either allozymes or single nucleotide polymorphisms (SNPs)) gave predictable differences in a variety of fitness and life history traits, including female fecundity and flight metabolic rate (Haag *et al.* 2005; Orsini *et al.* 2009; Saastamoinen *et al.* 2009), as well as the afore mentioned differential population growth rates (Hanski and Saccheri 2006). Niitepõld *et al.* (2009) used harmonic radar to track dispersal movements of adult *M. cinxia* butterflies, confirming the heterozygote advantage found previously by Watt (1983) whereby a fitness advantage is gained by heterozygotes that can fly at lower ambient temperatures than homozygotes. This latter study also confirmed predictions from previous studies (Haag *et al.* 2005; Hanski *et al.* 2004) that *Pgi* genotype affects flight metabolic rate (MR). Individuals with a specific *Pgi* allele were found to have elevated flight MR (Haag *et al.* 2005), which is further linked to the [ATP]/[ADP] ratio in flight muscles of females after forced flight (Hanski *et al.* 2004), suggesting that the increased ability to rephosphorylate ADP is linked to higher dispersal rate in females, which is therefore affected by MR.

As a species widely accepted to persist as metapopulations in fragmented landscapes (Hula *et al.* 2004; Wahlberg *et al.* 2002; Warren 1994) *E. aurinia* has shown dramatic declines in distribution over the past few decades (Asher *et al.* 2001; Fox *et al.* 2010; van Swaay and Warren 1999). At some sites this decline has been recently reversed and trends in density and distribution are now steadily increasing with favourable weather conditions during the flight period and intense habitat management (pers. comm. Butterfly Conservation), although at many sites the steep decline has continued (Smee *et al.* 2011) and could be influenced by the lowering of genetic diversity between highly genetically differentiated populations at all geographical scales of *E. aurinia*'s distribution in the UK (Smee *et al.*, in prep.: **Chapter 4**). Consequently, it would be of great benefit to be able to predict differential success and fitness at the organismal level by studying potential adaptive markers such as *Pgi* at the molecular level. If there are pleiotropic effects of one locus causing individuals of certain genotypes to have increased flight ability at higher temperatures, but to suffer lower fecundity, as found by Watt (1992), there may also be severe consequences for populations in the event of climate change. In terms of conservation, this can help predict how populations will respond to environmental change, and even identify source populations for potential translocations or re-introductions in combination with knowledge gained from neutral markers.

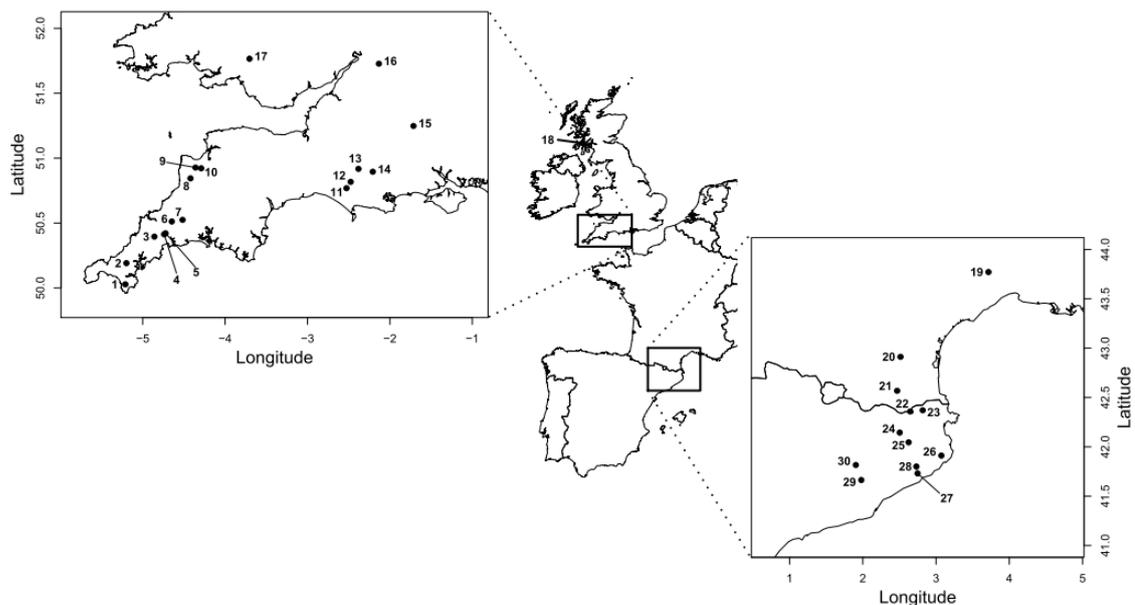
Hence, to compliment previous studies by the authors using neutral molecular markers to investigate levels of genetic diversity and population structure in *E. aurinia* (Smee *et al.*, in prep.: see **Chapter 4**), the current study aimed to survey putative functional variation in the candidate gene for dispersal, *Pgi*. The full-length coding sequence of *Pgi* in *E. aurinia* was cloned to identify non-synonymous single nucleotide polymorphisms (SNPs) across populations spanning the UK mainland distribution of *E. aurinia* as well as populations in northern Spain / southern France (Catalonia region). A single SNP of interest causing a charge change in the amino acid produced was surveyed and variation analysed against flight propensity and ability in the field, in addition to population trend. Results suggest that *Pgi* variants do not yield dispersal related variation, or population growth variation, in this species, and hence that *Pgi* is not suitable for use as an adaptive molecular marker in *E. aurinia*. Limitations, and the need for further studies, are discussed.

## 5.3 MATERIALS AND METHODS

### 5.3.1 SNP genotyping

#### 5.3.1.1 Sample collection

Larvae of the univoltine *E. aurinia* (Lepidoptera; Nymphalidae) live gregariously in silken webs from the date they hatch in early to mid-July, through winter diapause, until the penultimate instar when they separate off to feed individually and then pupate in late Spring (Porter 1981). Across southern England, a single larva was collected from each web (to avoid siblings being included) found over a period of two generations (under Natural England licence numbers 20082745 and 20091722) to gain samples from populations across the entire English distribution (N = 16). In autumn 2008, samples were also collected by volunteers in Wales and Scotland (under Countryside Council for Wales (CCW) licence OTH:SPCA:09:2008 and Scottish Natural Heritage (SNH) licence number 9154, respectively) to facilitate comparisons across the entire UK mainland distribution of *E. aurinia*. European samples (from the Catalonia area of Spain and southern France) of *E. aurinia* were collected by Wee (2004) and Singer, M. (pers. comm.) which allowed further comparisons with populations across the European distribution of *E. aurinia* (N = 12) (**Figure 5.1**).



**Figure 5.1** – Map of samples sites across the UK and Catalonia area of northern Spain / southern France. Population numbers refer to those in **Table 5.1**.

### 5.3.1.2 Sequencing *Pgi* and searching for SNPs

To identify any variation across *Pgi* and possible non-synonymous SNPs that could cause changes in the enzyme PGI produced, and hence changes in the physiology and metabolism of individuals, the coding sequence (open reading frame, ORF) of *Pgi* first had to be obtained. Initially it was attempted to gain the genomic length of *Pgi*, to allow all variation across the gene to be analysed and to allow primers to be designed in intron regions if needed, if SNPs were close to the start or end of exons (see **Appendix 2** for a description of attempts to sequence the entirety of the genomic length of *Pgi*). However, this was unsuccessful.

To obtain the coding sequence of *Pgi*, the RNA of two captive *E. aurinia* larvae was initially used to amplify and sequence the cDNA, following the protocol as reported in the Supplementary Material (**section 5.6**). The ORF of *Pgi* was obtained from a single individual from each of 27 sites across the UK, covering the UK mainland range of the butterfly (**Figure 5.S1**), which were then aligned in Geneious Pro 5.3.6 (Drummond *et al.* 2010) to identify SNPs (**Figure 5.S2**). Six non-synonymous SNPs were found out of a total of 48 SNPs across the ORF of *Pgi*. One non-synonymous SNP in particular, at amino acid 373 on exon 9, was very close to the location of the SNP(s) of interest in *M. cinxia* at 372 and 375. Another SNP, only 100 bp in the 5' direction, appeared to be in linkage with the first, and so it was decided that it would be just as informative to focus on the one SNP rather than amplify and analyse both. This was also due to the logistics of either amplifying across the 2 exons altogether, which had so far been unsuccessful (See **Appendix 2**), or designing new primers for the second SNP on exon 8 and doubling the effort and expense required for apparently no further gain.

The non-synonymous SNP at amino acid 373 causes a change in a guanine nucleotide to an adenosine nucleotide in some individuals, causing the amino acid produced to change from a non-polar glycine (G) to a negatively charged (polar) glutamic acid (E), which would have an impact on the biochemistry of the final dimer produced.

### 5.3.1.3 Genotyping all samples at the SNP of interest

Each larva collected from southern England, Scotland and Wales was starved for 24 hours to remove plant material from the digestive tract and then dissected to remove the midgut and potentially any parasitoids, before then being snap frozen. Following the methods of Martinez-Torres *et al.* (1998) (with some modifications – see **Appendix 1**),

genomic DNA was extracted, diluted to a working concentration (~15 ng/μl), and stored at -20°C until required. DNA from samples collected in Europe was extracted from the heads of larvae using DNeasy Tissue kits (Qiagen) (Wee 2004).

Primers for the SNP on exon 9 were as follows: M13\_PGIexon9\_F1: 5' – TGT AAA ACG ACG GCC AGT ATT CGC AGC GTA TTT CCA AC – 3' and M13\_PGIexon9\_R1: 5' – CAG GAA ACA GCT ATG ACC CTC GTC CCC TGG TGT ATC A – 3'. Underlined bases form the M13 tails added for ease of sequencing directly from PCR. Each 50 μl PCR mix contained approximately 1.5 ng genomic DNA, 25 μl *Taq* PCR Master Mix (Qiagen), and 10 pmol of both the forward and reverse primers. The PCR cycling conditions were as follows: 94°C for 1 min; followed by 35 cycles of 94°C for 30 secs, 58°C for 30 secs, 72°C for 1 min; followed by a final extension step of 72°C for 5 mins. Amplified products were first visualised on a 1.2% agarose gel for confirmation of amplification, then sequenced using an ABI3130 Genetic Analyzer (Applied Biosystems) after a clean-up with ExoSAP to remove residual nucleotides. Forward and reverse sequences were assembled in Geneious Pro v5.3.6 (Drummond *et al.* 2010) and SNP genotypes manually inferred from all 569 sequences (see **Table 5.1**). Dubious results were re-sequenced.

As a control to validate genomic DNA extractions, primers for a housekeeping gene *Elongation Factor 1-Alpha (EF1α)* were also tested on samples in the same PCR reactions (Eauri-EF1a-F: 5' – CCT GGC CAC AGA GAT TTC AT – 3' and Eauri-EF1a-R: 5' – CAC GAC GCA ATT CCT TAA CA – 3') and run on the same 1.2% agarose gels. These primers were designed originally from a 454 transcriptome of *E. aurinia* using the *EF1α* sequence from *Bombyx* as a BLAST query.

#### 5.3.1.4 Assessing whether PGI is under selection

For samples that had been genotyped at SNP 373 but also genotyped at nine microsatellite loci (N = 468) in a previous study by the authors (Smee *et al.*, in prep.: see **Chapters 3** and **4**), the programme LOSITAN (Antao *et al.* 2008; Beaumont and Nichols 1996) was used to identify putative loci under selection. LOSITAN uses an  $F_{ST}$  outlier method, plotting  $F_{ST}$  values against expected heterozygosity ( $H_E$ ), to determine possible loci under either positive or balancing selection, or that are neutral. Each population was also tested for departures from HWE due to excess heterozygotes.

The McDonald-Kreitman test (Egea *et al.* 2008; McDonald and Kreitman 1991), was also carried out to determine if balancing selection was acting across the entire coding sequence of *Pgi*. This test compares patterns of polymorphism within a species with divergence between closely related species using ‘silent’ synonymous SNPs and ‘replacement’ non-synonymous SNPs in protein-coding regions such as *Pgi* in this case. As in **Section 5.3.1.2**, the ORF of *Pgi* from a single individual from each of 27 sites across the UK were used for the within species comparison, and an allele (GenBank Accession: EU888473) of *Pgi* from a closely related species, *Melitaea cinxia*, was added for the between species comparison. The online webtool developed by Egea *et al.* (2008) was used to carry out the analysis.

### 5.3.2 Flight assays in the field

#### 5.3.2.1 Site choice and timing

Sites chosen for flight assays were restricted to southern England, primarily due to accessibility for the authors. These comprised of a mixture of the two main biotypes that *E. aurinia* occupies in the UK: damp, neutral or acidophilous grasslands and dry, calcicolous grasslands (Barnett and Warren 1995). Some sites that would have also been favourable to visit were not suitable due to their topography, which would have made tracking the adult butterflies across distances very difficult. The seven sites covered are highlighted in **Table 5.1**.

Flight assays in the field were carried out after the peak flight period at each site, to allow females time to have laid at least their first egg batch (which might be their sole egg batch) and hence so as not to reduce individual fitness and likelihood of mating and laying through disturbance of populations. The peak flight period at each site was discerned from both landowners and inspecting the first few individuals caught to identify gender (males emerge earlier) and evidence of wear on the scales of the wings. If the adult was freshly emerged, its wings should be intact and the colours should be bold and bright, whereas once an individual is a few days old the edges become damaged and colours are much more faded as the scales have worn off.

#### 5.3.2.2 Flight assays

At each site, butterflies were caught and then temporarily housed in a mesh pop-up bug dorm. Individually, but in close succession to keep environmental conditions as similar as possible, they were taken out and a single back leg was taken and placed immediately

in 100% ethanol for subsequent genomic DNA extraction, as in **Appendix 1**. Each adult was also sexed. Providing the environmental conditions were good (> 18°C and minimal wind), adults were then ‘dropped’ from an approximate height of 2 m and allowed to fly away. The number of times an individual was ‘dropped’ before actually flying off was used as a proxy for their propensity to fly – if after ten drops they had still not flown, they were excluded from the test. Once an individual had taken flight, a timer was started until the point at which they landed for the first time since being dropped, and this distance was then measured as a proxy for flight ability. Unfortunately, other measures of flight in the field could not be used as the adults could not be followed further than this initial flight. Although generally a sedentary butterfly, the adults will fly very fast once released, as if escaping from a predator, and dart in all directions. This also then produces confusion between the released individual and other butterflies also flying in the area at that time.

### 5.3.3 Statistical analysis of SNP genotype against ecological data

When analysing the genotypes of individuals at SNP 373 against flight data and population trends, analyses were restricted to sites in southern England due to accessibility for the authors, availability of data and a lack variation in the SNP over the Catalonia populations (see **section 5.4.1**).

Population trend was determined from a combination of larval web surveys conducted by the authors over three years along with data collected by landowners (usually Butterfly Conservation, Cornwall/Devon Wildlife Trust or Natural England) or contracted surveyors, in the years preceding and including 2009 when the samples were collected. Over the time span of 5 years in most cases, a population was deemed to be either: ‘stable’ if it had varied in the number of larval webs found each year, but remained relatively constant; ‘declining’ if the number of larval webs had consecutively decreased across the time span; or ‘increasing’ if the numbers of larval webs had consecutively increased over the years measured.

All statistical analyses were carried out using R 2.12.1 (R Development Core Team, 2010). For flight data (N = 99), distance and duration were log-transformed and analysed as response variables against the fixed effects and interaction of genotype (AA, AG or GG) and gender (male or female) in linear mixed effects models, including the population from where each individual was sampled as a random effect. Dispersal

propensity, inferred from the number of ‘drops’ taken until flight, was analysed using a censored survival analysis (Crawley 2007). The response variable, number of drops before flight, was modelled against explanatory variables genotype, site (each of seven populations) and gender. Censoring allowed butterflies that had not flown before terminating the trial (10 drops maximum) to contribute to the ‘flight’ hazard function. Exponential and Weibull error distributions were compared for best fit. To analyse the potential effect of population trend on genotype frequencies at each population, the frequency of glycine alleles (G) in each population was analysed in a general linear model with quasibinomial errors, against the trend of each population. The difference between the actual number of heterozygotes and the expected number of heterozygotes in each population if they were in Hardy-Weinberg equilibrium (HWE) was also analysed in a general linear model against population trend. Response variables were log-transformed where needed, to reduce skew and improve homogeneity of variance in residuals which was checked by visual inspection of residual against fitted values plots. Models were also checked for Gaussian errors. In all models, significance of explanatory variables was tested using likelihood ratio tests, using a significance threshold of  $p < 0.05$ . Simplification of mixed effects models required conversion from restricted maximum likelihood to maximum likelihood fits (Crawley 2007). To decipher significant differences in effect among the three genotypes on response variables, models were re-levelled to use a different genotype as the baseline comparison (GG, instead of AA which is alphabetically earlier). Paired-t-tests were then reported from the three pairwise comparisons, and tested for significance using Bonferroni correction (Rice 1989).

## 5.4 RESULTS

### 5.4.1 SNP variation

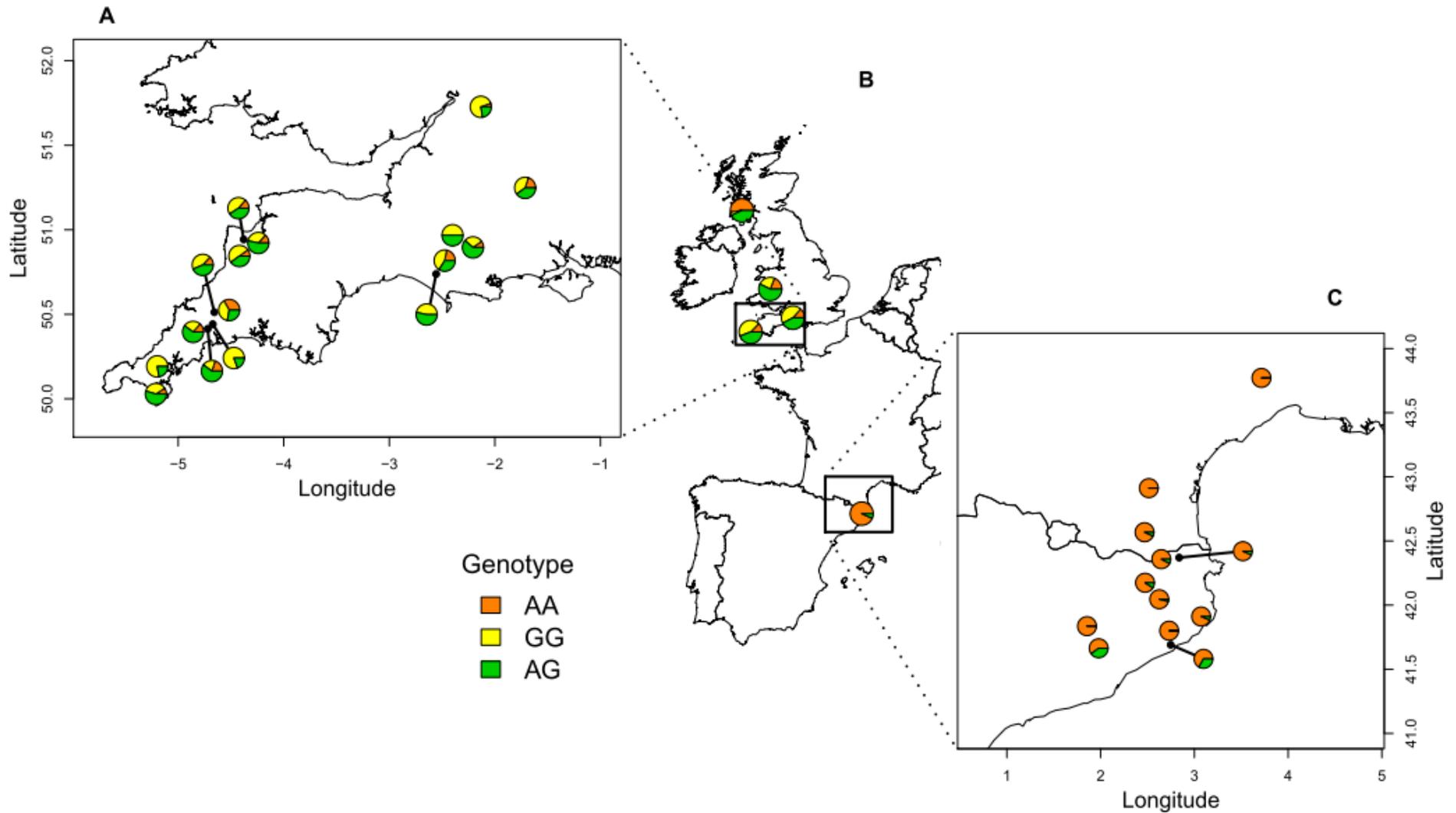
Throughout southern England, Wales and Scotland, there is much variation in the genotype frequencies found (**Table 5.1; Figure 5.2A and B**). In Catalonia, the glycine homozygote (GG) is completely absent from the sample (see **Table 5.1; Figure 5.2C**), and although several of the populations in Catalonia are monomorphic for the glutamic acid homozygote (AA), there is only one population over all 30 showing significant deviation from HWE. This deviation is not due to an excess of heterozygotes, but an excess of glycine homozygotes (**Table 5.1- Population 12**).

**Table 5.1** - The observed frequencies of each genotype at SNP 373 across 30 sampling sites.

Region	Map Number	Population	N	Genotype			$H_O$	$H_E$	$\chi^2$
				AA	GG	AG			
S. England	1	<b>Lizard</b> †	44	4	16	24	0.55	0.46	1.40
S. England	2	Stithians	9	0	7	2	0.22	0.20	0.14
S. England	3	Goss Moor	15	2	4	9	0.60	0.49	0.74
S. England	4	<b>Breney Common</b> †	30	6	6	18	0.60	0.50	1.20
S. England	5	<b>Redmoor</b> †	26	0	21	5	0.19	0.17	0.29
S. England	6	Colvannick	23	3	10	10	0.44	0.45	0.04
S. England	7	Carkeet	18	6	7	5	0.28	0.50	3.53
S. England	8	Dunsdon	20	2	10	8	0.40	0.42	0.05
S. England	9	Volehouse	22	3	10	9	0.41	0.45	0.18
S. England	10	Stowford	15	2	5	8	0.53	0.48	0.19
S. England	11	<b>Southfield</b> †	15	0	7	8	0.53	0.39	1.98
S. England	12	<b>Giants Hill</b> †	50	11	22	17	0.34	0.48	4.07 *
S. England	13	<b>Lydlinch</b> †	8	0	4	4	0.50	0.38	0.89
S. England	14	<b>Hod Hill</b> †	31	3	9	19	0.61	0.48	2.32
S. England	15	Salisbury	20	4	8	8	0.40	0.48	0.56
S. England	16	Strawberry Bank	18	1	13	4	0.22	0.28	0.72
Wales	17	Varied	21	4	5	12	0.57	0.31	0.44
Scotland	18	Varied	21	11	1	9	0.43	0.50	0.25
Catalonia	19	Pic St Loup	8	8	0	0	0.00	0.00	-
Catalonia	20	Col de la Redoulade	9	9	0	0	0.00	0.00	-
Catalonia	21	Col del Forn	15	14	0	1	0.07	0.06	0.02
Catalonia	22	Coustouges	13	12	0	1	0.08	0.07	0.02
Catalonia	23	Darnius	15	14	0	1	0.07	0.06	0.02
Catalonia	24	Can Jorda	11	10	0	1	0.09	0.09	0.02
Catalonia	25	LaBaraca	53	51	0	2	0.04	0.04	0.02
Catalonia	26	Mas Calc	16	15	0	1	0.06	0.06	0.02
Catalonia	27	Tordera	6	4	0	2	0.33	0.28	0.24
Catalonia	28	Sils	6	6	0	0	0.00	0.00	-
Catalonia	29	Col Estelales	5	3	0	2	0.40	0.32	0.31
Catalonia	30	El Guix	6	6	0	0	0.00	0.00	-

†**Populations** included in the flight assays. Sample size (N); observed ( $H_O$ ) and expected ( $H_E$ )

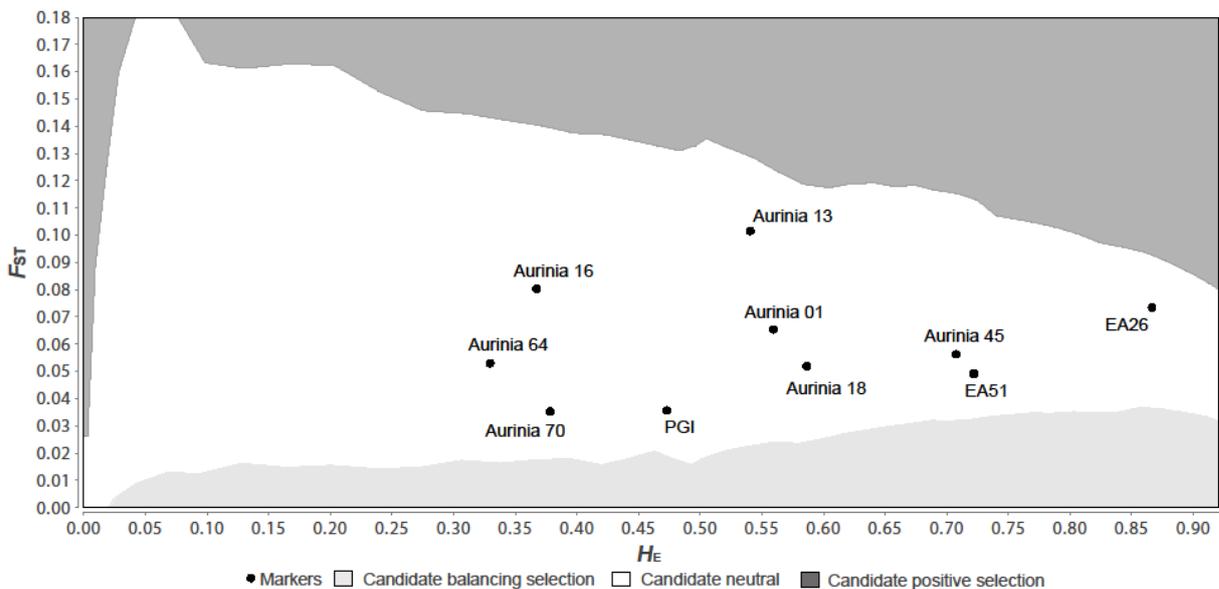
heterozygosities and chi-squared values ( $\chi^2$ ) of the tests for departure from HWE, based on one degree of freedom and calculated using a web-tool developed by Rodriguez *et al.* (2009). \* $P < 0.05$ .



**Figure 5.2** – Genotype frequencies of Pgi SNP 373 across the study area, including (A) southern England, as well as (B) Wales and Scotland and (C) Catalonia in northern Spain / southern France. Populations in the UK appear to have a mixed distribution, although frequencies in Scotland do show a closer similarity to those in Catalonia, continental Europe, with a higher frequency of glutamic acid homozygotes (AA).

### 5.4.2 Is PGI under selection?

From the results of the  $F_{ST}$  outlier method used by the programme LOSITAN (Antao *et al.* 2008; Beaumont and Nichols 1996) of the genotypes at *Pgi* alongside genotypes at nine microsatellite loci (Figure 5.3), it would appear that *Pgi* is a putative neutral marker in these samples of *E. aurinia*. Under both the infinite alleles mutation model (IAM) and the stepwise mutation model (SMM) the result is the same and *Pgi* is well within the boundaries of the ‘neutral’ region.

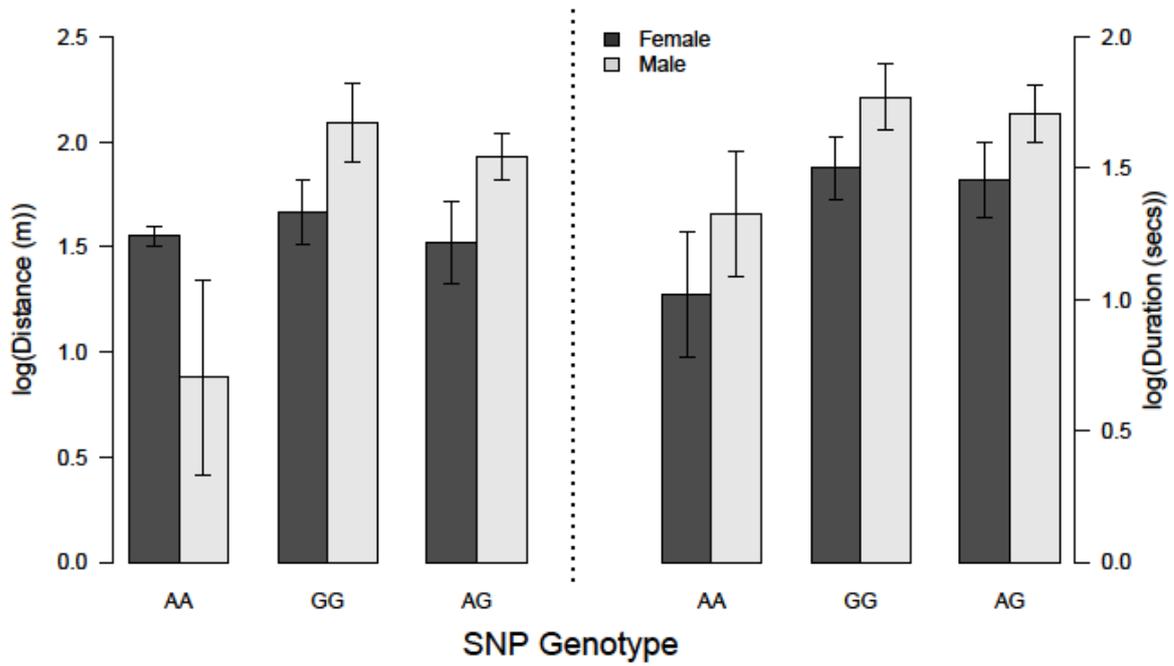


**Figure 5.3** - LOSITAN results for all samples in the UK for which there was data for both microsatellites and *Pgi* ( $N = 306$ ). This workbench uses an  $F_{ST}$  outlier method to detect potential loci under selection. Under a stepwise mutation model (SMM), all loci were found to be candidate neutral markers – located in the white region of the figure. The dark grey region above indicates where candidate positive selection markers would be found, and the lighter region below indicates where candidate balancing selection markers would be found. Under an infinite alleles model of mutation (IAM), the result is the same.

The McDonald-Kreitman test (Egea *et al.* 2008; McDonald and Kreitman 1991) comparing within species polymorphisms (27 samples) to between species divergence (using *M. cinxia* as a comparison) by analysing the ratio of ‘silent’ synonymous SNPs to ‘replacement’ non-synonymous SNPs, found no significant difference between the two ratios ( $\chi^2 = 0.043$ ,  $p = 0.51$ ) indicating that *Pgi* in *E. aurinia* is putatively neutral.

### 5.4.3 Genotypic effects on ecological variables

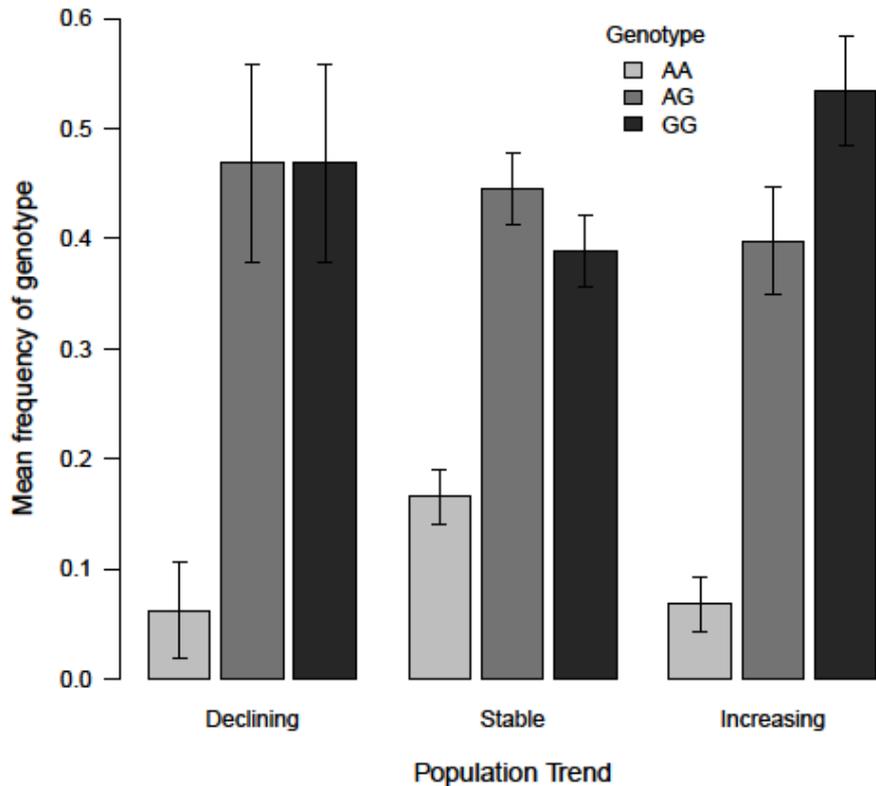
Out of the 99 flight assays performed, only 8 individuals were homozygote for glutamic acid (AA) at SNP 373, with 44 glycine homozygotes (GG) and 47 heterozygotes (AG).



**Figure 5.4** – Logged means  $\pm$  standard errors of distance flown (left panel) and duration of flight (right panel) for butterflies of differing genotypes. The genotype of each individual did not affect the distance or duration of flight; however, males in general did fly further and for longer.

At several different natural populations of *E. aurinia* in southern England ( $N = 7$ ), the genotype of individuals did not affect their flight ability, as measured by the distance flown on a single flight before alighting ( $\chi^2_2 = 4.55$ ,  $p = 0.10$ ; **Figure 5.4** – left panel). After Bonferroni correction for multiple tests (Rice 1989), none of the pairwise tests between genotypes were significant at the corrected level of  $p < 0.0167$ . Homozygotes of glutamic acid (AA) did not fly a significantly shorter distance than heterozygotes ( $t_{85} = 1.25$ ,  $p = 0.22$ ) or glycine homozygotes (GG;  $t_{85} = 2.03$ ,  $p = 0.05$ ), and the latter two did not fly significantly different distances ( $t_{85} = 1.52$ ,  $p = 0.13$ ). Similar, non-significant results ( $\chi^2_2 = 3.04$ ,  $p = 0.22$ ) were found for the effect of genotype on the duration of the same single flight (**Figure 5.4** – right panel). Glutamic acid homozygotes did not fly for a shorter time than either glycine homozygotes ( $t_{85} = 1.59$ ,  $p = 0.12$ ) or heterozygotes ( $t_{85} = 1.55$ ,  $p = 0.13$ ), and the latter two did not fly for a

significantly different length of time ( $t_{85} = 0.05$ ,  $p = 0.97$ ) after Bonferroni correction for multiple tests (Rice 1989). However, in both instances there was a significant effect of gender with males flying further ( $\chi^2_1 = 4.40$ ,  $p = 0.04$ ) and for longer ( $\chi^2_1 = 5.28$ ,  $p = 0.02$ ) – **Figure 5.4A and B**.



**Figure 5.5** – Population trend did not have an effect on the frequency of each genotype at a site, although there were significantly fewer glutamic acid homozygotes (AA) across all population trends.

Population trend does not influence the frequency of glycine alleles in each population ( $F_{2,13} = 2.05$ ,  $p = 0.17$ ) or the difference in the number of heterozygotes found compared to what would be expected in populations under HWE ( $F_{2,13} = 1.14$ ,  $p = 0.35$ ). Therefore, genotype frequencies are not linked to population trend (**Figure 5.5**), although the number of glutamic acid homozygotes found across all population trends is significantly lower ( $p < 0.001$ ).

Dispersal propensity, inferred from the number of ‘drops’ taken until an individual flew, was also not significantly affected by the genotype of the individual ( $\chi^2_2 = 1.40$ ,  $p = 0.50$ ), the gender ( $\chi^2_1 = 0.75$ ,  $p = 0.39$ ), or the site of provenance ( $\chi^2_6 = 10.54$ ,  $p = 0.10$ ),

with the majority of butterflies flying on the first drop. The Weibull error distribution (with scale parameter 0.83) provided a better fit than the Exponential ( $\chi^2_1 = 14.14$ ,  $p < 0.001$ ), suggesting genuine heterogeneity in the propensity of butterflies to fly when disturbed.

## 5.5 DISCUSSION

The findings from this study challenge the emerging perspective that PGI could be used as an adaptive molecular marker for arthropods, whereby variation in *Pgi* affects population level fitness. Although a non-synonymous SNP was found (at amino acid 373) close to the location of two SNPs in the coding region of *M. cinxia* (372 and 375) as well as in *C. eurytheme* (375), no evidence of balancing selection was detected, with no heterozygote excess or advantage across any of the variables measured, and no relationship between genotype and flight in the field was observed. More studies on PGI in *E. aurinia* are now essential to confirm and expand on the findings of this current study.

### 5.5.1 Variation in *Pgi* genotype frequencies

Although variation across the coding sequence of *Pgi* in *E. aurinia* showed similar levels to that in *M. cinxia*, (6 non-synonymous SNPs out of 48 SNPs altogether; and 10 non-synonymous SNPs out of 55 SNPs altogether, respectively), no balancing selection was detected when using an  $F_{ST}$  outlier method (Antao *et al.* 2008). Patterns of geographical variation were also surprising and further dispelled the possibility of heterozygote advantage (see **Figure 5.2**). In Catalonia, encompassing parts of northern Spain and southern France, populations illustrated incredibly low variation at the *Pgi* SNP 373, with 11 out of the 163 individuals being heterozygote, some populations completely homozygote for the glutamic acid variant, and a complete lack of glycine homozygotes (**Table 5.1**). It is possible that adaptive markers in the form of SNPs are non-transferable between subspecies or geographically distant populations, as this has not been tested in *M. cinxia*, and perhaps there is greater variation elsewhere in the coding region. It may be possible that either the homozygote variant, or glycine allele, is slowly being removed from these populations through the process of genetic drift, or that there is some selective advantage to the glutamic acid allele.

An obvious difference between the southern England populations and the Catalonia populations would be the climate they experience. There is a growing number of studies to date that demonstrate differential success of *Pgi* variants at varying temperatures, with *C. eurytheme* heterozygotes able to fly at a broader range of temperatures (Watt 1983), as well as *M. cinxia* heterozygotes being able to fly at lower ambient temperatures than homozygotes, but homozygotes performing better at higher temperatures (Niitepold 2010). In studies on a montane beetle, *Chrysomela aeneicollis*, different populations dominated by different PGI alleles seem to be locally adapted to temperature, and differ in their expression of a heat shock protein, Hsp70, used as an indicator of thermal tolerance (Dahlhoff and Rank 2000; Neargarder *et al.* 2003). It may therefore be that the glutamic acid homozygote in *E. aurinia* performs better at higher temperatures, hence its dominance in the warmer climes of Catalonia, and its low frequency in populations of southern England. However, this would not explain the high frequency in Scotland, whereby the glutamic acid homozygote displays a much higher frequency than would be expected by this hypothesis. Alternatively, as suggested by Joyce and Pullin (2001), the apparent higher similarity between Scottish populations and European populations of *E. aurinia* could be an artefact of post-glacial colonisation from southern refugia. This lower differentiation between Scottish populations of *E. aurinia* and populations in Catalonia was confirmed in a study by Smee *et al.* (in prep.: **chapter 4**) using microsatellite markers to infer population structure, which found higher differentiation between southern England populations and Catalonian populations, than between Scottish populations and Catalonian.

Historically, allozymes of PGI have long been used as putatively neutral markers in population genetics studies (e.g. Joyce and Pullin 2001; Joyce and Pullin 2003; Nève *et al.* 2000). Doubt concerning the neutrality of certain allozyme markers has increased as allozymes have been shown to be under the influence of natural selection in some cases and thus not ‘neutral’ (Begun and Aquadro 1994; Goulson 1993; Wells *et al.* 2009). However, as commented on by Hedrick (2004), due to the low effective population sizes in endangered species, genetic variants are more likely to be neutral as genetic drift plays a more important role than selection on allele frequencies. It is consequently possible that the distribution of *Pgi* genotypes across isolated populations of *E. aurinia* reflect patterns more indicative of neutral variation than adaptive variation.

### 5.5.2 Relating *Pgi* genotypes to flight and population trend

There was no significant effect of genotype on dispersal distance, duration, propensity, or even population trend. The main findings were that in general, and regardless of genotype, males fly further and for longer. This was not unexpected, as many studies have previously reported higher mobility in males (Hovestadt and Nieminen 2009; Junker and Schmitt 2010). When examining the role of population trend on the occurrence of *Pgi* genotypes across *E. aurinia* populations, there is a general difficulty in determining ‘old’ and ‘newly established’ populations to thoroughly answer questions relating to population dynamics. Unlike the *M. cinxia* metapopulation of ca. 4,000 discrete patches, of which approximately 500 are occupied in any given year, and for which there exists a long term dataset detailing the high population turnover each year of approximately 100 extinctions and equivalent colonisations (Hanski 1999; Hanski *et al.* 1995; Nieminen *et al.* 2004; Saastamoinen 2007), there doesn’t exist such a scenario for the vulnerable *E. aurinia*. Due to its high priority on conservation targets and status as a highly protected species in both the UK and Europe (Fox *et al.* 2010), even collecting adequate sample sizes from an adequate number of populations is not easy. The level of knowledge regarding fine-scale population dynamics at each site is also much lower as not all sites are monitored every year, and apparent ‘colonisations’ may be the result of very low densities of *E. aurinia* surviving at a site undetected, and then entering a phase of sudden population growth if environmental factors become suitable, for instance.

### 5.5.3 Further studies needed for *E. aurinia*

To confirm the above non-significant findings, many more studies on *E. aurinia* are needed, including further measures of dispersal ability and propensity, in combination with studies on metabolic rate, which have received great attention in other species. Although studies on *M. cinxia* have demonstrated higher flight metabolic rate in females from newly established, isolated populations combined with higher [ATP]/[ADP] ratios (Haag *et al.* 2005; Hanski *et al.* 2004), more studies are needed to confirm these findings in other species. It could be argued that a higher flight metabolic rate could be detrimental to a species dispersing across a fragmented landscape. A high propensity to fly initially is an under-emphasised trait of importance, followed by the characteristic in an adult to be ‘thrifty’ and maintain a low metabolic rate during flight, therefore using up energetic resources more slowly and being able to fly further and for longer,

consequently covering larger distances and improving the chance of locating suitable habitat patches to colonise.

Wheat *et al.* (2010) used three other genes involved in metabolism (malate dehydrogenase, *Mdh*; isocitrate dehydrogenase, *Idh*; and glyceraldehydes-3-phosphate dehydrogenase, *Gapdh*) to also evaluate nucleotide variation in samples also typed for *Pgi*. Results showed significantly lower levels of variation in these three genes in comparison to *Pgi*, whereby none had any non-synonymous variation at all. Although the current study has incorporated a comparison with microsatellite markers genotyped for the same collection of samples, it would offer further insight into the causes of variation at the *Pgi* locus if other loci involved in metabolic processes were also investigated across the same broad geographical range.

#### 5.5.4 Conclusions

The current study has been conducted on a much larger geographical scale than the previous studies on a sister checkerspot butterfly, *M. cinxia*, which covered an area of 50 x 70 km across the Åland Islands in Finland (Hanski 1999; Hanski *et al.* 1995; Nieminen *et al.* 2004; Saastamoinen 2007) compared to the few thousand km<sup>2</sup> covered here. Although a non-synonymous SNP was found in close proximity in the coding region of *Pgi* in *E. aurinia* to SNPs of interest in *M. cinxia* that show balancing selection and heterozygote advantage, these findings were not replicated here and no evidence for selection was found. *Pgi* allozymes were also found to be polymorphic in only three out of 26 populations of *Procllossiana eunomia* (Nève *et al.* 2000), which adds increasing doubt to the use of *Pgi* as a molecular marker for selection in arthropods.

It is clear that in the face of climate change, habitat change, and severe declines in biodiversity; more reliable knowledge is required for the successful conservation of these delicate and declining species. Model predictions for the 21<sup>st</sup> Century indicate a dire future for terrestrial species, with variation in extinction rates explained firstly by the degree of land-use and climate change (aggressive mitigation in both areas could reduce extinction rates, but is realistically unlikely) and secondly by the lack of understanding about species ecology and in particular, migration rates (Pereira *et al.* 2010). This latter influence on species extinction can, and will be, mitigated. An important step towards an understanding of migration and dispersal would be the detection of major genes ‘for’ dispersal. A search for such ‘adaptive’ genes would yield

a mixture of important alleles of value to conservation management programmes, alongside a set of more neutral molecular markers for population genetic analysis, as both neutral and adaptive genetic variations have their place in studies aimed at the conservation of declining species such as *E. aurinia* (Hedrick 2004; Holderegger *et al.* 2006; Primmer 2009). Neutral molecular markers cannot inform conservationists on the adaptive or evolutionary potential of their species of concern, but can inform on integral processes of gene flow, inbreeding, migration, and more. Adaptive molecular markers can provide the missing essential information into the general health and adaptive potential of populations across different environmental situations (Hedrick 2004; Primmer 2009). The combination of neutral and adaptive markers promises to be of extreme benefit in the conservation of declining species such as *E. aurinia*, even if *Pgi* is not under selection in this species.

## 5.6 SUPPLEMENTARY MATERIAL

### 5.6.1 RNA isolation and cDNA synthesis

RNA was extracted separately from two captive-bred larvae using TRIzol (Invitrogen) with any contamination of genomic DNA removed by DNase treatment (TURBO DNase, Ambion) for 30 min at 37 °C. Following the manufacturer's protocol, the RNeasy MinElute Clean up Kit (Qiagen) was used to further purify the RNA. Full-length, enriched, cDNAs were generated from 2 µg of total RNA using the SMART PCR cDNA synthesis kit (BD Clontech) following the manufacturer's protocol. Reverse transcription was performed with the PrimeScript reverse transcription enzyme (Takara) for 60 min at 42 °C and 90 min at 50 °C.

### 5.6.2 Sequencing of *Pgi* cDNA

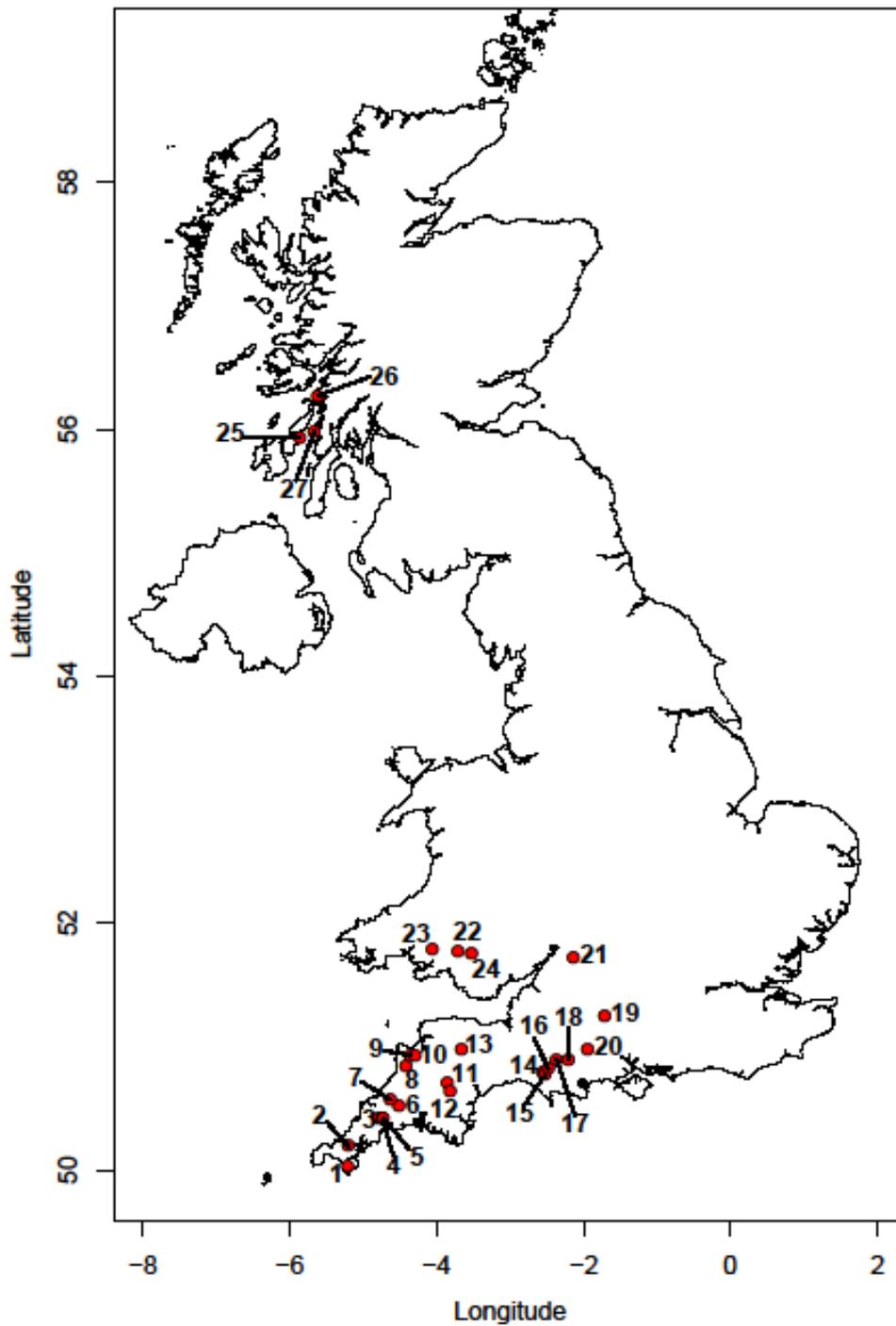
As *Pgi* is a highly conserved gene, the protein sequence from *Colias eurytheme* (GenBank Accession number ABB00833.1) was used in a BLAST (Basic Local Alignment Search Tool) against an assembled 454 transcriptome of *E. aurinia*, resulting in positive hits to four contigs containing the corresponding sequence. From these contigs, specific primers for the *Pgi* open reading frame (ORF) in *E. aurinia* were designed (*pgi\_F1*: 5' – TTT TTG CAC ACC GAG AAC AC – 3' and *pgi\_R1*: 5' – CAA GTT GCT TGC CAA GTT CA – 3'). Two 50 µl polymerase chain reactions (PCR) were set up with these primers and the prepared cDNAs, with an annealing temperature of 55°C and an extension step of two minutes as the expected length of the *Pgi* ORF was 1.6 Kb. After visualising the PCR products on a 1.2% agarose gel, they were cloned into the pCR4 TOPO/TA vector (Invitrogen) and ligations transformed into One Shot TOP10 *E. coli* competent cells. A PCR colony screen was carried out for each to confirm that the *Pgi* sequence was indeed inserted into the vector, and then three positive colonies from both were picked and incubated overnight at 37°C in LB medium containing 0.1 mg/ml Ampicillin. A QIAprep Spin Miniprep Kit (Qiagen) was used to purify the plasmid DNA and the *Pgi* insert was subsequently sequenced using an ABI 3130 Genetic Analyzer (Applied Biosystems).

The resulting 1,612 bp sequence was confirmed as *Pgi* by a BLAST against the NCBI databases. The sequence aligned between positions 1 and 1572 of the coding sequence of *Colias eurytheme* (Genbank DQ205063.1) and showed 78.8% nucleotide identity. From this partial sequence, gene-specific primers were designed for rapid amplification

of cDNA ends (RACE) with a high annealing temperature (~70°C) to increase specificity and allow ‘touch-down’ PCR (PGI\_3RACE\_F: 5’- CGC TGC TGG GTG TGT GGT ACG G - 3’ and PGI\_5RACE\_R: 5’ – TAG AAG GCG TGC TGC CCG TTG G - 3’). Using a SMART RACE cDNA Amplification Kit (Clontech), the 5’ and 3’ ends of the *Pgi* cDNA were amplified, cloned using the same protocol as previously described, and then sequenced. This yielded the remaining sequence of the *Pgi* ORF, including 85 bp of the 5’ untranslated region (UTR) and 484 bp of the 3’ UTR, almost to the poly-A tail.

### 5.6.3 Single nucleotide polymorphisms (SNPs) in *Pgi* across 27 sites

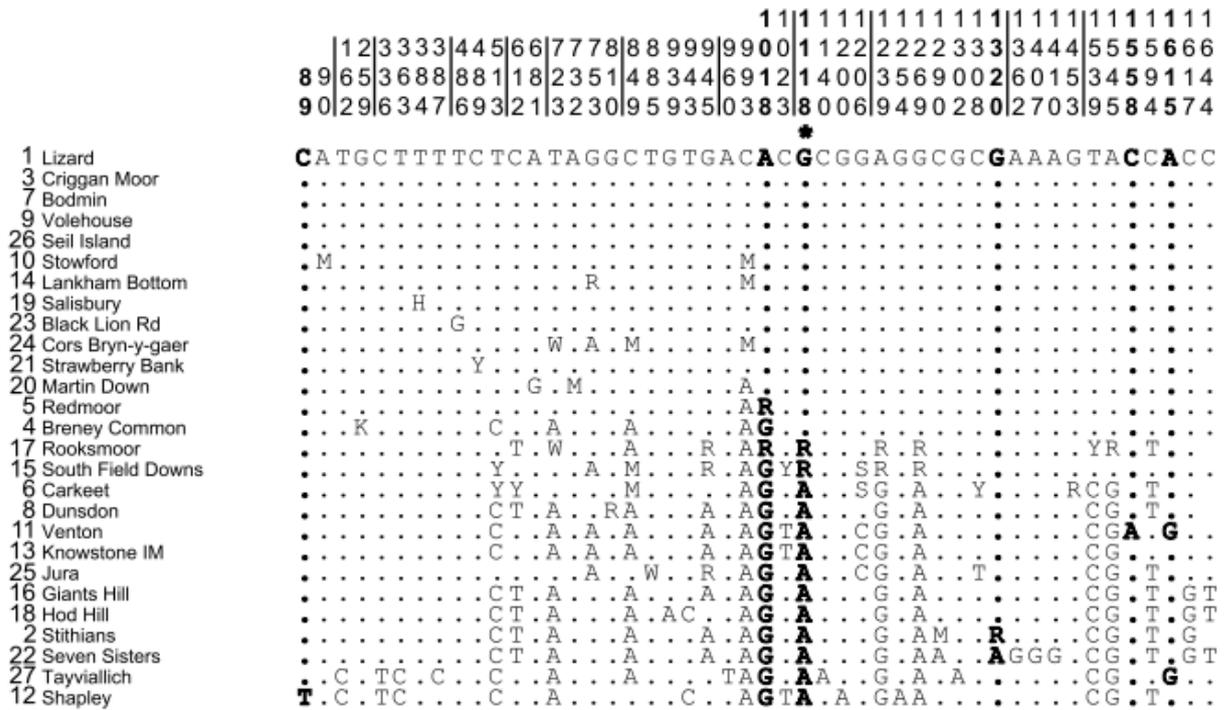
Using an annealing temperature of 55°C and an extension time of 2.5 mins in 30 PCR cycles, sequences of *Pgi* cDNA were obtained from an individual larva from each of twenty-seven sites across the UK (**Figure 5.S1**), encompassing the majority of *E. aurinia*’s distribution on mainland UK. The primer sequences used for the amplification were manually re-designed from the previously synthesized cDNA so as to start at the very beginning of exon 1, and finish at the very end of exon 12. Both were also designed with M13 sequencing tags added on to the 5’ end so that the PCR product could be sequenced directly without the need for the more expensive and time-consuming process of cloning each time (M13\_PGI\_F: 5’ – TGT AAA ACG ACG GCC AGT ATG GAA CCT AAG GTG AAC TTG – 3’ and M13\_PGI\_R: 5’ – CAG GAA ACA GCT ATG ACC CTA AGC AAA GTT TTC CTT TAA GA – 3’). The extension of the M13 tails (underlined) would also increase the likelihood of clean sequences at the very beginning and end of *Pgi*. To ensure the correct PCR products were sequenced (and not any of the shorter products seen on some of the gels), bands were cut out of gels and a MinElute Gel Extraction kit (Qiagen) used to purify the DNA fragments before sequencing.



**Figure 5.S1** – Map of the twenty seven sites from which an individual larvae was sampled and sequenced for the coding region of *phosphoglucose isomerase (Pgi)*.

Once the ORF of *Pgi* was sequenced from all twenty seven sites, they were aligned in Geneious Pro 5.3.6 (Drummond *et al.* 2010), revealing both synonymous and non-

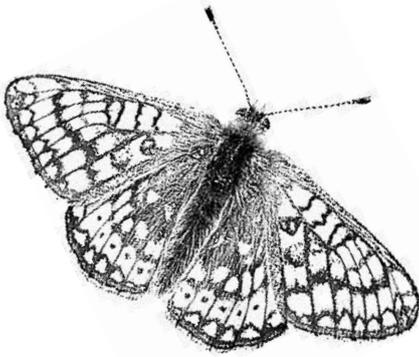
synonymous SNPs (**Figure 5.S2**). In total, 48 SNPs were found, although only 6 of these were non-synonymous.



**Figure 5.S2** – Synonymous and non-synonymous single nucleotide polymorphisms (SNPs) across the twelve exons of *Pgi*. Each row represents an individual from each of 27 sites across the UK (see **Figure 5.S1**) that was sequenced. Nucleotide positions are indicated at the top of each column, with exon boundaries marked by vertical lines. Non-synonymous sites are highlighted in bold and dots indicate identical nucleotides to the first entry in that column, or reference sequence here shown as the sequence from the Lizard site. Nucleotide ambiguity codes are as follows: M = A / C; R = A / G; W = A / T; S = C / G; Y = C / T; K = T / G and H = A / C / T. An asterisk (\*) indicates the SNP of possible interest (see Chapter 5 of this thesis).

## Chapter 6

### **Friend or foe? *Wolbachia* infection in the threatened British butterfly *Euphydryas aurinia***



*'What makes thing baffling is their degree of complexity, not their sheer size.....a star is simpler than an insect'*

*Martin Rees 1999  
Exploring Our Universe and Others*



## 6.1 ABSTRACT

The ‘unseen’ threat of pathogens and parasites to populations of endangered arthropods is often overlooked in management strategies and conservation research. *Wolbachia* are endosymbiotic bacteria capable of dramatically altering the reproductive system of their host, with potentially severe consequences for population dynamics and persistence. The invasion of *Wolbachia* infections into fragile populations of endangered species risks local extinction caused by reduction in effective population size. We identified 100% prevalence of an unknown single strain of *Wolbachia* across 19 UK populations of the marsh fritillary butterfly, *Euphydryas aurinia*, a vulnerable species of Nymphalidae, using a PCR-based diagnostic in conjunction with MLST (multi-locus sequence typing). Such a high infection frequency may have implications for the conservation of butterfly species existing in fragile metapopulations. We attempted to infer the phenotype (cytoplasmic incompatibility, sex ratio distorter, neutral or mutualistic) of this *Wolbachia* strain using sequence homology with known strains: closest matches were with sex ratio distorters in other host species. If the strain was originally a sex ratio distorter, then 100% prevalence implies the evolution of a phenotype suppressor in the host. If, however, the wider host population is not completely infected, then transfer of infected individuals among subpopulations should be avoided.

## 6.2 INTRODUCTION

Lepidoptera in general have experienced unparalleled declines in the last few decades. In the UK, 37% of regularly breeding butterflies are either extinct already or threatened (Fox *et al.* 2010), a conservation crisis that is worse than for the Aves (29% not including extinct species; Eaton *et al.* 2005), vascular plants (20%; Cheffings and Farrell 2005) and the Odonata (21%; Daguet *et al.* 2008). As sensitive ‘environmental indicators’ of habitat health, and with very short generations, the Lepidoptera are the first major taxon to visibly illustrate the effects of habitat deterioration or other unseen foes.

Selfish genetic elements (SGEs) inherited through the maternal line are common within eukaryotes (Hurst and Werren 2001) and can severely alter the phenotype of their host, with ramifications for population dynamics. The bacterial endosymbiont *Wolbachia*, one of the most intensively studied groups of SGEs, can cause a variety of alterations to their host’s reproductive system, from feminization (**F**) of genetic males (Kageyama *et al.* 2002) to inducing thelytoky (**T** – parthenogenesis in which only female offspring are produced; Russell and Stouthamer 2011), cytoplasmic incompatibility (**CI** - sperm from infected males cannot fertilise eggs from either uninfected females or females infected with a different strain; Narita *et al.* 2009) or even male killing (**MK** - death of male embryos; Hurst *et al.* 1999; Jiggins *et al.* 2001). With the prevalence of *Wolbachia* among arthropod species initially estimated at ~17% (Werren *et al.* 1995) and recently increased to around 66% (Hilgenboecker *et al.* 2008), with one study by Jeyaprakesh *et al.* (2000) even showing prevalence of 76%, it is unsurprising to encounter this SGE in endangered species. However, the presence of such potentially parasitic infections is extremely worrying in host species of conservation concern. The recent report of extensive *Wolbachia* infections in the endangered Karner Blue butterfly, *Lycaeides melissa samuelis*, in the Midwest of America (Nice *et al.* 2009) has raised the possibility that the presence of such bacterial endosymbionts may hinder the conservation of species whose distributions are fragmented due to changes in their habitat.

Due to vertical transmission, the route to proliferation for *Wolbachia* lies in the matriline of the host population. Consequently, there is intense selection on *Wolbachia* to produce efficient modes of spreading by manipulating host behaviour and biasing host populations towards female dominated sex ratios. F, T and MK strains directly influence

the sex ratio in a population by causing females to produce heavily female-biased broods. Whereas F and T strains turn potential males into *Wolbachia*-transmitting females, MK is the most extreme manipulator – killing male embryos early in development. These alterations result in a conflict of interest between autosomal genes – which are selected to invest equally in male and female production - and bacterial genes selected to favour production of females. The spread of the endosymbiont will therefore cause strong selection in the host for suppressors that either prevent transmission or reduce the action of the bacterium (Hornett *et al.* 2006), and which then spread rapidly through populations (Charlat *et al.* 2007). The final class of reproductive alteration known to be caused by *Wolbachia*, CI, does not affect the sex ratio of populations, but causes both uni- and bi-directional incompatibility between individuals (Engelstadter and Telschow 2009). The former occurs when sperm from infected males cannot fertilise the eggs of uninfected females, hence increasing the selective advantage of infected females. The latter occurs when differing strains are harboured by the male and female, rendering them incompatible. This final reproductive manipulation can reach high prevalence where host population structure allows, and can severely reduce gene flow between host populations differing in their infecting strain (Engelstadter and Telschow 2009).

Existing in fragile metapopulations, the marsh fritillary *Euphydryas aurinia* (Lepidoptera; Nymphalidae) has suffered serious declines across much of its Palaearctic distribution. The UK is currently considered a stronghold for the species, but even here there has been a distributional loss of 55% over a period of 30 years (Asher *et al.* 2001). *E. aurinia* has subsequently become a well-known example of a species facing local extinction due to fragmented or degraded landscape caused by anthropogenic activities. Consequently, large amounts of both effort and funding have been put into conservation projects across the UK, often using *E. aurinia* as an ‘umbrella’ target species for management that will benefit many other species. Despite extensive ecological and behavioural studies on habitat preferences (e.g. Konvicka *et al.* 2003; Liu *et al.* 2006), minimum viable metapopulation size (e.g. Bulman *et al.* 2007) and landscape-scale conservation (e.g. Anthes *et al.* 2003; Betzholtz *et al.* 2007), metapopulations of *E. aurinia* are shown to be still declining in south-west UK even after recommended management strategies have been implemented (Smee *et al.* 2011; **Chapter 2**). This necessitates prompt research beyond that of management requirements and butterfly habitat preferences.

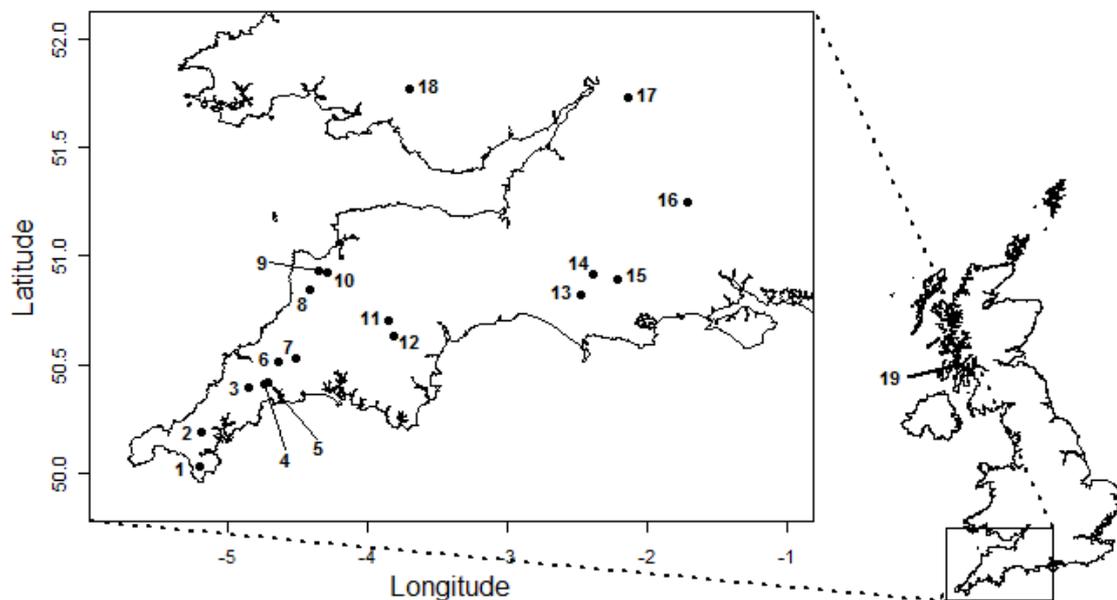
The implications of *Wolbachia* infection for declining metapopulation species of conservation concern are substantial – spread of such a strongly selected bacterium is perhaps unavoidable, but the costs to small, isolated populations could be catastrophic. Metapopulation dynamics rely on a balance between stochastic extinction of subpopulations and re-colonization of patches (Hanski 1999; Hanski and Thomas 1994), allowing the metapopulation to persist as a whole. This process depends largely on the ability of the species to disperse and colonise new areas, and to also keep immigration rates high enough to avoid reduced gene flow and potential inbreeding, which can heighten extinction risk (Saccheri *et al.* 1998). If *Wolbachia* causes extreme sex ratio distortion, reductions in population size due to male killing, or incompatibility between individuals within subpopulations of a threatened butterfly, then local extinction rates will increase via the combined influence of demographic stochasticity (small populations suffer a higher risk of complete collapse) and inbreeding (caused by reduced effective population sizes). Furthermore, if infected butterflies disperse from source populations to uninfected patches and spread infections, the benefits of connectivity among subpopulations may be lost entirely.

Here we show that a single strain of *Wolbachia* is 100% prevalent among 281 individuals sampled from 19 sites in the UK. The strain is novel, and therefore has unknown phenotype. Often considered a ‘foe’, particularly to threatened species (Nice *et al.* 2009), some studies have shown evidence of extreme mutualistic behaviour (Dedine *et al.* 2001; Hosokawa *et al.* 2010) whereby a host may be severely handicapped in the absence of the endosymbiont. Evidence of such extreme mutualism is lacking in Lepidoptera, but should not be ruled out. We therefore consider the implications of complete and incomplete prevalence of all possible *Wolbachia* phenotypes on *E. aurinia* metapopulations, and call for urgent action to determine its wider prevalence and the nature of the host-endosymbiont interaction.

## 6.3 MATERIALS AND METHODS

### 6.3.1 Sample collection and preparation

*E. aurinia* (Lepidoptera; Nymphalidae) is a univoltine species, with females usually laying an initial large batch of c. 300 eggs soon after emergence on the underside of leaves of the sole host plant used in this area, Devil's-bit scabious *Succisa pratensis* Moench (Dipsacaceae). Once emerged, the larvae live gregariously in silken webs from late July until winter diapause (Porter 1982). Under Natural England licence number 20091722, a single larva was collected from each communal web found in September (third or fourth instar), starved for 24 hours to remove plant material from the digestive tract and then dissected to remove the midgut and potentially any parasitoids, and then snap frozen. Genomic DNA was then extracted from all larvae following the methods of Martinez-Torres *et al.* (1998) with some modifications (see **Appendix 1** for protocol). A total of 281 larvae were collected from the 19 field sites shown in **Figure 6.1** and described in **Table 6.2**, covering the majority of the butterfly's range in mainland UK.



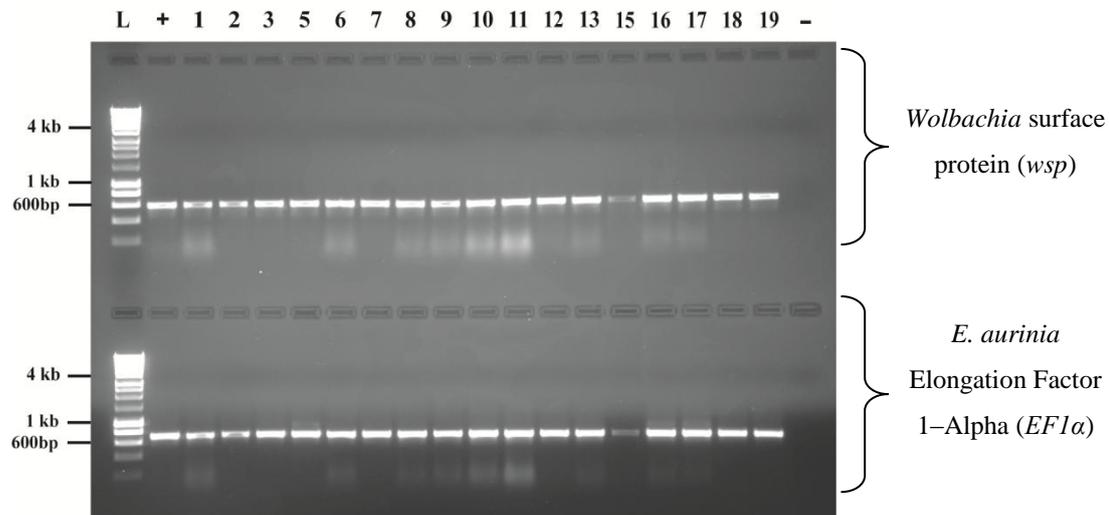
**Figure 6.1** – Sample sites in the south-west UK, Wales and Scotland. Population numbers correspond to **Table 6.2**. For the MLST, a single individual from each of the nineteen locations was sequenced for all five MLST genes and the surface protein (*wsp*). When screening with *wsp* primers, all samples available were tested – in Scotland and Wales this included samples from other populations within that country's distribution.

### 6.3.2 Detecting *Wolbachia*

An initial polymerase chain reaction (PCR) using primers specific to the *Wolbachia* surface protein gene (*wsp*) Wsp81F (5'-TGGTCCAATAAGTGATGAAGAAAC-3') and Wsp691R (5'-AAAAATTAACGCTACTCCA-3') (Braig *et al.* 1998; Zhou *et al.* 1998) was used to detect the presence of a *Wolbachia* infection. PCR products were visualised on a 1.2% agarose gel. The amplified *Wolbachia wsp* gene product was then cloned into the pCR4 TOPO/TA vector (Invitrogen) and ligations transformed into One Shot TOP10 *E. coli* competent cells. A PCR colony screen was carried out to confirm that the *wsp* sequence was indeed inserted into the vector, and positive colonies were then picked and incubated overnight at 37°C in LB medium containing 0.1 mg/ml Ampicillin. A QIAprep Spin Miniprep Kit (Qiagen) was used to purify the plasmid DNA. The *wsp* insert was subsequently sequenced using M13 forward (5'-TGTAACACGACGGCCAGT-3') and reverse (5'-CAGGAAACAGCTATGACC-3') primers and an ABI 3130 Genetic Analyzer (Applied Biosystems). Once the obtained (identical) sequences were confirmed as a *wsp* sequence by a BLAST (Basic Local Alignment Search Tool) against the NCBI databases, this sample was used as a positive control for further PCR screens.

To determine prevalence of *Wolbachia* across the British (ex. N. Ireland) range of *E. aurinia*, a diagnostic PCR-based test was carried out using the same *wsp* primers for the surface protein of *Wolbachia* (wsp81F and wsp691R) on samples in **Table 6.2**, using the methods of Zhou *et al.* (1998). In addition to the positive control gained from the previous sequencing efforts, both positive and negative *Wolbachia* controls for *Drosophila melanogaster* were included and gave the relevant results when tested. To avoid false negatives, samples were also put through the same PCR reaction with another set of primers specific to an *E. aurinia* housekeeping gene, Elongation Factor 1-Alpha (*EF1 $\alpha$*  – see **Table 6.1**). This set of reactions was used as DNA extraction verification – if the result was negative for EF1 $\alpha$  then that sample was discounted from the screened total. Problems were encountered with genomic DNA samples initially, likely due to an excess of template DNA concentration, or PCR inhibiting substances such as pigments. All samples were consequently diluted either 1 part genomic DNA in 100 parts MilliQ water (1:100 $\mu$ l), or 1:200 $\mu$ l, as dilution can lessen the effects of PCR inhibiting substances (Werren and Windsor 2000). This approach proved successful for the majority of samples. All PCR screening products were visualised on 1.2% agarose

gels and scored for presence/absence (See **Figure 6.2** for an example agarose gel). A subset of approximately 20 were re-run to confirm findings.



**Figure 6.2** – An example agarose gel for scoring presence/absence of *Wolbachia*. The first lane is HyperLadder I (Bioline) to size the fragments, followed by a positive control lane. Each numbered lane thereafter corresponds to the same population numbers in **Table 2**, with a negative control in the last column. The top row of bands indicate the successful amplification of the *Wolbachia* surface protein gene (*wsp*) with the lower row of bands confirming the quality of DNA extraction and ruling out false *wsp* positives by amplifying the housekeeping gene Elongation Factor 1-alpha (*EF1α*).

### 6.3.3 Multilocus sequence typing (MLST)

To identify the *Wolbachia* strain infecting UK populations of *E. aurinia*, multi-locus sequence typing (MLST) of one individual per population was conducted, as recommended by the *Wolbachia* MLST database (<http://pubmlst.org/wolbachia/>) (Baldo *et al.* 2006). This involved sequencing fragments of five *Wolbachia* genes: *gatB*, *hcpA*, *ftsZ*, *coxA* and *fbpA*, along with the surface protein *wsp* (containing four hyper-variable regions (HVR)) to give an allelic profile (ST). All degenerate primers (see **Table 6.1**) were designed with M13 tails at the 5' end (both forward and reverse) for ease when sequencing and PCR cycling conditions were followed as directed on the website indicated above. An ABI 3130 Genetic Analyzer (Applied Biosystems) was used to generate sequences which were then aligned using Geneious Pro 5.3.6 (Drummond *et al.* 2010) to templates of each gene. Consensus sequences were then submitted to both the *Wolbachia* MLST database for strain typing and GenBank under accession numbers JN116272 and JN116284 to JN116288.

**Table 6.1** – List of primers used during screening and MLST.

Gene	Primer sequences (5' – 3')	Annealing Temp. (°C)
<i>wsp</i>	<i>Wolbachia</i> surface protein wsp81F : TGG TCC AAT AAG TGA TGA AGA AAC wsp691R : AAA AAT TAA ACG CTA CTC CA	52
<i>gatB</i>	Aspartyl/glutamyl-tRNA(Gln) amidotransferase, subunit B gatB_F1: GAK TTA AAY CGY GCA GGB GTT gatB_R1: TGG YAA YTC RGG YAA AGA TGA	54
<i>coxA</i>	Cytochrome c oxidase, subunit I coxA_F1: TTG GRG CRA TYA ACT TTA TAG coxA_R1: CT AAA GAC TTT KAC RCC AGT	54
<i>hcpA</i>	Conserved hypothetical protein, NCBI COG0217 hcpA_F1: GAA ATA RCA GTT GCT GCA AA hcpA_R1: GAA AGT YRA GCA AGY TCT G	54
<i>ftsZ</i>	Cell division protein ftsZ_F1: ATY ATG GAR CAT ATA AAR GAT AG ftsZ_R1: TCR AGY AAT GGA TTR GAT AT	54
<i>fbpA</i>	Fructose-biphosphate aldolase fbpA_F1: GCT GCT CCR CTT GGY WTG AT fbpA_R1: CCR CCA GAR AAA AYY ACT ATT C	59
<i>EF1a</i>	Elongation Factor 1-Alpha EF1a_F : CCT GGC CAC AGA GAT TTC AT EF1a_R : CAC GAC GCA ATT CCT TAA CA	55

All primer sequences apart from *EF1a* and *wsp* were taken from the *Wolbachia* MLST database <http://pubmlst.org/wolbachia/> (Baldo *et al.* 2006) and ordered with M13 sequencing tags: Forward (5'-TGTAACGACGCGCCAGT-3') and Reverse (5'-CAGGAAACAGCTATGACC-3'). The *wsp* sequences are taken from Braig *et al.* (1998) and Zhou *et al.* (1998) and also ordered with M13 sequencing tags. *EF1a* primers were designed by the authors and ordered without M13 sequencing tags.

### 6.3.4 Phylogenetic analysis

To explore possible phenotypes of the *Wolbachia* strain isolated from *E. aurinia*, and to determine evolutionary relationships with strains isolated from other arthropods, a sequence alignment was constructed using MAFFT (Multiple Alignment using Fast Fourier Transform, [www.ebi.ac.uk/mafft/](http://www.ebi.ac.uk/mafft/)) for a total of 25 *wsp* gene sequences from strains where the phenotype consequently displayed by the host is known. The surface protein *wsp* gene was chosen for the analysis because it evolves at a faster rate than previously used genes such as *ftsZ* and consequently shows higher sequence variation (Van Meer *et al.* 1999; Zhou *et al.* 1998). A maximum likelihood (ML) phylogenetic tree was constructed using MEGA5 (Tamura *et al.* 2011) and the Tamura-Nei model (1993) of estimating evolutionary distance, with the bootstrap method of testing the support for each clade with 1000 replicates.

## 6.4 RESULTS

### 6.4.1 Screening for *Wolbachia*

Across the UK mainland distribution of *E. aurinia*, there was 100% prevalence of *Wolbachia* in a sample of 281 individuals (See **Table 6.2**) using the surface protein *wsp* as an identifier of *Wolbachia* presence.

**Table 6.2** – Population information from each site sampled (Sites correspond to those on the map in **Figure 6.1**) and numbers of individuals from each site screened with *wsp* primers.

Site number	Site Name	Population size*/ Trend	Number of individuals screened/Number infected
1	Lizard	38 / Increasing	21 / 21
2	Stithians	4 / Declining	8 / 8
3	Goss Moor	12 / Declining	12 / 12
4	Breney Common	35 / Stable	11 / 11
5	Redmoor	30 / Increasing	20 / 20
6	Colvannick	29 / Stable	22 / 22
7	Carkeet	100 / Stable	19 / 19
8	Dunsdon	186 / Stable	16 / 16
9	Volehouse	160 / Stable	22 / 22
10	Stowford	187 / Increasing	10 / 10
11	Venton	31 / Stable	2 / 2
12	Shapley	72 / Stable	1 / 1
13	Giants Hill	3272 / Stable	25 / 25
14	Lydlinch	9 / Declining	3 / 3
15	Hod Hill	191 / Stable	20 / 20
16	Salisbury	100 / Stable	18 / 18
17	Strawberry Bank	70 / Increasing	18 / 18
18	Wales	Varied	21 / 21
19	Scotland	Varied	12 / 12

\*Population size is the number of webs found during surveys at each site during 2009 (the year in which samples were collected). For screening purposes, samples were used from several populations across Wales and Scotland, but sites used for the MLST were: from Wales – Seven Sisters; from Scotland – Tayviallich, as shown in **Figure 6.1**.

There is no detectable link between prevalence of endosymbiotic infection and population dynamics across the sampling locations – there is a wide variety of population sizes and population trends, including large stable populations and severely declining populations on the brink of local extinction, but there is no variation in prevalence of infection - it is 100% everywhere. Although sex ratios were not measured in the field as part of this study, a separate study by the authors at a subset of the populations shows no evidence for female-biasing (See **Table 6.3**), contrary to what would be expected by the MK, F and T strains of the maternally inherited *Wolbachia*. There actually appears to be a larger percentage of males at some sites, although this may just be a consequence of the time of sampling within the flight period, as males emerge before females, or due to the ‘patrolling’ behaviour of males. In either instance, there is a clear presence of males.

**Table 6.3** – Putative sex ratios at a sub-sample of populations.

Site Name	Total N	% Females	% Males	Ratio F : M
Lizard	23	39	61	1 : 1.56
Breney Common	9	44	56	1 : 1.27
Redmoor	8	75	25	1 : 0.33
Lydlinch	8	37.5	62.5	1 : 1.67
Giants Hill	25	32	68	1 : 2.13
Southfield*	15	27	73	1 : 2.70
Hod Hill	15	53	47	1 : 0.89
Average		43.93	56.07	1 : 1.51

\*Southfield is located close to population 13 in **Figure 1**, but was not included in the present study into *Wolbachia* infection.

#### 6.4.2 Multilocus sequence typing (MLST)

For each of the five MLST genes and the surface protein *wsp* gene, the sequences from all 19 populations (one individual per population) were identical, indicating the presence of a single strain of *Wolbachia* across the UK mainland distribution of *E. aurinia*. Alleles for *gatB*, *hcpA* and *ftsZ* matched to alleles already present in the *Wolbachia* MLST database (<http://pubmlst.org/wolbachia/>), but *coxA*, *fbpA* and *wsp* sequences gave new alleles, and consequently a new strain as identified by the curator of the database as strain 235. The translated nucleotide sequence of *wsp* can further be split

into four Hyper-Variable Regions (HVR), of which three of the four matched to sequences already in the database, but HVR3 was deemed a new allele. In an allelic profile query to the database, the closest match to strain 235 (*E. aurinia*) was B-group strain 128, found in a singly infected chloropid fly (Diptera: Chloropidae) in USA, with only two of the five MLST genes identical to those of *E. aurinia*, and also with unknown phenotype.

### 6.4.3 Phylogenetic analysis

When comparing the partial *wsp* nucleotide sequence of *E. aurinia* to other hosts with known phenotype (**Figure 6.3A**), no obvious relationships emerged. It is evident that the *wsp* sequence alone cannot determine what phenotype a strain may exhibit. In a BLAST search against the NCBI database, the *wsp* sequence from *E. aurinia* revealed highest homology (100% coverage and 97% identity) to a B-group strain *Wolbachia pipientis* found in a Rice Crane fly *Tipula aino* (Diptera: Tipulidae) in Thailand (Kittayapong *et al.* 2003), and a strain found in a Bark beetle *Pityogenes chalcographus* (Coleoptera: Scolytidae) in Europe (Avtzis *et al.* 2008), neither of a known phenotype.



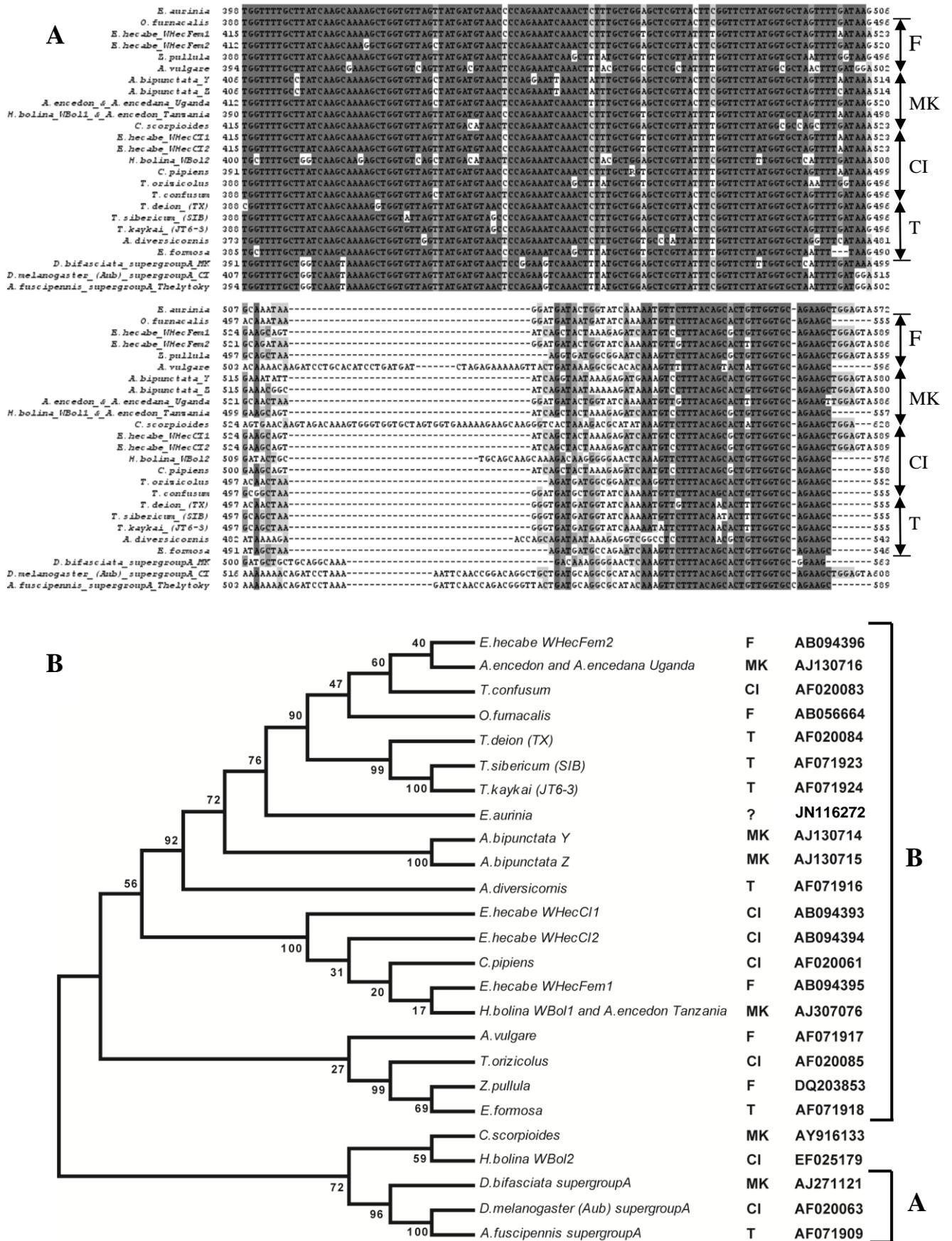


Figure 6.3 continued - F, Feminizing; MK, Male Killer; CI, Cytoplasmic Incompatibility; T, Thelytoky/parthenogenesis. Nucleotide sequences were aligned using MAFFT multiple

alignment programme. Conserved residues are boxed with light shading and identical residues are boxed with darker shading. The top sequence is the strain isolated from *E. aurinia*, and the bottom three sequences are strains from supergroup A as potential outgroups, showing three of the four phenotypes being compared here: MK, CI and T respectively. Sequences are grouped according to the phenotype produced – see groupings to the right of the alignment. **(B)** Phylogenetic analysis of the same *wsp* sequences, conducted in MEGA5 using Maximum Likelihood method. The results of bootstrap replicates are given next to the branches as a percentage of the number of replicate trees (1000) that produced the same cluster of taxa. The phenotype produced by each strain in that particular host is given to the right of the tree, along with GenBank accession numbers and the supergroup it belongs to, if known.

In a phylogenetic reconstruction with the same *wsp* gene sequences producing known phenotypes (**Figure 6.3B**), *E. aurinia* shows considerable divergence from all others, with nucleotide sequence distances ranging from 4.8% (*O. furnicalis*) to 21.5% (*A. vulgare*), with a mean of 13.8% overall. For ease, identical sequences (*A. encedon* and *A. encedana* Ugandan strains; *H. bolina* WBo11 and *A. encedon* Tanzanian strain) were treated as a single sequence. The consensus tree produced after bootstrapping shows very similar homology to the tree produced by Koop *et al.* (2009: See **Figure 6.2**) for the strains used in both diagrams. *E. aurinia* is firmly placed within supergroup B (See: Zhou *et al.* 1998), but seems to form a clade of its own after the male killing (MK) strains of *A. bipunctata* but before the thelytoky (T) strains of the parasitic wasps of the genus *Trichogramma* and the feminizer (F) strain of *O. furnacalis* – consequently providing no real clues towards a potential phenotype to be expected in infected *E. aurinia* butterflies.

## 6.5 DISCUSSION

Throughout the entirety of *E. aurinia*'s UK mainland distribution, a single strain of *Wolbachia* was found to be 100% prevalent. This is not a completely uncommon finding, with 100% of Karner Blue butterflies (*Lycaeides Melissa samuelis*) west of Lake Michigan also found to be infected (Nice *et al.* 2009); 95% of female *Acraea encedana* infected with a male-killer (Jiggins *et al.* 2000b); and over 99% of *Hypolimnas bolina* females infected with a male-killer in Independent Samoa (Dyson and Hurst 2004). However, in all of these species there does appear to exist either individuals or other populations in which *Wolbachia* does not reside, and therefore it would be of immediate importance to screen *E. aurinia* samples from other localities both inside and outside mainland UK to see if the phenomenon of complete prevalence continues across the remainder of its distribution from western Europe, Morocco and Algeria across Russia and temperate Asia to Korea (Tolman and Lewington 2009).

Unfortunately the results from this study do not enable us to uncover the mechanism of host manipulation, as in many other studies whereby the authors report the presence of *Wolbachia* but not the phenotype (e.g. Avtzis *et al.* 2008; Nice *et al.* 2009; Tagami and Miura 2004). Phenotype is not monophyletic, and instead seems dependent on interactions between the bacteria and the genomic make-up of the individual host, with some identical strains of *Wolbachia* causing different reproductive alterations in different hosts (Fujii *et al.* 2001) and some identical strains causing the same alteration in different hosts (Dyson *et al.* 2002). Sasaki *et al.* (2002) demonstrated this unreliability in predicting the phenotype produced by a single strain via micro-injection of *Wolbachia* directly from one species, in which it displayed CI, to another species in which it then displayed MK. Consequently, even when identical sequences are found in different species the phenotype cannot be inferred without experimental confirmation. The alignment and phylogenetic relationships between strains of known phenotype in **Figure 6.3** also strengthens the inability to infer the possible phenotype of the *E. aurinia* strain.

Often identical strains are found not in phylogenetically similar hosts, but in hosts with dependencies on each other, i.e. those engaged in host-parasite interactions, or those forming parts of the same community (Kittayapong *et al.* 2003) which gives strong evidence for occasional horizontal transmission of infection. An identical MK sequence in *H. bolina* to that of Tanzanian *A. encedon* also gives strong support for horizontal

transfer in Lepidoptera (Dyson *et al.* 2002). However, finding identical *wsp* sequences across taxa is uncommon, and even within the same host species; *Wolbachia* in *Adalia bipunctata* bear identical *ftsZ* sequences, but different *wsp* sequences (Hurst *et al.* 1999) and consequently form two different strains which are likely to have diverged recently as the differences are only indicated by the faster-evolving *wsp* gene (Schulenburg *et al.* 2000). Other hosts also seem capable of harbouring more than one strain of *Wolbachia* with the same (MK; Jiggins *et al.* 2001) or different (CI and F; Hiroki *et al.* 2004) phenotype.

Spread of *Wolbachia* in *E. aurinia* would appear to be a relatively recent phenomenon, as all sequences of the MLST genes and *wsp* gene are identical across the whole geographical distribution. The life history of *E. aurinia* makes it particularly susceptible to the effects of both vertical and horizontal transmission: a very short life span (c.a. 2 weeks) therefore encouraging maximum reproduction in the form of usually one large egg batch (perhaps a second smaller one further from the natal patch), followed by gregarious behaviour of the larvae, living and feeding on the same leaves of the single host plant in this area *S. pratensis*. The adult butterflies are generally sedentary, but adults have been known to disperse up to 15km (Warren 1994), which could have large implications for the spread of *Wolbachia*. General metapopulation dynamics would predict a rapid spread of infection independent of the phenotypic effects of the strain, as processes of extinction and re-colonisation are linked by dispersal. In a recent study using molecular markers to decipher population structure of *E. aurinia*, Smee *et al.* (in prep.: **Chapter 4**) found evidence of restricted gene flow between metapopulations and significant isolation-by-distance across the same geographical region, inferring low levels of local dispersal. Low levels of gene flow do not seem to explain the high prevalence of *Wolbachia* across these same metapopulations. An alternative explanation would involve a country-wide bottleneck in *E. aurinia* distribution, followed by a rapid expansion to its current distribution – due to the identical *wsp* sequences found across all sampled populations. This explanation is also unlikely: the low levels of contemporary gene flow reported by Smee *et al.* (in prep.: **Chapter 4**) do not complement a rapid expansion in distribution, and Fox *et al.* (2010) show that the long term trend (approx. 25 years) in area of occupancy (10 km grid) is - 46% with a population trend of - 73%.

All five of the MLST genes sequenced in *E. aurinia* do show, separately, high homology (95% < identity) to those found in MK strains of *A. encendon* or *H. bolina* –

both Nymphalidae butterflies, which potentially may indicate the phenotype of this strain. MK phenotypes enhance their transmission by reducing sibling competition, but this is extremely costly to the host, reducing densities of host males and therefore mating opportunities in the population. Consequently, selection should favour the evolution and spread of either resistance to infection, or suppressors, and thereby elicit a response from the symbiont once more. Hornett *et al.* (2006) describe the evolution of a MK suppressor in *H. bolina*, and show that if initial *Wolbachia* prevalence is in excess of 0.5 then the suppressor is likely to spread from 1% to 95% frequency in less than 100 generations. Subsequently, the suppressor enables uncovering of a second phenotype, CI, in surviving males (Hornett *et al.* 2008). This ‘hidden phenotype’ could help explain the complete prevalence of *Wolbachia* across *E. aurinia* populations: Hornett *et al.* (2010) show that when infection prevalence is high, CI can drive both the suppressor and bacterium to even higher frequency. Extant (or observable) MK strains may just be those that have not yet been suppressed. If suppression evolves regularly, and spreads rapidly, MK infection becomes quiescent. This is well known for other sex ratio distorters, with only hybrids or inter-population crosses revealing the sex ratio distortion. In addition, the majority of sex ratio-distorting elements are found already in the presence of host suppressors that inhibit their action (Jacobs and Wade 2003).

There are three possible scenarios for the infection in *E. aurinia*; the strain could be harmful, ‘parasitic’, and hence cause any of the aforementioned reproductive alterations (cytoplasmic incompatibility (CI); male killing (MK); induced parthenogenesis/thelytoky (T); or feminization (F)); the strain may be ‘neutral’ and therefore cause no selective pressures on either host or bacteria; or the strain may be ‘mutualistic’ and cause positive selection to allow it to spread throughout populations. We consider the ‘neutral’ option to be unlikely, as drift to fixation of a neutral infection would require a severe bottleneck in UK *E. aurinia* populations, followed by spread to its current distribution.

Influx of uninfected individuals (either from dispersal events or organised re-introductions from captive breeding programmes) would have differing implications in the ‘parasitic’ and ‘mutualistic’ scenarios. Populations with high prevalence of a mutualistic symbiont would be expected to perform well, perhaps growing large enough to form a source subpopulation. Emigrants from this source would populate uninfected subpopulations, causing spread of the infection. Immigration of uninfected individuals

into infected subpopulations would likely have little impact as uninfected offspring would quickly be selected against. In populations with a high prevalence of a harmful strain, immigrant uninfected females without a suppressor gene, mated to a resident infected male, will either produce no offspring (in the case of CI strains) or they will produce all infected daughters (in the case of F, T or MK strains): both outcomes enhance the transmission of the *Wolbachia* even further. Immigration of hosts infected with a harmful strain into fragile, uninfected subpopulations lacking suppressor genes, could have catastrophic consequences, reducing absolute or effective population size and exposing these small populations to exaggerated demographic stochasticity and inbreeding (Nice *et al.* 2009). Only complete linkage between parasite and suppressor (an unlikely scenario) could counteract this effect.

Current management strategies for rapidly declining species in fragmented landscapes (*E. aurinia* is a classic example) include improving habitat quality and increasing connectivity between patches either through the formation of ‘corridors’ or expanding the habitat that is already present. Captive propagation and re-introductions are also used in extreme cases of local extinctions (Porter 2007; Porter and Ellis 2011). Such intense management could aid the spread of *Wolbachia* - by opening up infection routes across the landscape, or if captive breeding stock is infected - and potentially do more harm than good. Consequently, it is important to gain more information on the phenotype of the strain through experimental approaches, and to search more widely for variation in the prevalence of this endosymbiont. Initial steps would be to gain reliable sex-ratios from the field, and determine if prevalence continues outside of the UK mainland. Investigations into the spread of mitochondrial genes may also give insights into the origin of the *Wolbachia* strain, as they are often co-inherited markers (Charlat *et al.* 2009). It would be interesting to note whether other arthropods existing in the same community displayed infection and similar prevalence levels of the same or closely-related strain, and what phenotypic effects are shown. *E. aurinia* is prey to a specialist parasitoid wasp, *Cotesia bignelli* (Bulman 2001; Porter 1981), which may be an extremely convenient horizontal transmission route for *Wolbachia* to spread between metapopulations. Other studies also give cause for concern to infected metapopulation species; Goodacre *et al.* (2009) show that the occurrence of some endosymbionts, such as *Rickettsia*, has a negative relationship with dispersal tendency in ‘ballooning’ spiders. If the same behaviour is noted in Lepidopteran species existing in metapopulations, whereby infected individuals are less likely to disperse, then infection frequency would

rise in isolated subpopulations and disrupt extinction-colonization dynamics. The impacts of this remain unexplored: long-term metapopulation survival requires a balance of extinction and re-colonisation via dispersal, but parasite-mediated lack of dispersal could prevent the spread of *Wolbachia* to uninfected subpopulations.

Complete prevalence of *Wolbachia* across the UK mainland distribution of *E. aurinia*, suggests several possibilities. The strain may be neutral, but the required severe bottleneck of the host (for fixation of infection) followed by rapid geographical spread (required to maintain the identity of the *wsp* sequence across the range) is unlikely as previously explained. The strain may be beneficial to the species, but to confirm this, infection phenotypes must be determined. Finally, the strain may be harmful in the absence of a suppressor – in the case of CI there could be complete prevalence of a single strain as males and females of the same strain are still able to mate successfully; but in all other cases of harmful infections, to achieve complete prevalence and no extreme sex-ratio biases in the field, a suppressor would have had to evolve and be strongly selected upon. It remains possible that the decline of *E. aurinia* is linked to *Wolbachia* infection, but 100% prevalence implies either the successful evolution of a suppressor gene in the host, or the attenuation of virulence (perhaps even change to a mutualistic phenotype) of the endosymbiont. However, complacency is unwise in conservation biology: urgent work is required to accurately characterise the host-endosymbiont relationship, and to survey for prevalence elsewhere in the natural range of *E. aurinia*.

# Chapter 7

## General Discussion



*'The caterpillar ate through one nice green leaf, and after that he felt much better.  
Now he wasn't hungry anymore - and he wasn't a little caterpillar anymore.*

*He was a big, fat caterpillar.'*

*Eric Carle 1969  
The Very Hungry Caterpillar*

## 7.1 GENERAL DISCUSSION

Each chapter has been written as a separate piece of research intended for publication, and hence are all explicitly separate in their aims and findings. However, all are also constituents of a common goal aimed towards the conservation of the marsh fritillary butterfly, *Euphydryas aurinia*. Here I aim to draw together all of these preceding chapters, and set the work in a broader conservation context whilst highlighting areas for future work.

### 7.1.1 Baseline research of species ecology is still needed

Over the past two decades the marsh fritillary butterfly, *Euphydryas aurinia*, has become an iconic species for conservation in the UK. Landscape scale projects such as the Two Moors Threatened Butterfly Project on Dartmoor in south-west England have proved extremely successful in encouraging this butterfly to breed and expand its range (pers. comm. Butterfly Conservation) and have reversed the short term population trend in the UK to a positive trend (Fox *et al.* 2010). Successes such as these emphasise the importance of knowledge on species' basic ecology and requirements for effective conservation, such as habitat preferences and the appropriate management tools to create favoured conditions.

Chapter two demonstrates this need for detailed knowledge on the basic ecology of populations in Cornwall, as steep declines in both adult and larval numbers are seen at all but one site, even after implementation of recommended management regimes (see **Figure 2.6**). This could be a spurious result: unfortunately, there are limitations to using only a five year dataset on a univoltine butterfly species - limitations which are increased further in species which demonstrate high levels of fluctuation in numbers even in 'good' periods (Ford and Ford 1930). Thomas *et al.* (2002) suggested that studies of four years or less cannot observe the higher levels of variation in population dynamics that are found over longer term studies, and hence may severely underestimate extinction threats. Furthermore, the severe overall decline across the mid-Cornwall moors may be another example of *E. aurinia* populations suffering from 'extinction debt' (Tilman *et al.* 1994). This continued decline of a population, regardless of current management practices, may be due to a time lag between habitat loss, fragmentation or degradation, and the ability of the species to reach a new equilibrium

with the new habitat distribution. Extinction debt has been documented for *E. aurinia* in other areas of the UK (Bulman *et al.* 2007).

Conversely, it is equally possible that management over these nine Cornish sites has been successful, but the butterfly has not had long enough to respond. Unmanaged habitat can be quickly restored to suitable conditions with the right management (Bulman 2001) but unfortunately, during the majority of the flight periods of this study, weather was not conducive to successful dispersal as temperatures were low and rainfall was high. Levels of successful matings were consequently likely to be low as the ability of patrolling males to find females was restricted. Such influential environmental effects may have outweighed the beneficial effects of habitat management, delaying potential positive response of the butterfly population. However, some beneficial results of management are already being observed: it was clear that in the autumn months, when larvae are gregarious in their first few instars (Porter 1981), vegetation height plays an important role in the densities of the butterfly at a site, with higher densities seen in localities of higher autumn sward height (**Section 2.4**). This, and the reliance on host plant abundance, is also supported by previous studies on *E. aurinia* populations (Bulman 2001; Porter 1981; Warren 1994) which show a preferential vegetation height of 5 – 20cm for *E. aurinia*. Over at least three of the sites in chapter two there is a clear increase in occurrence of the host plant *S. pratensis* (see **Figure 2.4**), which holds hope for an eventual positive response from the butterfly.

Baseline research on species ecology such as this is still of extreme importance for declining species and devising appropriate management plans. This importance has been well recognised for *E. aurinia*, as is evident from the wide range of ecological studies conducted across a wide geographical area, from Wales in the UK (Fowles and Smith 2006), to Spain (Munguira *et al.* 1997), Germany (Anthes *et al.* 2003), Czech Republic (Hula *et al.* 2004; Konvicka *et al.* 2003), Sweden (Betzholtz *et al.* 2007), Finland (Wahlberg *et al.* 2002) and even China (Chen *et al.* 2004; Liu *et al.* 2006).

### **7.1.2 Neutral molecular markers: useful, but questionable?**

The use of neutral molecular markers for studies of population genetics is not without problems. There are inherent advantages and disadvantages of all molecular markers, depending on the question being asked and the species being studied. Allozymes and mtDNA are often not polymorphic enough for more detailed population structure

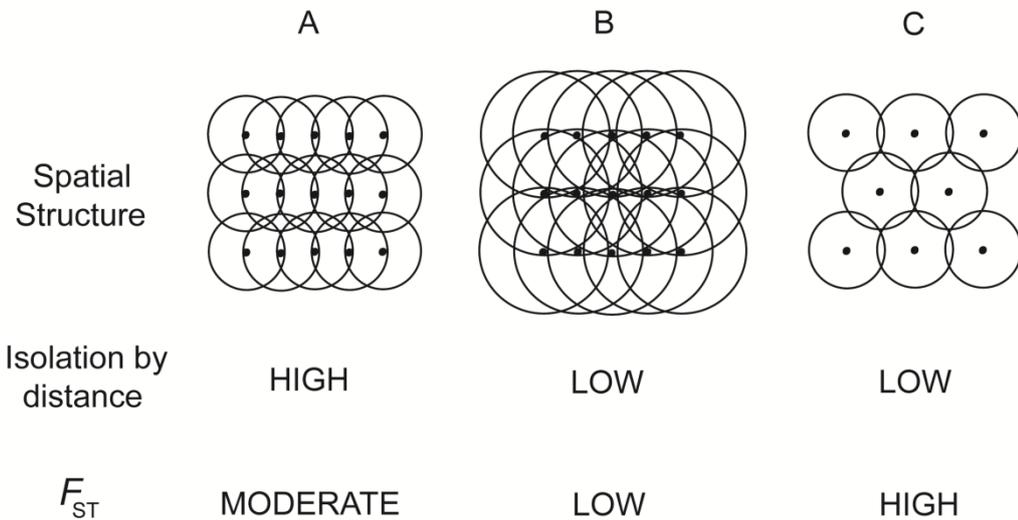
analysis and won't identify the entire variation present in the sample. Allozymes also need fresh samples for analysis and so exclude any data that might be gained from using historical museum specimens. There is an ever-increasing number of novel approaches to marker systems for use in entomological studies, however many have not gained widespread popularity (Behura 2006). Microsatellites remain the marker of choice for many studies as they show high levels of polymorphism, are abundant therefore allowing large numbers of markers for studies, and are easily genotyped (Goldstein and Schlötterer 1999).

Unfortunately, in lepidopteran species there are almost always problems in the isolation and characterisation of microsatellite loci (see **chapter 3** for more details). Furthermore, doubt has also been raised over the use of microsatellites as neutral markers, due to a suspected influence on gene spacing and secondary folding of DNA in the chromosomes as well as regulatory roles in gene expression (Behura 2006; Goldstein and Schlötterer 1999).

The use of EST-derived microsatellites (**chapter 3** and **chapter 4**) is also controversial, as loci may be under the influence of selection and hence may show deviations from Hardy-Weinberg equilibrium (HWE: **Figure 3.1**). However, this deviance from HWE should not be seen as a strong limiting factor on the use of these markers: first, in nature the assumptions of HWE are almost universally violated – no wild population exists in complete equilibrium from one generation to the next. Secondly, it has been demonstrated that these EST-SSRs (EST-derived simple sequence repeats, or microsatellites) show the same results and population differentiation as a set of AFLP markers on the same subset of samples (Mikheyev *et al.* in prep.). The deviations from HWE may be simply because there is actual structure amongst the populations.

Chapter four follows both the more traditional methods of population genetic analyses, allowing measures of population differentiation to inform on patterns and processes of genetic structure, such as IBD (Wright 1943) and  $F_{ST}$  measures (Wright 1931), as well as the more recent emerging field of maximum likelihood and Bayesian frameworks for inferring clusters, or genetic 'populations' as determined from allele frequencies across spatial populations (Gaggiotti 2004; Pritchard *et al.* 2000). Analyses of IBD can still be very informative, although it can be difficult to disentangle the actual message in the data. A high level of IBD is often shown by moderate dispersers who preferentially

mate with closer neighbours, but have the ability to disperse slightly further afield. Low levels of IBD are found in either highly mobile species, which are panmictic, or essentially isolated populations with extremely sedentary individuals, whereby each population becomes genetically distinct from all its neighbours as gene flow is significantly restricted, and so IBD is not found (**Figure 7.1**). Depending on geographical area, the same species may show different patterns of IBD: in the Southern Alps, populations of *Parnassius apollo* are large and there is occasional migration between them which leads to a lack of significant IBD, as more recent barriers have not yet caused differentiation. Yet, populations in the high Alps did show significant IBD, which is likely a result of stepping-stone post-glacial colonisation (Descimon *et al.* 2001). Similar reasoning can be applied to the pattern seen in *E. aurinia* across the two geographical areas studied here. In Catalonia, there is highly significant IBD possibly as a result of post-glacial stepping-stone movements of the butterfly (or linked to host-associated differentiation, which follows a North-South cline (Wee 2004; Mikheyev *et al.*, in prep.), whereas in southern England this is only visible once an ‘outlier’ population is removed from the analysis (discussed in **Section 7.1.5**). It is likely that populations in southern England were once linked by much higher levels of migration, until severe habitat fragmentation caused individual populations to become more differentiated, and dispersal across the matrix of unsuitable habitat more difficult and costly.



**Figure 7.1** – The configuration of populations across a landscape has a large influence on the spatial population genetics observed. Small, filled black circles indicate suitable habitat patches, with their potential genetic ‘neighbourhood’ indicated by the larger empty circles. At each generation in (A), each of the populations may exchange individuals with its direct neighbours, whereas due to perhaps higher dispersal rates or less barriers across the landscape, the larger genetic neighbourhoods in (B) facilitate exchange with direct and indirect neighbours and allow much greater gene flow across the whole landscape, and hence lower population differentiation as measured by Wright’s  $F_{ST}$  (Wright 1931). Even though in (C) the neighbourhoods are the same size as in (A), there will be considerably higher population differentiation as the number of habitat patches is almost halved, resulting in individual populations becoming increasingly isolated and subject to genetic drift, causing genetic isolation of each patch and no evidence of IBD as there are no successful dispersal events. Figure adapted from Nève (2009).

The implications for conservation approaches to each of these scenarios are diverse – see **Section 7.1.5** for management recommendations.

Population genetics programmes such as STRUCTURE (Pritchard *et al.* 2000) can be controversial, and yet many more similar programmes are being developed, all with slightly varying underlying assumptions (Excoffier and Heckel 2006). Kalinowski (2011) claims that STRUCTURE does not correctly predict the number of clusters in a dataset, whereas other authors claim that with added spatial location data, there is an increase in accuracy (François and Durand 2010; Hubisz *et al.* 2009). In the current study of *E. aurinia*, the combination of both approaches, IBD and STRUCTURE, seems highly beneficial. Analyses based on  $F$ -statistics show considerable IBD across both

geographical regions independently, as well as between the two (although the English Channel was shown to cause most of the difference between the UK and continental populations, as it likely presents a formidable barrier to dispersal), but this pattern can also be inferred from the STRUCTURE analyses, as the dominant population membership at each location in southern England can be seen to change from west to east (**Figure 4.7**) and the continental populations are highly differentiated from the UK populations.

### 7.1.3 Adaptive molecular markers in lepidopteran conservation

Although it would be of great benefit to be able to predict differential success and fitness at the organismal level by studying potential adaptive markers such as *Pgi* at the molecular level, what is true for one species is not always true for others, as is the case here. In **chapter 5**, similarities were discovered in *Pgi* between *E. aurinia* and *M. cinxia* in the coding region of the gene, whereby both species demonstrated high levels of SNP variation (in comparison to three other metabolic genes sequenced from *M. cinxia*). Charge-changing (non-synonymous) SNPs were found in very close proximity; amino acid 373 in *E. aurinia* and amino acids 372 and 375 in *M. cinxia*. Also, *C. eurytheme* is also known to have a non-synonymous SNP on the codon for amino acid 375, suggesting structural similarities among *Pgi* in the three species. Evidence of balancing selection via heterozygote advantage was however missing for *E. aurinia*, suggesting that PGI could function as a neutral marker rather than an adaptive marker in declining UK populations (Hedrick 2004).

It has recently been proposed that ‘cryptic’ neutral variation can slowly accumulate in an organism, which may be beneficial and cause the individual to be more predisposed to survive if then placed in an entirely new environment (Hayden *et al.* 2011). This may be an appropriate proposal in such a system as this, and hints that perhaps even though this genetic variation, albeit very similar to the variation in both *M. cinxia* and *C. eurytheme*, shows no evidence of selection or fitness benefits in southern UK populations, there may be a different result in continental populations who experience very different climatic situations. There may also be other SNPs of importance within the coding region of *Pgi* in *E. aurinia* which are only found in continental European populations and therefore have not been identified here (initial alignment was 27 individuals from sites in the UK only – see **Section 5.6** Supplementary Material).

For studies such as presented here, there are large implications of working on a declining protected species instead of a long-term thoroughly monitored species which can be manipulated, such as *M. cinxia* in the Åland Islands (Hanski 1999; Nieminen *et al.* 2004). Flight measures such as those implemented by Niitepold *et al.* (2009: harmonic radar tracking) may be far more thorough than the simple measures used in chapter five, but such techniques are not always appropriate for all species, and especially those under strict legislation.

#### 7.1.4 Out of sight, out of mind – unseen foes must not be forgotten

Although in **chapter 6** there was no apparent detrimental effect of the infection of the bacterial endosymbiont *Wolbachia* in *E. aurinia* populations, it cannot be ruled out entirely that such endosymbiotic bacteria have played some part in the decline of *E. aurinia* and may play a role in the decline of many arthropod species. There is no straight forward methodology to determine what phenotypic manipulation should arise from each specific strain, as the same strain can produce different phenotypes in different hosts (Fujii *et al.* 2001; Sasaki *et al.* 2002). It would seem possible that as the infection frequency is 100% of an identical strain but there is no visible phenotype, and no obvious sex-ratio differences in the field, that the strain could be a suppressed male-killing strain. Due to the great selective pressure on the organism to inhibit or resist the action of a sex ratio distorter such as male-killers, often a suppressor may evolve (Hornett *et al.* 2006) and spread through infected populations in rapid selective sweeps. In such circumstances it may be thought that the detrimental effects of such an infection are lost if the phenotype is suppressed, but it may actually cause further problems and complexity if the suppression of one phenotype leads to the expression of another, different phenotype (Hornett *et al.* 2008; Hornett *et al.* 2010) which in turn has even greater detrimental effects on the dynamics of populations. In any case, it is imperative that further studies resolve the phenotype of such infections so that future conservation initiatives may consider such infections alongside important variables such as habitat quality, size and isolation.

In **chapter 2**, as in Bulman (2001), no correlation was found between adult butterfly densities and patch quality, which suggests other limiting factors on population density. Selfish genetic elements (SGEs) such as *Wolbachia* infecting populations of Lepidoptera in fragmented landscapes can cause various reproductive alterations (Hurst and Werren 2001) and severe ramifications for population dynamics. However,

parasitoid wasps are another, equally catastrophic, potential cause of severe disturbance to population dynamics (Shaw *et al.* 2009) and have also been mostly neglected in conservation strategies. *Cotesia bignelli*, a specialist parasitoid of *E. aurinia*, appears to also have a metapopulation structure and Klapwijk (2008) emphasises the need for large and continuous habitat patches to support both butterfly and parasitoid. With enough host plant available, populations of larval webs should be large enough to support the dependent metapopulations of parasitoid, whilst also buffering the effect on population dynamics. The more isolated larval webs are, the higher the level of parasitism (Klapwijk 2008), which could have severe consequences for marginal populations of *E. aurinia* and eventually lead to local extinction. This may also cause local extinction of the parasitoid wasp due to its dependence on the presence of the host.

### 7.1.5 Management recommendations

The overarching aim in conservation research is always to inform management practices on the ground. For *E. aurinia* and other lepidopteran species surviving as metapopulations in fragmented landscapes, depending on priorities, there is much to consider before implementing plans. The findings of this study have strong implications for the conservation of *E. aurinia* and other such species.

Fragmentation and the reduction in migration, loss of genetic variation in small populations, and inbreeding depression are three of the seven major genetic issues in conservation biology, as realised by Frankham (1995). Conservation management has previously just considered the effects of fragmentation on populations and the consequent risk of inbreeding potentially leading to extinction (Saccheri *et al.* 1998). Management to this effect is aimed at increasing suitable habitat area, as well as increasing connectivity between sites, and in some cases extreme measures of reintroductions (Porter 2007; Porter and Ellis 2011; Thomas *et al.* 2009). Yet, if priorities are to conserve not only genetic diversity in populations of endangered lepidopteran species, but also genetic integrity of populations, then such management techniques could be counter-productive in certain circumstances. Management for genetic diversity might be favoured when local populations are suffering inbreeding depression. Management for genetic integrity might be favoured if local populations harbour rare alleles or show local adaptation to habitat, climate or host plant (e.g. Baxter *et al.* 2010; Mikheyev *et al.* in prep).

If, as is the case for *E. aurinia* in both southern England and Catalonia (see **chapter 4**), there is significant IBD across a landscape and populations are also significantly differentiated from one another, improving connectivity may actually reduce the genetic integrity of individual populations as gene pools become mixed. This may, or may not, result in the loss of potential adaptive variation in the form of ‘cryptic’ neutral variation accumulating over time (Hayden *et al.* 2011), causing all interlinked populations to be less able to adapt to changing environments, rather than only a few. Crucial conservation measures in these circumstances suggest only improvement of the habitat to ensure long-term suitability, if there is already high within-population diversity. If within-population diversity is low, it may be more appropriate to translocate individuals from distant populations to increase diversity and adaptive potential. If IBD is significant, but populations are not highly differentiated from one another, then improving connectivity is neither beneficial nor detrimental, as neighbouring populations are likely to be similar genetically. In this scenario, habitat improvement is again a key factor to encourage population growth and diversification, but if within-population diversity is low it may be necessary to reintroduce populations with higher diversity from elsewhere.

If there is no IBD detected across landscapes, the suggested management may be entirely different. In cases of highly dispersive species, this may suggest a panmictic situation, whereas for less mobile species a lack of IBD is more likely to illustrate a severe isolation of populations and their subsequent differentiation from one another (Nève 2009). In the latter case, in situations of high population differentiation and high within-population diversity, connectivity will again reduce the genetic integrity of populations and should be strongly avoided, using only habitat improvement to allow populations themselves to increase in size and natural occasional dispersal events to occur. On the other hand, with no IBD, low within-population diversity and low diversity among all populations in general although they may be significantly differentiated, a dire situation occurs. In this worst-case scenario, recommended options are to improve the habitat, and hence the population size (actual and effective), and reintroduce individuals from geographically distant populations, or captive sources that are genetically stronger (higher variation). In the case of *E. aurinia* in Cumbria, UK, the last few remaining webs of larvae (155 larvae in total) were taken into captivity in 2004 to try and artificially boost numbers to then reintroduce them back into a network of (improved) habitat (Porter 2007; Porter and Ellis 2011). Unfortunately, the pure

Cumbrian stock was potentially too genetically weak and stocks dwindled, resulting in the addition of Scottish lines (shown to be most genetically similar to the Cumbrian lines) to mate with the remaining few Cumbrian individuals, therefore increasing genetic diversity and population size. A re-introduction in 2007 to known historical network areas has so far been a success (Porter and Ellis 2011).

Frankham (1995) also proposed that genetic adaptation to captivity and the ensuing effects on reintroduction success as one of the seven major genetic issues in conservation biology. In the case of the Cumbrian re-introductions of *E. aurinia*, there was a specific goal in mind and different genetic lineages were purposefully mixed and not kept in captivity indefinitely. However, there is the case of one site in Cornwall that seemed to be significantly genetically differentiated from neighbouring populations in chapter four, suggesting an unnatural ‘colonisation’ event. This population, Colvannick (population 6, **Table 4.2, Figure 4.7**), is in relative close proximity to other large populations in the area such as Carkeet (population 7, **Table 4.2, Figure 4.7**), but may have gone through a severe bottleneck in recent years, or it may have been supplemented by keen butterfly conservationists hoping to lend a helping hand to a struggling population by adding some individuals from personal stock likely to be well adapted to captivity and lacking adaptive potential in the wild.

Such situations illustrate clearly the difficulties associated with charismatic endangered species, and the unfortunate lack of thorough knowledge that can be gained realistically. When studying dispersal in particular, which, aside from the spatial distribution of habitat is a key determinant of population structure, it can be extremely difficult to determine if ‘colonisations’ really are just that, or if populations have continued to remain in patches at very low, undetectable frequencies.

### 7.1.6 Conclusions

It is widely accepted that the conservation of declining lepidopteran species in fragmented landscapes requires research and knowledge not only on behaviour and life history traits, but the underlying adaptive resources of populations and species. The landscape scale approach to conservation has received much support, especially when dealing with metapopulations of declining Lepidoptera such as the marsh fritillary butterfly, *Euphydryas aurinia* (for example: Bulman 2001; Bulman *et al.* 2007; Early

2006; Early *et al.* 2008), and will continue to be a high conservation priority for all species in fragmented landscapes.

More recently, it has also become clear that the way forward for successful conservation may well lie with a partnering of ecological data with conservation genomics approaches (Allendorf *et al.* 2010; Kohn *et al.* 2006; Primmer 2009). The rapid increase in genomic resources for non-model organisms and the development of increasingly sophisticated molecular tools for a wide variety of purposes, from efficient searching for neutral markers such as microsatellites, to locating SNPs of adaptive significance, i.e. with functional purpose and fitness effects, is already transforming our ability to answer previously unanswerable questions (Stapley *et al.* 2010). With the ever-increasing realisation that dramatic interventions are needed if we are to conserve many of our declining charismatic lepidopteran species, it is clear that the future of conservation will combine the wealth of knowledge on species direct requirements, with the great potential to discover their underlying adaptability or vulnerability in the face of climate and environmental change.

# Appendices



# Genomic DNA extraction protocol

adapted from Martinez-Torres *et al.* (1998)

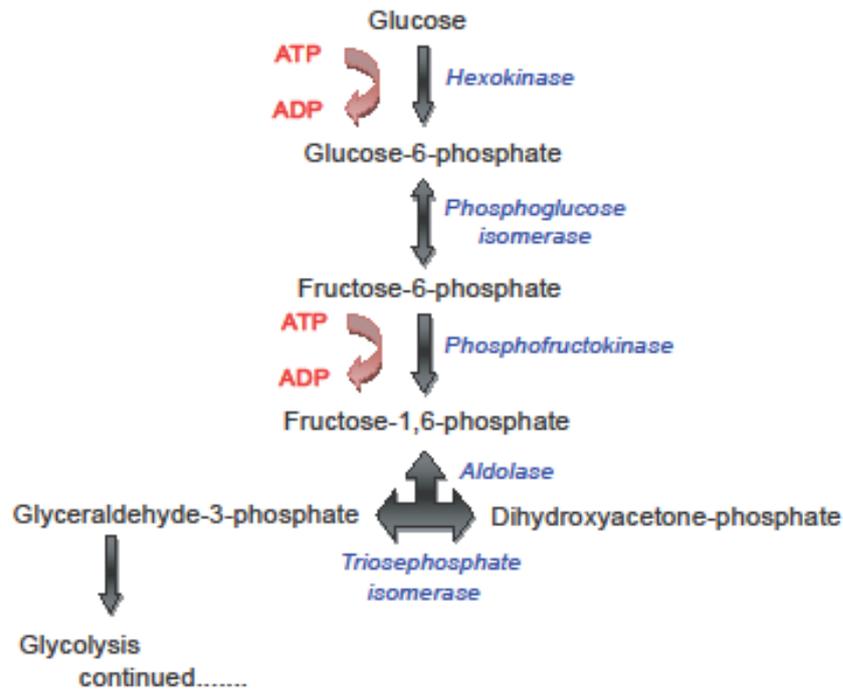
1. Source of gDNA – insect leg / half a larva / an abdomen etc – placed in 2ml eppendorf and placed in liquid nitrogen (N<sub>2</sub>).
2. Whilst kept frozen, homogenize the tissue with a sterile pestle until a fine powder is achieved.
3. Remove from liquid nitrogen and add **100µl of Solution 1** (10mM Tris-HCl, 60mM NaCl, 5% sucrose, 10mM EDTA (pH 8)) and continue to homogenize.
4. Add **125µl of Solution 2** (1.25% SDS, 300mM Tris-HCl, 5% sucrose, 10mM EDTA (pH 8)).
5. Mix by inverting and then incubate at **65°C** for 30 minutes.
6. Add **38µl of 3M potassium acetate** (pH 5.2).
7. Mix by briefly vortexing and incubate at **-80°C** for 15 minutes (can be extended).
8. Centrifuge at maximum speed for 10 minutes.
9. Transfer supernatant to a clean 1.5ml eppendorf (using 2 tips – set at 600µl).
10. Precipitate DNA by adding **300µl Isopropanol** and leaving at room temperature for 20 minutes.
11. Centrifuge for 10 minutes at **4°C**.
12. Remove supernatant and wash with **500µl -20°C 70% ethanol** (added along non-pellet side of eppendorf).
13. Centrifuge for 10 minutes at **4°C**.
14. Remove ethanol (using 2 tips – set at 600µl).
15. Leave to dry at room temperature for 25 minutes (or until ethanol has evaporated).
16. Re-suspend DNA pellet using **10µl RNase-free water** heated to **65°C**. Mix by pipetting, making sure to cover the whole pellet side of the eppendorf.
17. Store on ice until concentration is checked on a Nanovue, dilute appropriately, then store at -20°C.

## Sequencing the glycolytic enzyme phosphoglucose isomerase (*Pgi*) from *Euphydryas aurinia*.

### WHAT IS PGI?

The enzyme *phosphoglucose isomerase* (PGI) is just one of many enzymes involved in the metabolic pathway of glycolysis, which converts glucose to pyruvate. PGI catalyses the second step of glycolysis, which is the conversion of glucose-6-phosphate to fructose-6-phosphate (see **Figure A1**). Throughout the process, a variety of three- and six-carbon intermediate compounds are produced which can also be directly useful in other cellular functions, as well as the high-energy compounds adenosine triphosphate (ATP) and reduced nicotinamide adenine dinucleotide (NADH).

Glycolysis is consequently a very important metabolic pathway across a wide variety of taxa, including Lepidoptera. The main components of flower nectar, a main food source of adult butterflies, are glucose and fructose, which are converted to pyruvate through glycolysis - releasing the high-energy compounds ATP and NADH which are required for flight, as well as other essential cellular functions.



**Figure A1** – The first part of the metabolic pathway of glycolysis through which glucose is converted to pyruvate. Various enzymes catalyse steps in the pathway, including phosphoglucose isomerase (PGI) at the second step. The high-energy compounds adenosine triphosphate (ATP) and reduced nicotinamide adenine dinucleotide (NADH) are formed throughout the process.

### WHY IS *PGI* OF INTEREST?

A string of inaugural papers by Watt and co-authors through the 1970s and 80s on *Colias* butterflies demonstrated evidence for direct natural selection on polymorphisms found at the *pgi* locus. Electrophoretic variants, or different genotypes identified by allozymes, were shown to differ in metabolic capacity (Watt 1977) leading to heterosis whereby heterozygotes demonstrated the ability to fly under a broader range of environmental conditions (Watt 1983). Further studies revealed differences in flight ability, survivorship and mating success between *pgi* variants (Watt 1992; Watt *et al.* 1985; Watt *et al.* 1983). More recent studies by Wheat and co-authors have further proposed the utility of *pgi* as an adaptive molecular marker for arthropods (Wheat 2010), demonstrating that in the Glanville fritillary butterfly (*Melitaea cinxia*), haplotypes (determined by either allozymes or single nucleotide polymorphisms (SNPs)) gave predictable differences in a variety of fitness and life history traits, including female fecundity and flight metabolic rate (Haag *et al.* 2005; Orsini *et al.* 2009; Saastamoinen *et al.* 2009).

In current climes, understanding the genetic variation that underlies adaptation to changing environments is an important and fundamental topic, and critical for species of conservation concern. If there are pleiotropic effects of one locus causing individuals of certain genotypes to have increased flight ability at higher temperatures, but suffer lower fecundity, there may be severe consequences for populations in the event of climate change. It is of huge benefit to be able to predict such differential success and fitness at the organismal level by studying at the molecular level. In terms of conservation, this can help inform as to how populations will respond to environmental change, and even identify source populations for potential translocations or re-introductions.

### ATTEMPTS TO OBTAIN THE WHOLE GENOMIC LENGTH OF *PGI*

To be able to identify all SNPs across the entirety of the *pgi* gene, the genomic length first had to be obtained. In other Lepidopteran species, the entire length was found to be just shy of 11 Kb (*Colias eurytheme*, GenBank Accession number: DQ205092.1). Having the genomic sequence of *pgi* would also allow primers to be designed within introns for SNPs close to the beginning or end of exons. Table A1 gives details of all primer pairs tested to amplify regions of the genomic length, and their success or failure.

**Table A1** – All primer sequences tested whilst attempting to gain the genomic length of *pgi* from *E. aurinia*. It is noted as to which combinations successfully amplified the correct product.

Primer 1 Sequence 5' – 3'	Primer 2 Sequence 5' – 3'	Amplified region	(Correct) product?
MF_PGI_F5P TTTTTGCACACCGAGAACAC	MF_PGI_R3P ACATTTTGCTCTAAATTTCTAATATGG	Genomic length	✗
PGI-ORF-F ATGGAACCTAAGGTGAACTTG	PGI-ORF-R CTAAGCAAAGTTTTCTTTAAGA	Genomic length	✗
PGI-ORF-F ATGGAACCTAAGGTGAACTTG	PGI_gR1 ATCCCGGGCTTGTCTACTC	Exon 1 to Exon 2	✗
PGI-ORF-F ATGGAACCTAAGGTGAACTTG	PGI_gR2 CCTTCCAATTGCCACTGATAA	Exon 1 to Exon 3	✗
PGI_gF1 CATCCCCACACCAAATGAC	PGI_gR3 GGGCCTCAGTAACCATGAGA	Exon 2 to Exon 4	✓

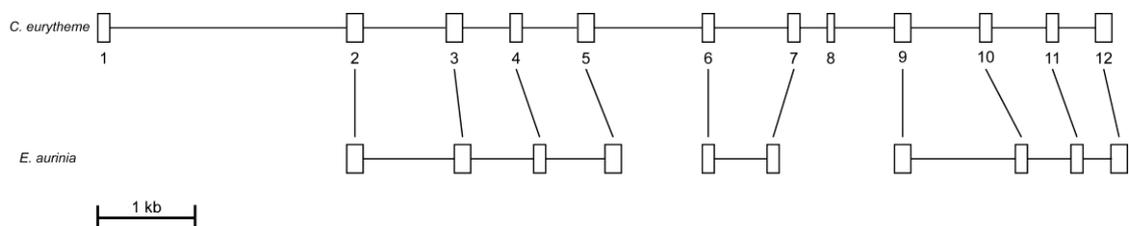
Appendix 2

PGI_gF2	PGI_gR4	Exon 3 to	✓
CGCACTGCGTAACAGAAAAA	GTGAAGGTTTTGGACGCAAT	Exon 5	✓
PGI_gF3	PGI_gR5	Exon 4 to	✗
CACAGCAAACCAATAACAGA	AACATGTTCTTGGGGTTCGAT	Exon 6	✗
PGI_gF4	PGI_gR6	Exon 5 to	✗*
AACCCACCTAGCCGAAGTCT	GAGTCTCGGCTCCGTAGAAG	Exon 8	✗*
PGI_gF5	PGI_gR7	Exon 7 to	✗*
CTGGTCTGCGATCGGTCTAT	TTGGCGAGCAGTATCTTGTG	Exon 10	✗*
PGI_gF6	PGI-ORF-R	Exon 9 to	✓
ATTCGCAGCGTATTTCCAAC	CTAAGCAAAGTTTTCTTTAAGA	Exon 12	✓
PGI-ORF-F	PGI_gR8	Exon 1 to	✗*
ATGGAACCTAAGGTGAACTTG	GACTCTGGCCAGCTCAAAC	Exon 2	✗*
PGI_gF7	PGI_gR9	Exon 5 to	✗
AACCCACCTAGCCGAAGTCT	TAGTACCTCGTCCCCTGGTG	Exon 9	✗
PGI_gF8	PGI_gR10	Exon 5 to	✗
ATTGCGTCCAAAACCTTCAC	CGTGTAGGGTCTCGAGGGTA	Exon 9	✗
PGI-ORF-F	PGI_gR11	Exon 1 to	✗*
ATGGAACCTAAGGTGAACTTG	GCCAGCAGTGGCTACTAGA	Exon 2	✗*
PGI_gF8	PGI_gR12	Exon 5 to	✗
ATTGCGTCCAAAACCTTCAC	ACCCAGTCCCAAAAACCAAACA	Exon 6	✗
PGI_gF9	PGI_gR13	Exon 6 to	✓
CGTCTGCTGTATCAAAACATTTTCGT	ACGGCGCTGTTGTAAAGTGGTT	Exon 7	✓
PGI_gF10	PGI_gR14	Exon 7 to	✗
GCTCTGGTCTGCGATCGGTCTA	CGATGCAAGTACTGGTCGTAGGG	Exon 8	✗
PGI_gF11	PGI_gR10	Exon 8 to	✗
TATCTTAGCGCTGCTGGGTGTG	CGTGTAGGGTCTCGAGGGTA	Exon 9	✗
PGInlcFor	PGInlcRev	Exon 8 to	✗
GTGTGGTACGGAAACTTCT	CCCTGGTGTATCAGCTGGTAGAA	Exon 9	✗
PGI_gF12	PGI_gR15	Exon 5 to	✗*
AACATCGACGGAACCCACCT	TTGGGGTTCGATACCGAATGC	Exon 6	✗*
PGI_EX1_F1	PGI_EX2_R1	Exon 1 to	✗*
ATGGAACCTAAGGTGAACTTGAAGC	TCCCGGGCTTGTCTACTCCACGA	Exon 2	✗*
PGI_EX5_F1	PGI_EX6_R1	Exon 5 to	✗
CGACGGAACCCACCTAGCCGAAG	TTCTCCCCGTTTCGTCGACAGTGC	Exon 6	✗

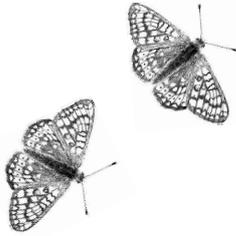
\*Primer pairs for which the desired product did amplify originally, but smaller bands were present or problems were encountered either in the cloning process or sequencing, resulting in no sequence data.

Genomic DNA was extracted using the protocol in Appendix 1; an adaptation of the protocol used by Martinez-Torres *et al.* (1998). Extension times of PCRs were adapted according to the primer pair used and hence the expected length of the region being amplified. The annealing temperature was also adjusted accordingly. As a control to validate genomic DNA extractions, and prevent false negatives when attempting to amplify regions of *pgi*, primers for a housekeeping gene *Elongation Factor 1-Alpha* (*EF1 $\alpha$* ) were also tested on samples in the same PCR reactions (Eauri-EF1a-F: 5' – CCT GGC CAC AGA GAT TTC AT – 3' and Eauri-EF1a-R: 5' – CAC GAC GCA ATT CCT TAA CA – 3'). These primers were also designed originally from the 454 transcriptome of *E. aurinia* using the *EF1 $\alpha$*  sequence from *Bombyx* as a BLAST query.

As the initial attempts to amplify the entire genomic length in one PCR failed (first two primer pairs in Table A1), primers were designed from the *Colias eurytheme* genomic sequence (GenBank Accession number: DQ205092.1) to amplify shorter regions of the *pgi* gene that overlapped, and hence would additively result in the entire genomic length of *pgi* for *E. aurinia*. This approach proved successful for some regions of the gene, but unfortunately not for others. Occasionally the right size fragment would amplify, but other smaller bands made cloning difficult. As the successful amplifications were sequenced, new primers were designed specific to *E. aurinia*. Figure A2 illustrates the progress made to gain the genomic length of *pgi* in *E. aurinia* before deciding to go ahead with just cDNA to assess SNPs. Another factor in this decision was the discovery of the sequence for a transposable element between exons 1 and 2 when sequencing the product of primers PGI\_gF12 and PGI\_gR15.



**Figure A2** – Progress made to gain the genomic length of *pgi* in *E. aurinia*. Hollow boxes represent exons, and connecting lines represent non-coding introns. The top half of the diagram illustrates the intron/exon structure of *C. eurytheme*, with exons numbered. The lower half of the diagram illustrates the progress made in the current study to obtain the genomic length in *E. aurinia*.



## References

- Abbitt RJ, Scott JM, Wilcove DS (2000) The geography of vulnerability: incorporating species geography and human development patterns into conservation planning. *Biol Conserv* 96:169-175.
- Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nat Rev Genet* 11:697-709. DOI: 10.1038/nrg2844.
- Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* 9:323.
- Anthes N, Fartmann T, Hermann G, Kaule G (2003) Combining larval habitat quality and metapopulation structure: The key for successful management of pre-alpine *Euphydryas aurinia* colonies. *J Insect Conserv* 7:175-185.
- Asher J, Warren M, Fox R, Harding P, Jeffcoate G, Jeffcoate S (2001) The millenium atlas of butterflies in Britain and Ireland. Oxford University Press, Oxford.
- Avtzis DN, Arthofer W, Stauffer C (2008) Sympatric occurrence of diverged mtDNA lineages of *Pityogenes chalcographus* (Coleoptera, Scolytinae) in Europe. *Biol J Linn Soc* 94:331-340. DOI: 10.1111/j.1095-8312.2008.01004.x.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* 3. DOI: 10.1371/journal.pone.0003376.
- Baldo L, Hotopp JCD, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MCJ, Tettelin H, Werren JH (2006) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol* 72:7098-7110. DOI: 10.1128/aem.00731-06.
- Barnett LK, Warren MS (1995) Species Action Plan: Marsh fritillary, *Eurodryas aurinia*. Butterfly Conservation.
- Bauerfeind SS, Theisen A, Fischer K (2009) Patch occupancy in the endangered butterfly *Lycaena helle* in a fragmented landscape: effects of habitat quality, patch size and isolation. *J Insect Conserv* 13:271-277. DOI: 10.1007/s10841-008-9166-1.
- Baxter SW, Nadeau NJ, Maroja LS, Wilkinson P, Counterman BA, Dawson A, Beltran M, Perez-Espona S, Chamberlain N, Ferguson L, Clark R, Davidson C, Glithero

- R, Mallet J, McMillan WO, Kronfurst M, Joron M, French-Constant R, Jiggins CD (2010) Genomic Hotspots for Adaptation: The Population Genetics of Müllerian Mimicry in the *Heliconius melpomene* Clade. *PLoS Genet* 6: e1000794.
- Beaumont M, Nichols R (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B* 263:1619 - 1626.
- Begun DJ, Aquadro CF (1994) Evolutionary Inferences From DNA Variation at the 6-Phosphogluconate Dehydrogenase Locus in Natural Populations of *Drosophila*: Selection and Geographic Differentiation. *Genetics* 136:155-171.
- Behura S (2006) Molecular marker systems in insects: current trends and future avenues. *Mol Ecol* 15:3087-3113.
- Betzholtz PE, Ehrig A, Lindeborg M, Dinnetz P (2007) Food plant density, patch isolation and vegetation height determine occurrence in a Swedish metapopulation of the marsh fritillary *Euphydryas aurinia* (Rottensburg, 1775) (Lepidoptera, Nymphalidae). *J Insect Conserv* 11:343-350. DOI: 10.1007/s10841-006-9048-3.
- Botham MS, Brereton TM, Middlebrook I, Cruickshanks KL, Harrower C, Beckmann B, & Roy DB (2008) United Kingdom Butterfly Monitoring Scheme report for 2008. CEH Wallingford.
- Braig HR, Zhou WG, Dobson SL, O'Neill SL (1998) Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. *J Bacteriol* 180:2373-2378.
- Brereton TM, Brook S, Hobson R (2005) Habitat condition monitoring for butterflies: 2004 pilot study., *Butterfly Conservation Report No. S05-06*, Wareham.
- Brereton TM, Warren MS, Roy DB, Stewart K (2008) The changing status of the Chalkhill Blue butterfly *Polyommatus coridon* in the UK: the impacts of conservation policies and environmental factors. *J Insect Conserv* 12:629-638. DOI: 10.1007/s10841-007-9099-0.
- Broquet T, Ray N, Petit E, Fryxell JM, Burel F (2006) Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landsc Ecol* 21:877-889. DOI: 10.1007/s10980-005-5956-y.
- Buhler C, Schmid B (2001) The influence of management regime and altitude on the population structure of *Succisa pratensis*: implications for vegetation monitoring. *J Appl Ecol* 38:689-698.

- Bulman CR (2001) Conservation biology of the marsh fritillary butterfly *Euphydryas aurinia*. PhD Thesis, University of Leeds, Leeds.
- Bulman CR, Wilson RJ, Holt AR, Bravo LG, Early RI, Warren MS, Thomas CD (2007) Minimum viable metapopulation size, extinction debt, and the conservation of a declining species. *Ecol Appl* 17:1460-1473.
- Carroll MJ, Anderson BJ, Brereton TM, Knight SJ, Kudrna O, Thomas CD (2009) Climate change and translocations: The potential to re-establish two regionally-extinct butterfly species in Britain. *Biol Conserv* 142:2114-2121. DOI: 10.1016/j.biocon.2009.04.010.
- Charlat S, Hurst GDD, Mercot H (2003) Evolutionary consequences of *Wolbachia* infections. *Trends Genet* 19:217-223. DOI: 10.1016/s0168-9525(03)00024-6.
- Charlat S, Hornett EA, Fullard JH, Davies N, Roderick GK, Wedell N, Hurst GDD (2007) Extraordinary flux in sex ratio. *Science* 317:214-214. DOI: 10.1126/science.1143369.
- Charlat S, Duploux A, Hornett EA, Dyson EA, Davies N, Roderick GK, Wedell N, Hurst GDD (2009) The joint evolutionary histories of *Wolbachia* and mitochondria in *Hypolimnas bolina*. *BMC Evol Biol* 9. DOI: 10.1186/1471-2148-9-64.
- Cheffings CM, Farrell L (2005) The Vascular Plant Red Data List for Great Britain. Species Status 7, Joint Nature Conservation Committee, Peterborough, UK.
- Chen J-J, Wang Y-F, Lei G-c, Wang R-j, Xu R-M (2004) Impact of habitat quality on metapopulation structure and distribution of two melitaeine butterfly species. *Acta Entomol Sin* 47:59-66.
- Clobert J, Ims RA, Rousset F (2004) Causes, Mechanisms and Consequences of Dispersal, in: I Hanski and OE Gaggiotti (Eds.), *Ecology, Genetics and Evolution of Metapopulations*, Elsevier Academic Press, London. pp. 307 - 336.
- Conradt L, Roper TJ, Thomas CD (2001) Dispersal Behaviour of Individuals in Metapopulations of Two British Butterflies. *Oikos* 95:416-424.
- Conradt L, Bodsworth EJ, Roper TJ, Thomas CD (2000) Non-random dispersal in the butterfly *Maniola jurtina*: implications for metapopulation models. *Proc R Soc Lond Ser B Biol Sci* 267:1505-1510.
- Crawley MJ (2007) *The R Book* Wiley, Chichester.
- Daguet CA, French GC, Taylor P (2008) The Odonata Red Data List for Great Britain. Species Status Assessment 11, Joint Nature Conservation Committee, Peterborough, UK.

- Dahlhoff EP, Rank NE (2000) Functional and physiological consequences of genetic variation at phosphoglucose isomerase: Heat shock protein expression is related to enzyme genotype in a montane beetle. *Proceedings of the National Academy of Sciences* 97:10056-10061. DOI: 10.1073/pnas.160277697.
- Davies ZG, Wilson RJ, Brereton TM, Thomas CD (2005) The re-expansion and improving status of the silver-spotted skipper butterfly (*Hesperia comma*) in Britain: a metapopulation success story. *Biol Conserv* 124:189-198. DOI: 10.1016/j.biocon.2005.01.029.
- Dedeine F, Vavre F, Fleury F, Loppin B, Hochberg ME, Bouletreau M (2001) Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *Proc Natl Acad Sci U S A* 98:6247-52.
- Department of the Environment (1995) *Biodiversity: The UK Steering Group Report*. HMSO, London.
- Descimon H, Zimmermann M, Cosson E, Barascud B, Neve G (2001) Genetic variation, geographic variation and gene flow in some French butterfly species. *Genetics Selection Evolution* 33:S223-S249.
- Dhuyvetter H, Gaublomme E, Desender K (2004) Genetic differentiation and local adaptation in the salt-marsh beetle *Pogonus chalceus* : a comparison between allozyme and microsatellite loci. *Mol Ecol* 13:1065-1074.
- Dolek M, Geyer A (1997) Influence of management on butterflies of rare grassland ecosystems in Germany. *J Insect Conserv* 1:125-130.
- Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) Geneious v5.3, Available from <http://www.geneious.com>.
- Dubut V, Grenier R, Megléczy E, Chappaz R, Costedoat C, Danancher D, Descloux S, Malausa T, Martin J-F, Pech N, Gilles A (2010) Development of 55 novel polymorphic microsatellite loci for the critically endangered *Zingel asper* L. (Actinopterygii: Perciformes: Percidae) and cross-species amplification in five other percids. *Eur J Wildl Res* 56:931-938.
- Dyson EA, Hurst GDD (2004) Persistence of an extreme sex-ratio bias in a natural population. *Proc Natl Acad Sci U S A* 101:6520-6523. DOI: 10.1073/pnas.0304068101.
- Dyson EA, Kamath MK, Hurst GDD (2002) *Wolbachia* infection associated with all-female broods in *Hypolimnas bolina* (Lepidoptera : Nymphalidae): evidence for

- horizontal transmission of a butterfly male killer. *Heredity* 88:166-171. DOI: 10.1038/sj/hdy/6800021.
- Earl DA (2011) Structure Harvester v.0.6.5 Available at [http://taylor0.biology.ucla.edu/struct\\_harvest/](http://taylor0.biology.ucla.edu/struct_harvest/) (accessed 12th May 2011).
- Early R (2006) Conservation at the Landscape Scale; the Marsh Fritillary and Other British Butterflies. PhD Thesis, University of York.
- Early R, Anderson B, Thomas CD (2008) Using habitat distribution models to evaluate large-scale landscape priorities for spatially dynamic species. *J Appl Ecol* 45:228-238. DOI: 10.1111/j.1365-2664.2007.01424.x.
- Eaton MA, Gregory RD, Noble DG, Robinson JA, Hughes J, Procter D, Brown AF, Gibbons DW (2005) Regional IUCN Red Listing: the Process as Applied to Birds in the United Kingdom. *Conserv Biol* 19:1557-1570.
- Egea R, Casillas S, Barbadilla A (2008) Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res* 36:W157-W162. DOI: 10.1093/nar/gkn337.
- Ehrlich PR (1961) Intrinsic Barriers to Dispersal in Checkerspot Butterfly. *Science* 134:108-109.
- Eichel S, Fartmann T (2008) Management of calcareous grasslands for Nickerl's fritillary (*Melitaea aurelia*) has to consider habitat requirements of the immature stages, isolation, and patch area. *J Insect Conserv* 12:677-688. DOI: 10.1007/s10841-007-9110-9.
- Endersby NM, Hoffmann AA, McKechnie SW, Weeks AR (2007) Is there genetic structure in populations of *Helicoverpa armigera* from Australia? *Entomol Exp Appl* 122:253-263.
- Engelstadter J (2010) The Effective Size of Populations Infected With Cytoplasmic Sex-Ratio Distorters. *Genetics* 186:309-320. DOI: 10.1534/genetics.110.120014.
- Engelstadter J, Telschow A (2009) Cytoplasmic incompatibility and host population structure. *Heredity* 103:196-207. DOI: 10.1038/hdy.2009.53.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611-2620.
- Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nat Rev Genet* 7:745 - 758.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10:564-567.

- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* 131:479-491.
- Fahrig L (2003) Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 34:487-515.
- Falush D, Stephens M, Pritchard JK (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* 164:1567-1587.
- Fauvelot C, Cleary DFR, Menken SBJ (2006) Short-term impact of disturbance on genetic diversity and structure of Indonesian populations of the butterfly *Drupadia theda* in East Kalimantan. *Mol Ecol* 15:2069-2081.
- Fitzpatrick MJ, Ben-Shahar Y, Smid HM, Vet LEM, Robinson GE, Sokolowski MB (2005) Candidate genes for behavioural ecology. *Trends Ecol Evol* 20:96-104.
- Ford HD, Ford EB (1930) Fluctuation in numbers, and its influence on variation, in *Melitaea aurinia*, Rott. (Lepidoptera). *Trans R Entomol Soc Lond* 78:345-352.
- Fowles AP, Smith RG (2006) Mapping the habitat quality of patch networks for the marsh fritillary *Euphydryas aurinia* (Rottemburg, 1775) (Lepidoptera, Nymphalidae) in Wales. *J Insect Conserv* 10:161-177. DOI: 10.1007/s10841-006-6291-6.
- Fox R, Warren MS, Brereton TM, Roy DB, Robinson A (2010) A new Red List of British butterflies. *Insect Conservation and Diversity*. DOI: 10.1111/j.1752-4598.2010.00117.x.
- François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources* 10:773-784.
- Frankham R (1995) Conservation Genetics. *Annu Rev Genet* 29:305-327. DOI: doi:10.1146/annurev.ge.29.120195.001513.
- Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge.
- Fujii Y, Kageyama D, Hoshizaki S, Ishikawa H, Sasaki T (2001) Transfection of *Wolbachia* in Lepidoptera: the feminizer of the adzuki bean borer *Ostrinia scapulalis* causes male killing in the Mediterranean flour moth *Ephestia kuehniella*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268:855-859. DOI: 10.1098/rspb.2001.1593.

- Gaggiotti OE (2004) Multilocus genotype methods for the study of metapopulation processes, in: I Hanski and OE Gaggiotti (Eds.), *Ecology, Genetics and Evolution of Metapopulations*, Elsevier Academic Press, London. pp. 367 - 386.
- Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. *Methods in Enzymology* 395:202-222.
- Goldstein DB, Schlötterer C (1999) *Microsatellites: Evolution and Applications* Oxford University Press, New York.
- Goodacre SL, Martin OY, Bonte D, Hutchings L, Woolley C, Ibrahim K, Thomas CFG, Hewitt GM (2009) Microbial modification of host long-distance dispersal capacity. *BMC Biol* 7. DOI: 10.1186/1741-7007-7-32.
- Goulson D (1993) Allozyme variation in the butterfly, *Maniola jurtina* (Lepidoptera: Satyridae) (L.): evidence for selection. *Heredity* 71:386-393.
- Guo SW, Thompson EA (1992) Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* 48:361-372.
- Haag CR, Saastamoinen M, Marden JH, Hanski I (2005) A candidate locus for variation in dispersal rate in a butterfly metapopulation. *Proc R Soc Lond B Biol Sci* 272:2449-2456. DOI: 10.1098/rspb.2005.3235.
- Habel JC, Schmitt T (2009) The genetic consequences of different dispersal behaviours in Lycaenid butterfly species. *Bull Entomol Res* 99:513-523. DOI: 10.1017/s0007485308006640.
- Hanski I (1994) A Practical Model of Metapopulation Dynamics. *J Anim Ecol* 63:151 - 162.
- Hanski I (1998) Metapopulation dynamics. *Nature* 396:41-49.
- Hanski I (1999) *Metapopulation Ecology* Oxford University Press, Oxford.
- Hanski I, Thomas CD (1994) Metapopulation dynamics and conservation - a spatially explicit model applied to butterflies. *Biol Conserv* 68:167-180.
- Hanski I, Gaggiotti OE (2004) *Ecology, genetics, and evolution of metapopulations*. Academic Press, San Diego.
- Hanski I, Saccheri I (2006) Molecular-level variation affects population growth in a butterfly metapopulation. *PLoS Biol* 4:719-726. DOI: 10.1371/journal.pbio.0040129.
- Hanski I, Kuussaari M, Nieminen M (1994) Metapopulation structure and migration in the butterfly *Melitaea cinxia*. *Ecology* 75:747-762.
- Hanski I, Pakkala T, Kuussaari M, Lei GC (1995) Metapopulation persistence of an endangered butterfly in a fragmented landscape. *Oikos* 72:21-28.

- Hanski I, Breuker CJ, Schops K, Setchfield R, Nieminen M (2002) Population history and life history influence the migration rate of female Glanville fritillary butterflies. *Oikos* 98:87-97.
- Hanski I, Erälahti C, Kankare M, Ovaskainen O, Sirén H (2004) Variation in migration propensity among individuals maintained by landscape structure. *Ecol Lett* 7:958-966. DOI: 10.1111/j.1461-0248.2004.00654.x.
- Hayden EJ, Ferrada E, Wagner A (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474:92-95. DOI: <http://www.nature.com/nature/journal/v474/n7349/abs/10.1038-nature10083-unlocked.html#supplementary-information>.
- Hedrick PW (2004) Recent developments in conservation genetics. *For Ecol Manage* 197:3-19.
- Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH (2008) How many species are infected with *Wolbachia*? - a statistical analysis of current data. *Fems Microbiology Letters* 281:215-220. DOI: 10.1111/j.1574-6968.2008.01110.x.
- Hill JK, Thomas CD, Huntley B (1999) Climate and habitat availability determine 20th century changes in a butterfly's range margin. *Proc R Soc Lond Ser B Biol Sci* 266:1197-1206.
- Hill JK, Thomas CD, Fox R, Telfer MG, Willis SG, Asher J, Huntley B (2002) Responses of butterflies to twentieth century climate warming: implications for future ranges. *Proc R Soc Lond Ser B Biol Sci* 269:2163-2171. DOI: 10.1098/rspb.2002.2134.
- Hiroki M, Tagami Y, Miura K, Kato Y (2004) Multiple infection with *Wolbachia* inducing different reproductive manipulations in the butterfly *Eurema hecabe*. *Proc R Soc Lond Ser B Biol Sci* 271:1751-1755. DOI: 10.1098/rspb.2004.2769.
- Hobson R, Budd P (2001) The Marsh Fritillary in Cornwall: Site Dossier. Incorporating the Bodmin Moor Larval Survey 2000. Report No. S01-32, Butterfly Conservation, Wareham.
- Hobson R, Bourn NAD, Warren MS, Brereton TM (2001) The Marsh Fritillary in England: A Review of Status and Habitat Condition. Report no. S01-31, Butterfly Conservation, Wareham.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet* 6. DOI: 10.1371/journal.pgen.1000862.

- Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecol* 21:797-807. DOI: 10.1007/s10980-005-5245-9.
- Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6:65-70.
- Hornett EA, Engelstadter J, Hurst GDD (2010) Hidden cytoplasmic incompatibility alters the dynamics of male-killer/host interactions. *J Evol Biol* 23:479-487. DOI: 10.1111/j.1420-9101.2009.01872.x.
- Hornett EA, Charlat S, Wedell N, Jiggins CD, Hurst GDD (2009) Rapidly Shifting Sex Ratio across a Species Range. *Curr Biol* 19:1628-1631. DOI: 10.1016/j.cub.2009.07.071.
- Hornett EA, Charlat S, Duploup AMR, Davies N, Roderick GK, Wedell N, Hurst GDD (2006) Evolution of male-killer suppression in a natural population. *PLoS Biol* 4:1643-1648. DOI: 10.1371/journal.pbio.0040283.
- Hornett EA, Duploup AMR, Davies N, Roderick GK, Wedell N, Hurst GDD, Charlat S (2008) You can't keep a good parasite down: Evolution of a male-killer suppressor uncovers cytoplasmic incompatibility. *Evolution* 62:1258-1263. DOI: 10.1111/j.1558-5646.2008.00353.x.
- Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T (2010) *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci U S A* 107:769-774. DOI: 10.1073/pnas.0911476107.
- Hovestadt T, Nieminen M (2009) Costs and benefits of dispersal in butterflies, in: J Settele et al. (Eds.), *Ecology of butterflies in Europe*, Cambridge University Press, Cambridge. pp. 97 - 106.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9:1322-1332.
- Hula V, Konvicka M, Pavlicko A, Fric Z (2004) Marsh Fritillary (*Euphydryas aurinia*) in the Czech Republic: monitoring, metapopulation structure, and conservation of an endangered butterfly. *Entomol Fenn* 15:231-241.
- Hurst GDD, Werren JH (2001) The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet* 2:597-606.
- Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited

- symbionts. *Proceedings of the Royal Society B-Biological Sciences* 272:1525-1534. DOI: 10.1098/rspb.2005.3056.
- Hurst GDD, Jiggins FM, von der Schulenburg JHG, Bertrand D, West SA, Goriacheva, II, Zakharov IA, Werren JH, Stouthamer R, Majerus MEN (1999) Male-killing *Wolbachia* in two species of insect. *Proc R Soc Lond Ser B Biol Sci* 266:735-740.
- Jacobs MS, Wade MJ (2003) A Synthetic Review of the Theory of Gynodioecy. *The American Naturalist* 161:837-851.
- Jensen J, Bohonak A, Kelley S (2005) Isolation by distance, web service. *BMC Genetics* 6:13. v.3.16 <http://ibdws.sdsu.edu/>
- Jeong H, Kim J, Im H, Jeong H, Hong M, Lee J, Han Y, Kim I (2009) Mitochondrial DNA Sequence Variation of the Swallowtail Butterfly, *Papilio xuthus*, and the Cabbage Butterfly, *Pieris rapae*. *Biochem Genet* 47:165-178.
- Jeyaprakash A, Hoy MA (2000) Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of sixty-three arthropod species. *Insect Mol Biol* 9:393-405. DOI: 10.1046/j.1365-2583.2000.00203.x.
- Ji YJ, Zhang DX (2004) Characteristics of microsatellite DNA in lepidopteran genomes and implications for their isolation. *Acta Zool Sin* 50:608-614.
- Ji YJ, Zhang DX, Hewitt GM, Kang L, Li DM (2003) Polymorphic microsatellite loci for the cotton bollworm *Helicoverpa armigera* (Lepidoptera : Noctuidae) and some remarks on their isolation. *Mol Ecol Notes* 3:102-104. DOI: 10.1046/j.1471-8286.2003.00366.x.
- Jiggins FM, Hurst GDD, Majerus MEN (2000a) Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. *Proceedings of the Royal Society B-Biological Sciences* 267:69-73.
- Jiggins FM, Hurst GDD, Dolman CE, Majerus MEN (2000b) High-prevalence male-killing *Wolbachia* in the butterfly *Acraea encedana*. *J Evol Biol* 13:495-501.
- Jiggins FM, Hurst GDD, Schulenburg J, Majerus MEN (2001) Two male-killing *Wolbachia* strains coexist within a population of the butterfly *Acraea encedon*. *Heredity* 86:161-166.
- Joyce DA, Pullin AS (2001) Phylogeography of the marsh fritillary *Euphydryas aurinia* (Lepidoptera : Nymphalidae) in the UK. *Biol J Linn Soc* 72:129-141.
- Joyce DA, Pullin AS (2003) Conservation implications of the distribution of genetic diversity at different scales: a case study using the marsh fritillary butterfly

- (*Euphydryas aurinia*). Biol Conserv 114:453-461. DOI: 10.1016/s0006-3207(03)00087-9.
- Junker M, Schmitt T (2010) Demography, dispersal and movement pattern of *Euphydryas aurinia* (Lepidoptera: Nymphalidae) at the Iberian Peninsula: an alarming example in an increasingly fragmented landscape? J Insect Conserv 14:237-246. DOI: 10.1007/s10841-009-9250-1.
- Kageyama D, Nishimura G, Hoshizaki S, Ishikawa Y (2002) Feminizing *Wolbachia* in an insect, *Ostrinia furnacalis* (Lepidoptera: Crambidae). Heredity 88:444-449.
- Kalinowski ST (2011) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity 106:625-632.
- Kandpal RP, Kandpal G, Weissman SM (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. Proc Natl Acad Sci U S A 91:88-92.
- Kankare M, van Nouhuys S, Gaggiotti O, Hanski I (2005) Metapopulation genetic structure of two coexisting parasitoids of the Glanville fritillary butterfly. Oecologia 143:77-84. DOI: 10.1007/s00442-004-1782-1.
- Keyghobadi N, Roland J, Strobeck C (1999) Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). Mol Ecol 8:1481-1495.
- Kittayapong P, Jamnongluk W, Thipaksorn A, Milne JR, Sindhusake C (2003) *Wolbachia* infection complexity among insects in the tropical rice-field community. Mol Ecol 12:1049-1060. DOI: 10.1046/j.1365-294X.2003.01793.x.
- Klapwijk M (2008) Spatial ecology of host-parasitoid interactions: a threatened butterfly and its specialized parasitoid. PhD Thesis chapter.
- Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. Trends Ecol Evol 21:629-637.
- Kong L, Li Q (2008) Development of expressed sequence tag-derived microsatellite markers for the sea urchin *Hemicentrotus pulcherrimus*. Molecular Ecology Resources 8:152-154.
- Konvicka M, Hula V, Fric Z (2003) Habitat of pre-hibernating larvae of the endangered butterfly *Euphydryas aurinia* (Lepidoptera : Nymphalidae): What can be learned from vegetation composition and architecture? Eur J Entomol 100:313-322.

- Koop JL, Zeh DW, Bonilla MM, Zeh JA (2009) Reproductive compensation favours male-killing *Wolbachia* in a live-bearing host. *Proceedings of the Royal Society B-Biological Sciences* 276:4021-4028. DOI: 10.1098/rspb.2009.1230.
- Kuussaari M, Saccheri I, Camara M, Hanski I (1998) Allee effect and population dynamics in the Glanville fritillary butterfly. *Oikos* 82:384-392.
- Lande R (1988) Genetics and Demography in Biological Conservation. *Science* 241:1455-1460.
- Leakey R, Lewin R (1996) *The Sixth Extinction: Biodiversity and its Survival*. Phoenix, London.
- Lee TM, Jetz W (2008) Future battlegrounds for conservation under global change. *Proceedings B: Biological Sciences* 275:1261-1270.
- Legendre P, Legendre L (1998) *Numerical Ecology*, 2<sup>nd</sup> Edition. Elsevier, New York.
- Lehmann T, Hawley WA, Kamau L, Fontenille D, Simard F, Collins FH (1996) Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. *Heredity* 77:192-200.
- Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull Entomol Soc Am* 15:237-240.
- Lewis OT, Hurford C (1997) Assessing the status of the marsh fritillary butterfly (*Eurodryas aurinia*): An example from Glamorgan, UK. *J Insect Conserv* 1:159-166.
- Liu WH, Wang YF, Xu RM (2006) Habitat utilization by ovipositing females and larvae of the Marsh fritillary (*Euphydryas aurinia*) in a mosaic of meadows and croplands. *J Insect Conserv* 10:351-360. DOI: 10.1007/s10841-006-9009-x.
- Louis EJ, Dempster ER (1987) An Exact Test for Hardy-Weinberg and Multiple Alleles. *Biometrics* 43:805-811.
- Lushai G, Fjellsted W, Marcovitch O, Aagaard K, Sherratt TN, Allen JA, Maclean N (2000) Application of molecular techniques to non-lethal tissue samples of endangered butterfly populations (*Parnassius apollo* L.) in Norway for conservation management. *Biol Conserv* 94:43-50.
- Maes D, Van Dyck H (2001) Butterfly diversity loss in Flanders (north Belgium): Europe's worst case scenario? *Biol Conserv* 99:263-276.
- Marden JH, Fescemyer HW, Saastamoinen M, MacFarland SP, Vera JC, Frilander MJ, Hanski I (2008) Weight and nutrition affect pre-mRNA splicing of a muscle gene associated with performance, energetics and life history. *J Exp Biol* 211:3653-3660. DOI: 10.1242/jeb.023903.

- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembien LA, Berka J, Braverman MS, Chen YJ, Chen ZT, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu PG, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-380. DOI: 10.1038/nature03959|ISSN 0250-6971.
- Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Berge JB, Devonshire AL, Guillet P, Pasteur N, Pauron D (1998) Molecular characterization of pyrethroid knockdown resistance (*kdr*) in the major malaria vector *Anopheles gambiae* s.s. *Insect Mol Biol* 7:179-184.
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652-654.
- Meglec E, Solignac M (1998) Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). *Hereditas* 128:179-180.
- Meglec E, Pecsénye K, Varga Z, Solignac M (1998) Comparison of differentiation pattern at allozyme and microsatellite loci in *Parnassius mnemosyne* (Lepidoptera) populations. *Hereditas* 128:95-103.
- Meglec E, Petenian F, Danchin E, D'Acier AC, Rasplus JY, Faure E (2004) High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Mol Ecol* 13:1693-1700.
- Metzger C, Ferchaud A-L, Geiser C, Ursenbacher S (2011) New polymorphic microsatellite markers of the endangered meadow viper (*Vipera ursinii*) identified by 454 high-throughput sequencing: when innovation meets conservation. *Conservation Genetics Resources*. DOI: 10.1007/s12686-011-9411-x.
- Mikheyev AS, Vo T, Wee B, Singer MC, Parmesan C (2010) Rapid microsatellite isolation from a butterfly by *de novo* transcriptome sequencing: performance and a comparison with AFLP-derived distances. *PLoS ONE* 5:e11212.

- Moilanen A, Hanski I (1998) Metapopulation dynamics: Effects of habitat quality and landscape structure. *Ecology* 79:2503-2515.
- Munguira ML, Martin J, GarciaBarros E, Viejo JL (1997) Use of space and resources in a Mediterranean population of the butterfly *Euphydryas aurinia*. *Acta Oecologica-International Journal of Ecology* 18:597-612.
- Narita S, Shimajiri Y, Nomura M (2009) Strong cytoplasmic incompatibility and high vertical transmission rate can explain the high frequencies of *Wolbachia* infection in Japanese populations of *Colias erate poliographus* (Lepidoptera: Pieridae). *Bull Entomol Res* 99:385-391. DOI: 10.1017/s0007485308006469.
- Neargarder G, Dahlhoff EP, Rank NE (2003) Variation in thermal tolerance is linked to phosphoglucose isomerase genotype in a montane leaf beetle. *Funct Ecol* 17:213-221. DOI: 10.1046/j.1365-2435.2003.00722.x.
- Nève G (2009) Population genetics of butterflies, in: J Settele et al. (Eds.), *Ecology of Butterflies in Europe*, Cambridge University Press, Cambridge. pp. 107 - 129.
- Nève G, Meglec E (2000) Microsatellite frequencies in different taxa. *Trends Ecol Evol* 15:376-377.
- Nève G, Barascud B, Descimon H, Baguette M (2000) Genetic structure of *Proclossiana eunomia* populations at the regional scale (Lepidoptera, Nymphalidae). *Heredity* 84:657-666.
- Nève G, Barascud B, Descimon H, Baguette M (2008) Gene flow rise with habitat fragmentation in the bog fritillary butterfly (Lepidoptera: Nymphalidae). *BMC Evol Biol* 8:84.
- New TR, Pyle RM, Thomas JA, Thomas CD, Hammond PC (1995) Butterfly Conservation Management. *Annu Rev Entomol* 40:57-83. DOI: doi:10.1146/annurev.en.40.010195.000421.
- Nice CC, Gompert Z, Forister ML, Fordyce JA (2009) An unseen foe in arthropod conservation efforts: The case of *Wolbachia* infections in the Karner blue butterfly. *Biol Conserv* 142. DOI: :10.1016/j.biocon.2009.08.020.
- Nieminen M, Siljander M, Ilkka H (2004) Structure and dynamics of *Melitaea cinxia* metapopulations, in: P Ehrlich and H Ilkka (Eds.), *On the wings of checkerspots: a model system for population biology*, Oxford University Press, New York. pp. 63 - 91.
- Nieminen M, Singer MC, Fortelius W, Schops K, Hanski I (2001) Experimental confirmation that inbreeding depression increases extinction risk in butterfly populations. *Am Nat* 157:237-244.

- Niitepold K (2010) Genotype by temperature interactions in the metabolic rate of the Glanville fritillary butterfly. *J Exp Biol* 213:1042-1048. DOI: 10.1242/jeb.034132.
- Niitepold K, Smith AD, Osborne JL, Reynolds DR, Carreck NL, Martin AP, Marden JH, Ovaskainen O, Hanski I (2009) Flight metabolic rate and Pgi genotype influence butterfly dispersal rate in the field. *Ecology* 90:2223-2232.
- Orsini L, Wheat CW, Haag CR, Kvist J, Frilander MJ, Hanski I (2009) Fitness differences associated with Pgi SNP genotypes in the Glanville fritillary butterfly (*Melitaea cinxia*). *J Evol Biol* 22:367-375. DOI: 10.1111/j.1420-9101.2008.01653.x.
- Ostrander EA, Jong PM, Rine J, Duyk G (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences* 89:3419-3423.
- Ovaskainen O, Hanski I (2004a) Metapopulation dynamics in highly fragmented landscapes, in: I Hanski and OE Gaggiotti (Eds.), *Ecology, Genetics and Evolution of Metapopulations*, Elsevier Academic Press, London. pp. 73 - 104.
- Ovaskainen O, Hanski I (2004b) From individual behavior to metapopulation dynamics: Unifying the patchy population and classic metapopulation models. *Am Nat* 164:364-377.
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288-295.
- Peet NB, Watkinson AR, Bell DJ, Sharma UR (1999) The conservation management of *Imperata cylindrica* grassland in Nepal with fire and cutting: an experimental approach. *J Appl Ecol* 36:374-387.
- Pereira HM, Leadley PW, Proença V, Alkemade R, Scharlemann JPW, Fernandez-Manjarrés JF, Araújo MB, Balvanera P, Biggs R, Cheung WWL, Chini L, Cooper HD, Gilman EL, Guénette S, Hurtt GC, Huntington HP, Mace GM, Oberdorff T, Revenga C, Rodrigues P, Scholes RJ, Sumaila UR, Walpole M (2010) Scenarios for Global Biodiversity in the 21st Century. *Science* 330:1496-1501. DOI: 10.1126/science.1196624.
- Petenian F, Meglecz E, Genson G, Rasplus JY, Faure E (2005) Isolation and characterization of polymorphic microsatellites in *Parnassius apollo* and *Euphydryas aurinia* (Lepidoptera). *Mol Ecol Notes* 5:243-245.

- Peterson MA, Denno RF (1998) The Influence of Dispersal and Diet Breadth on Patterns of Genetic Isolation by Distance in Phytophagous Insects. *The American Naturalist* 152:428-446.
- Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The Future of Biodiversity. *Science* 269:347-350. DOI: 10.1126/science.269.5222.347.
- Porter K (1981) The population dynamics of small colonies of the butterfly *Euphydryas aurinia*. PhD Thesis, Oxford University, Oxford.
- Porter K (1982) Basking behaviour in larvae of the butterfly *Euphydryas aurinia*. *Oikos* 38:308-312.
- Porter K (1983) Multivoltinism in *Apanteles bignellii* and the influence of weather on synchronization with its host *Euphydryas aurinia*. *Entomol Exp Appl* 34:155-162.
- Porter K (2007) The marsh fritillary *Euphydryas aurinia* (Rott.) in Cumbria: turning the tide. *Atropos* 31:34-43.
- Porter K, Ellis S (2011) Securing viable metapopulations of the Marsh Fritillary butterfly, *Euphydryas aurinia*, (Lepidoptera: Nymphalidae) in Northern England. *J Insect Conserv* 15:111-119. DOI: 10.1007/s10841-010-9328-9.
- Pöyry J, Lindgren S, Salminen J, Kuussaari M (2005) Responses of butterfly and moth species to restored cattle grazing in semi-natural grasslands. *Biol Conserv* 122:465-478. DOI: 10.1016/j.biocon.2004.09.007.
- Prasad MD, Muthulakshmi M, Madhu M, Archak S, Mita K, Nagaraju J (2005) Survey and Analysis of Microsatellites in the Silkworm, *Bombyx mori*: Frequency, Distribution, Mutations, Marker Potential and Their Conservation in Heterologous Species. *Genetics* 169:197-214. DOI: 10.1534/genetics.104.031005.
- Primmer CR (2009) From Conservation Genetics to Conservation Genomics. *Ann N Y Acad Sci* 1162:357-368. DOI: 10.1111/j.1749-6632.2009.04444.x.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Pritchard JK, Wen X, Falush D (2010) *Documentation for Structure software: Version 2.3*. Available from: <http://pritch.bsd.uchicago.edu/structure.html> (accessed 12th May 2011).
- R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>. Accessed 21 April 2011.

- Rice WR (1989) Analyzing Tables of Statistical Tests. *Evolution* 43:223-225.
- Rodriguez S, Gaunt TR, Day INM (2009) Hardy-Weinberg Equilibrium Testing of Biological Ascertainment for Mendelian Randomization Studies. *Am J Epidemiol* 169:505-514. DOI: 10.1093/aje/kwn359.
- Rokas A (2000) *Wolbachia* as a speciation agent. *Trends Ecol Evol* 15:44-45.
- Rousset F (1997) Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. *Genetics* 145:1219-1228.
- Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8:103-106. DOI: 10.1111/j.1471-8286.2007.01931.x.
- Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for General Users and for Biologist Programmers, in: S Krawetz and S Misener (Eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Humana Press, Totowa, NJ. pp. 365–386.
- Russell JE, Stouthamer R (2011) The genetics and evolution of obligate reproductive parasitism in *Trichogramma pretiosum* infected with parthenogenesis-inducing *Wolbachia*. *Heredity* 106:58-67. DOI: 10.1038/hdy.2010.48.
- Saastamoinen M (2007) Dispersal and related life history traits in the Glanville fritillary butterfly. PhD Thesis, University of Helsinki, Helsinki.
- Saastamoinen M, Ikonen S, Hanski I (2009) Significant effects of Pgi genotype and body reserves on lifespan in the Glanville fritillary butterfly. *Proceedings of the Royal Society B-Biological Sciences* 276:1313-1322. DOI: 10.1098/rspb.2008.1464.
- Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature* 392:491-494.
- Saccheri IJ, Wilson IJ, Nichols RA, Bruford MW, Brakefield PM (1999) Inbreeding of Bottlenecked Butterfly Populations: Estimation Using the Likelihood of Changes in Marker Allele Frequencies. *Genetics* 151:1053-1063.
- Sasaki T, Kubo T, Ishikawa H (2002) Interspecific Transfer of *Wolbachia* Between Two Lepidopteran Insects Expressing Cytoplasmic Incompatibility: A *Wolbachia* Variant Naturally Infecting *Cadra cautella* Causes Male Killing in *Ephesia kuehniella*. *Genetics* 162:1313-1319.
- Saunders DA, Hobbs RJ, Margules CR (1991) Biological Consequences of Ecosystem Fragmentation: A Review. *Conserv Biol* 5:18-32. DOI: 10.1111/j.1523-1739.1991.tb00384.x.

- Schtickzelle N, Baguette M (2009) (Meta)population viability analysis: a crystal ball for the conservation of endangered butterflies?, in: J Settele et al. (Eds.), *Ecology of Butterflies in Europe*, Cambridge University Press, Cambridge.
- Schtickzelle N, Mennechez G, Baguette M (2006) Dispersal depression with habitat fragmentation in the bog fritillary butterfly. *Ecology* 87:1057-1065.
- Schtickzelle N, Chouff J, Goffart P, Fichet V, Baguette M (2005) Metapopulation dynamics and conservation of the marsh fritillary butterfly: Population viability analysis and management options for a critically endangered species in Western Europe. *Biol Conserv* 126:569-581. DOI: 10.1016/j.biocon.2005.06.030.
- Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18:233-234.
- Schulenburg J, Hurst GDD, Huigens TME, van Meer MMM, Jiggins FM, Majerus MEN (2000) Molecular evolution and phylogenetic utility of *Wolbachia* *ftsZ* and *wsp* gene sequences with special reference to the origin of male-killing. *Mol Biol Evol* 17:584-600.
- Severns PM, Boldt L, Villegas S (2006) Conserving a wetland butterfly: quantifying early lifestage survival through seasonal flooding, adult nectar, and habitat preference. *J Insect Conserv* 10:361-370. DOI: 10.1007/s10841-006-9011-3.
- Shaw MR, Stefanescu C, van Nouhuys S (2009) Parasitoids of European butterflies, in: J Settele et al. (Eds.), *Ecology of Butterflies in Europe*, Cambridge University Press, Cambridge. pp. 130 - 156.
- Shoemaker DD, Ross KG, Keller L, Vargo EL, Werren JH (2000) *Wolbachia* infections in native and introduced populations of fire ants (*Solenopsis* spp.). *Insect Mol Biol* 9:661-673.
- Sigaard P, Pertoldi C, Madsen AB, Sogaard B, Loeschcke V (2008) Patterns of genetic variation in isolated Danish populations of the endangered butterfly *Euphydryas aurinia*. *Biol J Linn Soc* 95:677-687. DOI: 10.1111/j.1095-8312.2008.01078.x.
- Sinama M, Dubut V, Costedoat C, Gilles A, Junker M, Malausa T, Martin J-F, Neve G, Pech N, Schmitt T, Zimmermann M, Meglecz E (2011) Challenges of microsatellite development in Lepidoptera: *Euphydryas aurinia* (Nymphalidae) as a case study. *Eur J Entomol* 108:261-266.
- Singer MC, Ilkka H (2004) Dispersal behaviour and evolutionary metapopulation dynamics, in: P Ehrlich and H Ilkka (Eds.), *On the wings of checkerspot: a model system for population biology* Oxford University Press, New York. pp. 181 - 198.

- Slatkin M (1995) A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics* 139:457-462.
- Smee M, Smyth W, Tunmore M, Ffrench-Constant R, Hodgson D (2011) Butterflies on the brink: habitat requirements for declining populations of the marsh fritillary (*Euphydryas aurinia*) in SW England. *J Insect Conserv* 15:153-163. DOI: 10.1007/s10841-010-9334-y.
- Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J (2010) Adaptation genomics: the next generation. *Trends Ecol Evol* 25:705-712.
- Stewart KEJ, Bourn NAD, Thomas JA (2001) An evaluation of three quick methods commonly used to assess sward height in ecology. *J Appl Ecol* 38:1148-1154.
- Tagami Y, Miura K (2004) Distribution and prevalence of *Wolbachia* in Japanese populations of Lepidoptera. *Insect Mol Biol* 13:359-364.
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512-526.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* (submitted).
- Tay WT, Behere GT, Heckel DG, Lee SF, Batterham P (2008) Exon-primed intron-crossing (EPIC) PCR markers of *Helicoverpa armigera* (Lepidoptera: Noctuidae). *Bull Entomol Res* 98:509-518.
- Thomas C, Hanski I (2004) Metapopulation dynamics in changing environments: butterfly responses to habitat and climate change, in: I Hanski and OE Gaggiotti (Eds.), *Ecology, Genetics and Evolution of Metapopulations*, Elsevier Academic Press, London. pp. 489 - 514.
- Thomas CD (1994) Extinction, colonization, and metapopulations - environmental tracking by rare species. *Conserv Biol* 8:373-378.
- Thomas CD, Wilson RJ, Lewis OT (2002) Short-term studies underestimate 30-generation changes in a butterfly metapopulation. *Proc R Soc Lond Ser B Biol Sci* 269:563-569. DOI: 10.1098/rspb.2001.1939.
- Thomas J (1980) Why Did the Large Blue Become Extinct in Britain? *Oryx* 15:243-247. DOI: doi:10.1017/S0030605300024625.

- Thomas JA (1995) The conservation of declining butterfly populations in Britain and Europe: priorities, problems and successes. *Biol J Linn Soc* 56:55-72. DOI: 10.1111/j.1095-8312.1995.tb01120.x.
- Thomas JA, Simcox DJ, Clarke RT (2009) Successful Conservation of a Threatened Maculinea Butterfly. *Science* 325:80-83. DOI: 10.1126/science.1175726.
- Thomas JA, Bourn NAD, Clarke RT, Stewart KE, Simcox DJ, Pearman GS, Curtis R, Goodger B (2001) The quality and isolation of habitat patches both determine where butterflies persist in fragmented landscapes. *Proc R Soc Lond Ser B Biol Sci* 268:1791-1796.
- Tilman D, May RM, Lehman CL, Nowak MA (1994) Habitat destruction and the extinction debt. *Nature* 371:65-66.
- Tolman T, Lewington R (2009) *Collin's Butterfly Guide*, Harper Collins Publishers, London.
- Travis JMJ (2003) Climate change and habitat destruction: a deadly anthropogenic cocktail. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270:467-473. DOI: 10.1098/rspb.2002.2246.
- Tunmore M, Robinson F, Hodgson D (2008) Marsh fritillary monitoring on nine sites in mid-Cornwall: Flight and larval web surveys. A report to Mid-Cornwall Moors LIFE Project, Cornwall Environmental Consultants Ltd, Truro.
- Turner EC, Granroth HVM, Johnson HR, Lucas CBH, Thompson AM, Froy H, German RN, Holdgate R (2009) Habitat preference and dispersal of the Duke of Burgundy butterfly (*Hamearis lucina*) on an abandoned chalk quarry in Bedfordshire, UK. *J Insect Conserv* 13:475-486. DOI: 10.1007/s10841-008-9194-x.
- United Nations (2010) *World Urbanization Prospects, The 2009 Revision: Highlights*. New York, Department of Economic and Social Affairs, Population Division
- Van't Hof AE, Brakefield PM, Saccheri IJ, Zwaan BJ (2007) Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. *Heredity* 98:320-328.
- Van't Hof AE, Zwaan BJ, Saccheri IJ, Daly D, Bot ANM, Brakefield PM (2005) PRIMER NOTE Characterization of 28 microsatellite loci for the butterfly *Bicyclus anynana*. *Mol Ecol Notes* 5:169-172.
- Van Dyck H, Van Strien AJ, Maes D, Van Swaay CAM (2009) Declines in Common, Widespread Butterflies in a Landscape under Intense Human Use

- Declinaciones en Mariposas Comunes, con Distribución Amplia, en un Paisaje Bajo Uso Humano Intenso. *Conserv Biol* 23:957-965. DOI: 10.1111/j.1523-1739.2009.01175.x.
- Van Meer MMM, Witteveldt J, Stouthamer R (1999) Phylogeny of the arthropod endosymbiont *Wolbachia* based on the *wsp* gene. *Insect Mol Biol* 8:399-408. DOI: 10.1046/j.1365-2583.1999.83129.x.
- Van Nouhuys S, Lei GC (2004) Parasitoid-host metapopulation dynamics: the causes and consequences of phenological asynchrony. *J Anim Ecol* 73:526-535.
- van Nouhuys S, Hanski I (2004) Natural Enemies of Checkerspots, in: PR Ehrlich and I Hanski (Eds.), *On the Wings of Checkerspots: A Model System for Population Biology*, Oxford University Press, Oxford. pp. 161 - 180.
- van Swaay CAM, Warren MS (1999) Red Data book of European butterflies (*Rhopalocera*). No. 99 ed. Council of Europe Publishing, Strasbourg.
- Varshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A (2007) Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. *Plant Science* 173:638-649.
- Vianna P, Schama R, Russo CAM (2003) Genetic divergence and isolation by distance in the West Atlantic sea anemone *Actinia bermudensis* (McMurrich, 1889). *J Exp Mar Biol Ecol* 297:19-30.
- Wahlberg N, Moilanen A, Hanski I (1996) Predicting the occurrence of endangered species in fragmented landscapes. *Science* 273:1536-1538.
- Wahlberg N, Klemetti T, Hanski I (2002) Dynamic populations in a dynamic landscape: the metapopulation structure of the marsh fritillary butterfly. *Ecography* 25:224-232.
- Wang RJ, Wang YF, Lei GC, Xu RM, Painter J (2003) Genetic differentiation within metapopulations of *Euphydryas aurinia* and *Melitaea phoebe* in China. *Biochem Genet* 41:107-118.
- Wang RJ, Wang YF, Chen JJ, Lei GC, Xu RM (2004) Contrasting movement patterns in two species of chequerspot butterflies, *Euphydryas aurinia* and *Melitaea phoebe*, in the same patch network. *Ecol Entomol* 29:367-374.
- Warren MS (1994) The UK status and suspected metapopulation structure of a threatened European butterfly, the marsh fritillary *Eurodryas aurinia*. *Biol Conserv* 67:239-249.

- Warren MS, Hill JK, Thomas JA, Asher J, Fox R, Huntley B, Roy DB, Telfer MG, Jeffcoate S, Harding P, Jeffcoate G, Willis SG, Greatorex-Davies JN, Moss D, Thomas CD (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature* 414:65-69.
- Watt WB (1977) Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: biochemical and population aspects. *Genetics* 87:177-194.
- Watt WB (1983) Adaptation at specific loci. II. Demographic and biochemical elements in the maintenance of the *Colias* PGI polymorphism. *Genetics* 103:691-724.
- Watt WB (1992) Eggs, enzymes, and evolution: natural genetic variants change insect fecundity. *Proceedings of the National Academy of Sciences* 89:10608-10612.
- Watt WB, Cassin RC, Swan MS (1983) Adaptation at specific loci. III. Field behavior and survivorship differences among *Colias* PGI genotypes are predictable from *in vitro* biochemistry. *Genetics* 103:725-739.
- Watt WB, Carter PA, Blower SM (1985) Adaptation at specific loci. IV. Differential mating success among glycolytic allozyme genotypes of *Colias* butterflies. *Genetics* 109:157-175.
- Wee P-S (2004) Effects of Geographic Distance, Landscape Features and Host Association on Genetic Differentiation of Checkerspot Butterflies. PhD Thesis, The University of Texas, Austin.
- Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38:1358-1370.
- Wells CN, Williams RS, Walker GL, Haddad NM (2009) Effects of Corridors on Genetics of a Butterfly in a Landscape Experiment. *Southeast Nat* 8:709-722.
- Werren JH, Windsor DM (2000) *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proc R Soc Lond Ser B Biol Sci* 267:1277-1285.
- Werren JH, Windsor D, Guo L (1995) Distribution of *Wolbachia* among Neotropical Arthropods. *Proceedings B: Biological Sciences* 262:197-204.
- Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology* 6:741-751. DOI: 10.1038/nrmicro1969.
- Wheat C, W. , Haag C, R. , Marden J, H. , Hanski I, Frilander M, J. (2010) Nucleotide Polymorphism at a Gene (Pgi) under Balancing Selection in a Butterfly Metapopulation. *Mol Biol Evol* 27:267-282.

- Wheat CW (2010) Phosphoglucose isomerase (Pgi) performance and fitness effects among Arthropods and its potential role as an adaptive marker in conservation genetics. *Conserv Genet* 11:387-397. DOI: 10.1007/s10592-009-0042-7.
- Wheat CW, Watt WB, Boutwell CL (2005) A reconnaissance of population genetic variation in arctic and subarctic sulfur butterflies (*Colias* spp.; Lepidoptera, Pieridae). *Canadian Journal of Zoology-Revue Canadienne De Zoologie* 83:1614-1623. DOI: 10.1139/z05-159.
- Wilson RJ, Thomas CD (2002) Dispersal and the spatial dynamics of butterfly populations.
- Wilson RJ, Roy DB (2009) Butterfly population structure and dynamics, in: J Settele et al. (Eds.), *Ecology of Butterflies in Europe*, Cambridge University Press, Cambridge. pp. 81 - 96.
- Wilson RJ, Gutierrez D, Gutierrez J, Monserrat VJ (2007) An elevational shift in butterfly species richness and composition accompanying recent climate change. *Glob Change Biol* 13:1873-1887.
- Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M, Powell W (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. *Mol Ecol* 14:1681-1695.
- Wright S (1931) Evolution in Mendelian Populations. *Genetics* 16:97 - 159.
- Wright S (1943) Isolation by Distance. *Genetics* 28:114-138.
- Wright S (1978) *Evolution and the Genetics of Populations*. University of Chicago Press, Chicago.
- Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends Ecol Evol* 11:413-418.
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11:1-16.
- Zhang DX (2004) Lepidopteran microsatellite DNA: redundant but promising. *Trends Ecol Evol* 19:507-509. DOI: 10.1016/j.tree.2004.07.020.
- Zhang M, Cao TW, Zhong Y, Guo YP, Ma EB (2011) Phylogeny of Limenitidinae Butterflies (Lepidoptera: Nymphalidae) Inferred from Mitochondrial Cytochrome Oxidase I Gene Sequences. *Agric Sci Chin* 10:566-575. DOI: 10.1016/s1671-2927(11)60038-2.

- Zhou WG, Rousset F, O'Neill S (1998) Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc R Soc Lond Ser B Biol Sci 265:509-515.
- Zhou X, Faktor O, Applebaum SW, Coll M (2000) Population structure of the pestiferous moth *Helicoverpa armigera* in the Eastern Mediterranean using RAPD analysis. Heredity 85:251-256.
- Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA, Shagin DA (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res 32:e37.
- Zimmermann M, Wahlberg N, Descimon H (2000) Phylogeny of *Euphydryas* checkerspot butterflies (Lepidoptera : Nymphalidae) based on mitochondrial DNA sequence data. Ann Entomol Soc Am 93:347-355.