UNIVERSITY OF SOUTHAMPTON

The Adonis Blue (*Polyommatus bellargus*) in the UK: A Molecular Investigation into its Colonisation and Population Structure.

Georgina Harper

Thesis submitted for the degree of Doctor of Philosophy

Division of Biodiversity and Ecology
School of Biological Sciences
Faculty of Science

May 2001

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE

BIODIVERSITY AND ECOLOGY RESEARCH DIVISION

Doctor of Philosophy

THE ADONIS BLUE IN THE UK: A MOLECULAR INVESTIGATION INTO ITS COLONISATON AND POPULATION STUCTURE

By Georgina Harper

Populations of *P. bellargus* were examined to assess whether the genetic structuring reflects a recent colonisation event in the UK, as is inferred by the historical data for the species. To achieve this, the AT-rich mitochondrial control region was sequenced from individuals throughout the UK range and southern France. Genetic diversity was found to be much lower in the UK, with only three closely related haplotypes among fifty UK specimens, whereas six were identified from eight butterflies from France. The predominant UK haplotype, which was at fixation in most UK populations, was almost identical to one from France, suggesting that the UK population has a French origin. The only reasonable explanation for the observed patterns of MtDNA diversity is a recent and rapid colonisation event from France via one or a few female butterflies.

Five microsatellite DNA markers were isolated and used to examine the relationship between spatial distribution and population genetic structure among a subset of UK populations of *P. bellargus*. The butterfly is at its northern range limit in the UK, where it exists in a highly fragmented metapopulation structure on isolated pockets of calcareous grassland. These populations were affected by a severe population bottleneck in the late 1970's, when a drought caused large numbers of the host plant (*Hippocrepis comosa*) to wilt. Mantel tests indicated a significant effect of isolation by distance among the UK populations, a relationship that broke down at greater geographic scales (>37.5km), probably because of large areas of unsuitable habitat presenting barriers to gene flow. A significant relationship was also found between population size and levels of genetic diversity, inferring elevated levels of inbreeding within smaller populations. However, populations that were isolated or had undergone a bottleneck or founder event did not exhibit a similar reduction in genetic diversity. The results indicate that population connectivity and size are the predominant characteristics that must be maintained for the conservation of *P. bellargus* in the UK and other invertebrates with similar metapopulation structures.

Investigations into the suitability of museum preserved lepidopteran material for DNA analysis revealed that small mtDNA amplicons and microsatellites are amplifiable from all ages of specimens tested (0 to 105 years). Dried specimens were also found to be suitable for a RAPD identification of microsatellites (see Ender *et al*, 1996). This work demonstrates the enormous potential of dried, stored lepidopteran material for use in molecular ecological studies. In a pilot study, microsatellites were used to investigate the genetic consequences of a population bottleneck, by comparing pre-bottleneck and post-bottleneck measurements of genetic variability from a natural system. Analysis of DNA from both modern and 105-year-old museum specimens of *P. bellargus* from a population at Folkestone revealed the loss of specific alleles (known to have been present in 1896) following a demographic contraction. Lost alleles included common ones present in other sampled populations and others unique to the Folkestone population. Whilst trends were found for a post-bottleneck reduction in heterozygosity and allele number per locus, the relationships were not found to be significant.

CONTENTS

ACKNOWLEDGEMENTS

This work was funded by a University of Southampton PhD studentship. I would particularly like to thank my supervisors: Dave Goulson, for his advice and guidance over the last three years, and Norman Maclean for both guidance and providing most of the facilities that enabled me to do this research. I am grateful to everyone who helped in identifying *P. bellargus* populations and the accompanying historical data (particularly: Jeremy Thomas; Joyce Gay; Mike Fuller; Colin Pope; Tony Tutton; Nigel Bourn; Martin Warren; Steven Davis; David Shepphard; Nick Johanssen). I would especially like to thank Butterfly Conservation, English Nature, The National Trust, English Heritage, and the many landowners that allowed me to capture butterflies on their land.

I would have achieved much smaller sample sizes without the dedicated help of Sue Thomas (who tirelessly accompanied me to most of the field sites), Dave (for his unparalleled ability to catch adonis blues), Jane Stout (for magnificent leaps to catch adonis blues!), Matt Henderson (who pretended to help by pointing out butterflies while he sunbathed!), Andrea Kells, and Simon Bryant (for the French butterflies), whilst George McGavin and David Smith both kindly donated preserved material from their butterfly collections.

The following people all helped me get to grips with the molecular techniques and gave me advice about the many approaches to analysing the data: Arati Iyengar, Mark Dixon, Alex Rogers, Stuart Piertney, Matt Henderson, Gugs Lushai, Sanit Piyapattanakorn, Charlotte Allender, and Federica Pannacciulli.

I would also like to thank all of my friends in Southampton, especially Lauryn Bradley, for making it an enjoyable place to work. Finally I would like to thank my Family for their support and enthusiasm over the years, they are all now able to identify an adonis blue butterfly!

1.0 GENERAL INTRODUCTION

1.1. INTRODUCTION

This project seeks to investigate temporal and spatial variations in genetic diversity of the adonis blue butterfly (*Polyommatus* (*Lysandra*) *bellargus* Rottemburg) across its UK range and to assess the effects of habitat fragmentation on the population structure of the species. The study incorporates both wild caught and museum preserved specimens, and examined genetic variation using both mitochondrial and nuclear (microsatellite) DNA markers.

1.2. HABITAT FRAGMENTATION, INBREEDING AND EXTINCTION

Habitat destruction caused by intensive agriculture and urban development is fragmenting the distribution of many formerly widespread species. This is particularly the case in the United Kingdom, where many species are now confined to small isolated pockets of habitat. This situation has affected many butterfly species, including *P. bellargus*, causing them to exist within a metapopulation structure. A metapopulation is where small fragmented populations are partially isolated from one another, and hence are prone to undergo a repeated cycle of extinction and recolonisation. The ecology of metapopulations has been extensively discussed in theory and also with regard to its relevance for conservation (Gilpin & Hanski, 1991; Gilpin, 1991; Hanski and Gilpin, 1996). The population dynamics of metapopulations are also likely to cause repeated "bottlenecks" or "founder effects" within populations, events which have also been considered at length in theory (Wright, 1931 & 1978; Nunney & Campbell, 1993; Lynche *et al.*, 1995; Lande, 1995). These demographic changes are likely to exacerbate the effects of genetic drift, tending to encourage geographically separated populations to become increasingly genetically distinct from one another, whilst also encouraging inbreeding and probably an associated loss of genetic variation, particularly in the smaller, isolated, fringe populations (Wright, 1951; Frankham, 1995a). This reduction in

genetic variation, in particular allelic diversity, seems likely to have a profound effect upon a population's ability to adapt to habitat changes, such as climatic extremes, diseases, pests and parasites. Potentially it may lead to the eventual extinction of the colony (Hoffmann and Parsons, 1991; Pray *et al.*, 1994; Frankham, 1995a & 1995b; Heschel and Paige, 1995).

Although it may seem obvious that inbreeding will depress the reproductive fitness of outbreeding species (Wright, 1977), there has been much controversy as to whether inbreeding depression has significant effects on wild populations and its relevance to conservation (Caughley, 1994; Wauters *et al.*, 1994; Frankham 1995b). However, there is increasing evidence for inbreeding causing a reduction in fitness: a number of studies claim to have found strong evidence for inbreeding depression in captive populations (Bryant *et al.*, 1990; Ralls & Ballou, 1986; Laikre & Ryman, 1991; Saccheri *et al.*, 1996; Saccheri, 1999) and from a few experimental (Jiminez *et al.*, 1994; Newman & Pilsen, 1997) and observational field studies (Stockley, *et al.*, 1993; Thornhill, 1993). Work on captive populations is likely to underestimate the situation, because the effects on fitness are typically more severe in harsher environments, and therefore likely be more acute in the wild than in captivity (Hoffman & Parsons, 1991; Miller, 1994), whereas experimental investigations inevitably involve anthropogenically induced changes in a population. For example, experiments have demonstrated a distinct advantage to the dynamics of mice (*Mus*) and topminnow (*Poeciliopsis*) populations after the transfer of individuals from genetically diverse populations to those with a narrow gene base (Berry *et al.*, 1991; Vrijenhoek, 1994). However, conclusive evidence for inbreeding depression in wild populations has been more elusive (e.g. Caro & Laurenson, 1994; Caughley, 1994; Frankham, 1995b; Mallett, 1996 & 1997), for example in the cheetah (*Acinonyx jubatus*), where there has been much debate over the effects of inbreeding on fitness (Merola, 1994; O'Brien, 1994), a disagreement that stems from the lack of non-inbred cheetahs as controls. However, one published study of natural populations has demonstrated the underlying relationship between inbreeding and extinction. This work investigated the metapopulation structure of the glanville fritillary butterfly (*Melitaea*

cinxia), and clearly showed that the risk of extinction increased dramatically with decreasing heterozygosity, an indication of inbreeding (Saccheri *et al.*, 1998).

The majority of these published studies to date involving natural populations have investigated inbreeding in vertebrate species. This seems somewhat short-sighted when it is considered that the vast majority of animal species in the world are invertebrates.

These studies do include research on a few lepidopteran species, using both natural and captive bred populations. For example: the silver studded blue (*Plebejus argus*) (Brookes *et al.*, 1997); *Bicyclus anynana* (Saccheri *et al.*, 1996 & 1999); the small tortoiseshell (*Aglais urticae*) (Vandewoestijine, 1999); *Helicoverpa zea* (Mallett *et al.*, 1993) and, as already mentioned, the glanville fritillary (*Melitaea cinxia*) (Saccheri *et al.*, 1998). All of these studies present some evidence for a reduction in heterozygosity associated with inbreeding as a result of founder events or isolation of small populations and one study indicates the negative effect this has on levels of fitness and population survival (Saccherri *et al.*, 1998). Most of these investigations into the effects of habitat fragmentation have used allozyme electrophoresis and in some cases a mitochondrial DNA marker. However, it is likely that strong selection controls protein polymorphism, so that allozymes may not give an accurate representation of past demographic events such as bottlenecks and instead reflect current selection on the loci (Avise, 1994; Goulson, 1993). Only in one case (Saccheri, 1998) was the allozyme data backed up by a selectively neutral marker, such as a microsatellite, and even in this study the power of the analysis was reduced because only a one microsatellite was used. Ideally, a set of five or six markers should be used in parallel (Hughes and Queller, 1993; Ruzzante, 1998).

There is growing evidence for detrimental effects of inbreeding, which is causing increasing concern over the loss of genetic variation in small, isolated populations and the possible effects on the long-term viability of endangered species. Clearly more research is necessary, particularly within invertebrate species, to investigate the effects of population declines, habitat fragmentation and associated inbreeding. This project will

investigate these effects in a butterfly species, *P. bellargus*, which is under increasing threat from the destruction and fragmentation of its UK habitat. This species is also recovering from the effects of a severe genetic bottleneck in 1978-1981, an event that may have had severe consequences for the genetic diversity of the species. The project will not use allozyme markers, but a suite of polymorphic microsatellite markers alongside a mitochondrial DNA marker.

1.3. THE ECOLOGY OF THE ADONIS BLUE

The adonis blue (*P. bellargus*) is one of the more spectacular of the butterfly species native to the British Isles. The butterfly is strongly dimorphic, with the upper side of the male's wings an iridescent cobalt blue, while the female is a deep brown colour (see plate 1.1). The wings of the species are fringed with a chequered border, due to the dark ends of the wing veins, a feature that clearly distinguishes the butterfly from the very similar common blue (*Polyommatus icarus*). The underside of the wings of both sexes is characterized by a unique spotted pattern.

Plate 1.1. An adult male *Polyommatus bellargus* (Rottemburg).
Note the distinctivly chequered wing margin.

The butterfly is at the north western limit of its range in England, where it is confined to the warmer southern counties of Berkshire, Dorset, Hampshire, Isle of Wight, Kent, Sussex, Surrey and Wiltshire (Thomas, 1983a; Emmet & Heath 1990; Bourn & Warren, 1998; Bourn *et al.*, 1999; Stewart *et al.*, 2000; Asher *et al.*, 2001). The former range of

the butterfly also included Berkshire, Bedfordshire, Oxfordshire, Gloucestershire, Hertfordshire and Somerset. A large proportion of the butterflie's colonies are in Dorset, which seems to be the main-stay of the butterfly (Bourn *et al.*, 1999; Stewart *et al.*, 2000), though there are significant populations elsewhere, and the number of populations, though fluctuating, does appear to be on the increase (Pearman *et al.* 1998). The butterfly occurs only on south facing calcareous grassland (see plate 1.2), where its larval food plant, horseshoe vetch (*Hippocrepis comosa*), grows in the short vegetation (Thomas, 1983a).

Plate 1.2. A typical example of the habitat occupied by *P. bellargus.* This site is a south-facing escarpment above Folkestone, Kent.

The species is bi-voltine, with adults on the wing from mid-May to June and August to mid-September. Populations are almost always present as discrete colonies, with only a few larger than 1000 butterflies (Emmet & Heath, 1990; Bourn *et al.*, 1999; Stewart *et al.*, 2000). The butterflies are sedentary, with little movement or interchange between the populations; even the more mobile males will rarely venture in to unsuitable habitat (Thomas, 1983a & b). The females are extremely sedentary, spending long periods resting on the ground in sheltered hollows while their eggs develop (Thomas 1983a). After mating the females then search for suitable plants on which they will lay their egg load. They are extremely particular when it comes to sites for laying their eggs, choosing only *H. comosa*, which must be growing amidst short vegetation (typically 1-4cm) or

bare earth. This choice of short vegetation is in contrast to sites chosen by females of the species in warmer climates, where a taller mean sward length is chosen. It is likely that the sheltered short turf is necessary because it provides a warm microclimate for larval development, and because it is favoured by ants which have a strong mutualistic relationship with both the pupae and the larvae (Thomas, 1983a & b, Heath *et al*, 1984).

The ants (*Lasius alienus* and *Myrmica sabuleti*) tend the larvae, milking them via the Newcomer's organ for the sweet honeydew that is produced from the second instar onwards. Up to 30 ants have been seen tending a single larva, though numbers are generally around four to six. The larvae are often found inside the nests of the ants that tend them (Thomas, 1983a).

The butterfly's reliance on mutualism with ants, alongside the very specific climate and habitat requirements of the species has meant that it is very susceptible to sudden declines. Severe changes in climate, such as drought, may adversely affect the species, because south facing slopes are likely to be the where the food plant will wilt first – being warmer and drier. *P. bellargus* numbers were badly hit by an extreme drought in 1976, when its food plant wilted badly, having a knock-on fatal effect on larval development (Thomas, 1983a), resulting in around 90% loss of colonies nationwide (Asher *et al.*, 2001). The butterfly is also directly susceptible to climatic changes, causing butterfly numbers to fluctuate from year to year depending on climatic conditions – even a particularly wet summer can cause a decline.

1.4. WHY CHOOSE THE ADONIS BLUE?

Polyommatus bellargus is, for several reasons, a particularly interesting species to study. As a result of its sedentary nature and very specific habitat requirements, many populations of the species are isolated from one another. However, there are situations where neighbouring groups of populations within a geographical area are almost contiguous. This tends to be the case in localities, such as on the Isle of Purbeck in Dorset, where a chalk or limestone escarpment creates a fragmented length of south facing calcareous grassland. However, for the majority of the UK, the distances between suitable habitats can be large, with most or all of the intervening land being unsuitable habitat for the species. There are six main geographic regions where populations can be found: the North Downs, the South Downs, northern and southern Dorset, Salisbury Plain, and the Isle of Wight. Although the majority of populations across Dorset and Salisbury Plain are within reasonably close proximity to one another, these populations are well separated from other suitable sites across southern England. Populations on the Isle of Wight and Portland are particularly well isolated from other colonies, not only by unsuitable land habitat, but also by expanses of water.

P. bellargus underwent a chronic decline from the 1950's to the late 1970's, probably at least in part as a result of a variety of environmental pressures, including: an increase of land allocated to intensive farming, plus reductions in stock grazing and a decline in rabbits, caused by myxamatosis, both resulted in an increase in mean sward length. The number of populations in the UK plunged to its lowest point shortly after the drought of 1976, after a rapid decline of over 90% (Thomas, 1983a; Emmet and Heath, 1990; Pearman *et al*, 1998; Asher *et al*., 2001), because of the severe effects encountered by the host plant, *Hippocrepis comosa*. Since 1981, the species has been recovering its former distribution, with the number of colonies in 1997 being almost double that documented in 1978-1981 (Pearman *et al*, 1998; Emmet and Heath, 1990, Asher *et al*., 2001). Many populations must have recovered from just a few surviving individuals, while others that had become extinct may have been recolonised by just a single gravid female. Butterflies are unusual that, in the UK at least, a great deal is known about the history of individual

populations. In the case of *P. bellargus* we know with some certainty the details of which populations became extinct, and which were little affected by the 1976 drought.

1.5. AIMS OF THE PROJECT

This project will employ microsatellite and mitochondrial DNA analysis, to investigate the current and historical genetic structure of *P. bellargus* populations in the UK. This will be combined with information from documented historical data to draw conclusions about how current trends and past events, such as genetic bottlenecks, are affecting the population dynamics of *P. bellargus*. The main aims are as follows:

1) Using microsatellites and mitochondrial DNA, a survey of population structure at both local and national levels will be carried out. Populations will be chosen according to both their locality within the butterflies range, and according to their documented history.

2) Populations will be assessed for levels of genetic diversity and inbreeding, this will be correlated with their documented histories and size.

3) The relationship between the UK and French populations will be elucidated, and a possible route of colonisation for the UK population will be investigated.

4) Preserved specimens will be analysed to elucidate the putative effects of a decline on temporal genetic variation.

5) Measures to improve the conservation of the species within the UK will be discussed.

2.0 THE USE OF MOLECULAR METHODS IN ENTOMOLOGICAL STUDIES

2.1. INTRODUCTION

Though it has long been recognised that individuals vary phenotypically, variation at the molecular level has been described only relatively recently. Molecular methods have a number of applications within the field of entomology just as they do in many other biological areas, from investigations into relationships between individuals to relationships between populations, even relationships between species. Extensive use has already been made of molecular techniques in the field of population biology.

The study of population biology was radically revolutionised when Harris (1966), and Lewontin and Hubby (1966) showed that genetic diversity within a population could be estimated by the use of electrophoresis to demonstrate protein variation (allozyme electrophoresis), and thus to begin to reveal elements of the population structure. This technique has spread rapidly within the field of population genetics. However, in more recent times a new and more exciting approach to population biology has become available for use. Rather than assessing genetic variation by analysing proteins, a number of more direct approaches have been devised that are based upon examining variations at the level of the DNA sequence, consequently the data obtained are informative because they display all of the variation present (Avise, 1994). Another advantage of using DNA markers is that only a minute amount of tissue is necessary, whereas allozyme electrophoresis necessitates the homogenising of the entire insect. The former is invaluable when studying insects, particularly endangered species. This requirement for only a small amount of material also lends itself particularly well to investigations using preserved insects, such as museum specimens, because tissue removal will cause a minimum amount of damage to valuable specimens (Cano & Poinar, 1993; Mitten, 1994; Roy *et al.*, 1994). Allozyme analysis is impossible with preserved specimens because the functional enzymes are viable only for a short while after death, unless preserved at –70°C. Nevertheless, information from the protein polymorphisms (allozymes) is still relevant and may be the most suitable method available for many studies (Berg &

Hamrick, 1997). The costs of DNA based molecular techniques may limit their applications, as could the need for the expensive equipment necessary for carrying them out.

The array of techniques currently available is vast. This includes allozyme electrophoresis; restriction fragment length polymorphisms (RFLP); single and multi-locus (minisatellite) fingerprinting, using a variety of probes; directed amplification of minisatellite regions (DAMD); mitochondrial and nuclear DNA sequencing; microsatellite analysis (simple sequence repeats); randomly amplified polymorphic DNA (RAPD), and PCR-ELISA (enzyme linked immuno-sorbent assay). The technique chosen from this plethora of methods will depend heavily on the nature of the investigation being undertaken (Parker *et al.*, 1998). It is critical to any investigation that the technique used is suited to the research in terms of both costs and results. This review will examine the different techniques available and will assess their individual merits and applications.

2.2. REVIEW OF TECHNIQUES AVAILABLE

2.2.1. Allozyme electrophoresis

Allozymes are protein polymorphisms that represent themselves as band differences on a starch, cellulose actetate, or acrylamide gel. Each enzyme can be stained using a unique reaction mixture containing the necessary substrates and co-factors for visualisation. Many of the enzymes are ubiquitous between organisms, though generally only a subset is detectable in each and not all of these will be variable. The polymorphisms result from slight allelic differences in enzymes at particular loci; differences that manifest themselves as changes in the shape and/or charge to mass ratio, therefore affecting the speed at which a protein is electrophoresed through a gel. The samples used are crude tissue homogenates from fresh material; no further purification is needed because of the specificity of the staining procedure.

All organisms will be heterozygous or homozygous for a particular allele, represented by two bands or a single band respectively on a gel. This banding pattern may become more complex if enzymes contain multiple sub-units i.e. dimers or trimers *etc.*, as the markers will segregate in a co-dominant Mendelian fashion.

The technique has been used for three decades and has advantages in that it is cheap, easy and relatively fast to carry out (Watt, 1994). However, its applications are limited. Only a portion of the genetic variation can be resolved (approximately one third) (Lewontin, 1991); many insects, especially parasitic Hymenoptera, show very little variability in allozymes, possibly as a result of the haplodiploid population structure (Crespi, 1991; Blanchetot & Packer, 1992). Inbreeding and small effective population size may also affect levels of variation (Graur, 1985). Any bands that co-migrate will be assumed to be homozygous and patterns of ancestry and descent between alleles will be impossible to define (Roderick, 1996). The fact that the enzymes are functional proteins means that they may be subject to selection, and as such cannot be classed as a truly neutral marker (Avise, 1994). This will have some impact on population studies. Another disadvantage is that freshly killed or deep-frozen insects must be homogenised for analysis, therefore, destructive sampling is unavoidable for the procedure.

There are many situations where allozyme electrophoresis can be used to elucidate useful genetic information and they have been applied in an enormous number of studies. The method has many practical applications, including:

(i) The analysis of population structure, including the assessment of levels of genetic variation within and between populations. The technique has been extensively used to investigate this in a huge variety of organisms including many studies on insects (e.g. Lewontin and Hubby, 1966; Krafsur *et al.*, 1995; Scarpassa *et al.*, 1996; Crease *et al.*, 1997; Chapman *et* al., 1999; Rodriguez *et* al., 2000; Wang & Grace, 2000). A number of these studies have investigated population structure in lepidopteran species, including the silver studded blue (*Plejebus argus*); the golden sun moth (*Synemon plana*); *Parnassius mnemosyne*; the glanville fritillary (*Melitaea cinxia*); the small ermine moth

Yponomeuta padellus (Brookes *et al.*, 1997; Meglecz *et al.*, 1997 & 1999; Saccheri *et al.*, 1998; Raijmann & Menken, 2000; Clarke & O'Dwyer, 2000; Sula & Spitzer, 2000).

(ii) To reveal paternity and kinship relationships between individuals (e.g. Page & Robinson, 1994; Calos & Sakaluk, 1998).

(iii) The examination of taxonomic relationships (e.g. Jackson & Resh, 1998; Audisio *et* al., 2000; Colin, 2000; Fanciulli *et al.*, 2000) and investigation of speciation (e.g. Filchak *et al.*, 1999)

(iv) The identification of and analysis of variation in clonal lineages (e.g. Simon, *et al.*, 1995).

It is clear from this brief summary that allozymes have been used in an immense number of population studies. In the majority of more recent studies they are used in parallel with a second, generally more variable marker – often a region of the mitochondrial genome.

2.2.2. DNA extraction from fresh and preserved specimens

The extraction of DNA is an important aspect of any molecular investigation, since the purity and amount obtained determine the success of subsequent procedures. For example if the DNA is to be digested by restriction enzymes or to be used in the construction of a genomic library, it is important to remove proteins and RNA from the sample. This can be achieved by using a stringent phenol-chloroform extraction, possibly followed by a caesium chloride gradient purification (Cockburn & Fritz, 1996). If the DNA is required for basic PCR amplification, then its purity from contaminating proteins and organic matter may not be as crucial (Hammond *et al.*, 1996). Nevertheless, contamination with other DNA sources will cause problems, and the presence of nuclease enzymes are likely to result in DNA degradation.

The technique of extracting DNA from fresh insect material is relatively simple, and a multitude of protocols used for DNA isolation from other organisms will usually suffice (Cockburn & Fritz, 1996). However, extraction from older specimens may give more inconsistent results. This can be due to a number of factors, the most important of which is likely to be the degradation of the DNA (Tan and Orrego, 1992; Post *et al.*, 1993; Cameron, 1993; Cooper, 1994; Oakenfull, 1994; Whitfield, 1996).

A variety of techniques have been used to purify DNA from preserved specimens. For example, Cano *et al.*, (1993) used Chelex$^{@}$-100 (BioRad, UK) to release DNA from amber-entombed weevils. It has previously been noted that DNA extracted via this method is prone to rapid degradation, possibly because the DNA is not separated from proteins, particularly nucleases, which will quickly break down the DNA. To avoid this, an extraction with Glassmilk Spinbuffer can be useful (Cockburn & Fritz, 1996). Aljanabi and Martinez (1997) demonstrated that a salt-extraction method could isolate DNA suitable for PCR from a variety of preserved organisms including a grasshopper species (*Schistocerca pallens*). Hammond *et al.* (1996) used a guanidium thiocyanate extraction to isolate DNA from a variety of preserved air-dried insects, demonstrating that the DNA was suitable for a subsequent PCR amplification.

The extraction of DNA from older preserved specimens presents a variety of problems. In the majority of cases, the insects from these sources have not been stored in an ideal manner for the preservation of DNA. Butterflies for instance, are generally stored air dried, leaving the DNA susceptible to oxidative processes that may modify pyrimidines, sugar residues, baseless sites and intermolecular links (Paabo, 1989; Cockburn & Fritz, 1996). After death, the period prior to desiccation is a crucial stage in storage because hydrolytic enzymes will rapidly degrade the DNA (Roy *et al.*, 1994; Dillon *et al.*, 1996). Thus, desiccation speed will influence the level of degradation (Paabo, 1989). The result of this damage to the DNA means it is generally only useful for PCR-based studies (Thomas, 1989; Ellegren, 1991; Cano and Poinar, 1993; Mitton, 1994; Hammond *et al.*, 1996).

In older samples, fragmentation of the DNA will limit the amplifiable range of PCR to less than 300 base pairs (Paabo, 1989; Cockburn & Fritz, 1996). This will make it impossible to isolate entire genes, except by building up sequence data from contiguous PCR amplified fragments. This method was successfully used by Cano *et al.* (1993) to sequence the mitochondrial 16S rRNA gene and its internal spacers and thereby deduce the taxonomy of ancient amber entombed weevil. Hammond *et al.* (1996) reproducibly used random amplified polymorphic DNA (RAPD) analysis, a technique requiring significant lengths of DNA (300-3,000bp), in seven-year-old air-dried butterflies. Markers such as microsatellites have also been successfully amplified from preserved specimens (Roy *et al.*, 1994; Meglecz *et* al., 1998). The small size of these markers lend themselves particularly well to studies using preserved specimens. The mitochondrial genome has a high copy number within each cell (around 200 copies), as a result of the numbers of mitochondria present. This increases the likelihood of relevant mitochondrial DNA (mtDNA) regions being intact and available for amplification (Paabo, 1989; Thomas *et al.*, 1989; Wayne & Jenks, 1991; Moritz, 1994; Taylor *et al.*, 1994; Cockburn and Fritz, 1996).

One proviso of PCR amplification, particularly when using preserved specimens, is that because of the sensitivity of the technique, the presence of contaminating DNA may cause erroneous results. This is of paramount importance if a "universal" PCR primer is being used (i.e. primers designed for highly conserved DNA sequences that are not species specific). For example, DNA from necrophagous beetles, mites or fungal growths found on dried insects may contaminate the DNA extracted from museum specimens. If this results in enough contaminating DNA, subsequent PCRs may amplify the younger, less damaged DNA, resulting in spurious PCR products. Even DNA from human sources can cause problems. In this situation, DNA sequencing would be a useful method to identify the true origin of any PCR amplified products (Lister, 1994).

The use of DNA from museum specimens adds another dimension to the study of genetics: changes over time as well as space can be investigated. For example, examining evolutionary changes within a species occurring over decades or even

centuries. Even extinct populations or species that are hard to access in the wild can be studied (Cano & Poinar, 1993). Taxonomic questions can also be addressed (Cano *et al.*, 1993).

When working on preserved insects it is important to realise that few collectors will welcome destructive sampling, or removal of significant amounts of tissue from their specimens, because of the damage this would inflict. Philips and Simon (1995) have suggested a non-destructive method to extract DNA from dried insects, which may be acceptable to collectors: they have recovered DNA by submersing the insect exoskeleton in an extraction solution and soaking overnight at 68°C. However, this method would be of limited use with butterflies because of the delicacy of the specimens, especially the wings. Perhaps by minimising the tissue amounts required, the problem of damaging preserved specimens could be overcome. An obvious choice for butterflies would be to remove one or more legs for DNA extraction.

The problems with obtrusive sampling exist not just for museum specimens, but also with wild populations, particularly in studies of rare or endangered organisms. Rose *et al.* (1994) successfully PCR amplified DNA from wing clippings of silver studded blue butterflies (*Plebejus argus*). This technique may be equally as effective in stored specimens as it is in fresh, because butterfly wings dry out soon after pupation. In a study of the glanville fritillary butterflies (*Melitaea cinxia*) in Finland, the problems with obtrusive sampling were overcome by capturing gravid females from a wild population, then allowing larval development in the laboratory. The offspring were then released into the wild and the female used for molecular analysis (Saccheri *et al.*, 1998).

2.2.3. Mitochondrial DNA (MtDNA)

The mitochondrial genome is usually 16-18kb in size, with one genotype per individual, although there are exceptions to this. Bark weevils of the genus *Pissodes* have a 30-36kb mitochondrial genome and are often heteroplasmic for mtDNA (Taylor *et al.*, 1993;

Mitton, 1994; Lunt *et al.*, 1998). In contrast to the diploid nature of nuclear DNA, the mitochondrial genome is haploid. This is because the genome is maternally inherited, and thus the genotype of any individual is identical to that of its female parent. As a result of this, the genome does not undergo recombination (Avise, 1991) and consequently variation accumulates by mutation alone, with no subsequent reshuffling. This means that the mitochondrial genome can only represent a single segregating unit, reducing the power of the analysis; therefore mtDNA should always be used in conjunction with nuclear DNA markers (Moritz, 1994; Roderick, 1996). However, this non-mendelian inheritance pattern does mean that mtDNA can be a powerful tool in the detection of founder effects and population bottlenecks (Roderick, 1996; Taylor *et al.*, 1997), because unlike nuclear DNA, the genetic evidence in the mtDNA for these events will not subsequently be confused by the recombination and segregation of alleles. One important proviso of this maternal inheritance pattern is that gene flow from male migration cannot be quantified, and the effective population size will be reduced (Avise, 1991 & 1994).

When both nuclear and mitochondrial markers are used in parallel, their results do not always concur. Evidence has shown that there is not always a clear correlation between diversity in the nuclear and mitochondrial genomes. For example, low mitochondrial and high nuclear diversity has been reported in rapidly expanding populations of species such as geckos, (Moritz, 1991), whereas high mitochondrial and low nuclear diversity has been observed in declining species such as humpback whales (Baker *et al.*, 1993). Both these examples represent cases where there have been recent demographic changes in the populations; the disparity between the nuclear and mitochondrial markers may be attributable to their differing inheritance patterns and mutation rates.

The mtDNA genome has a number of rapidly evolving regions, which have been used to investigate the population structure and dynamics of many organisms, including insects. These parts of the genome, many of which code for functional genes, have been shown to have a higher mutation rate than equivalent nuclear DNA regions. However, there are exceptions to this: some *Drosophila* species appear to have an equal rate of mutation

between nuclear and mtDNA, although this may result from their extremely high nuclear mutation rate (Vawter & Brown, 1986). Interspersed between these highly variable mitochondrial regions are extremely conserved sequences. Primers designed within these conserved areas can be used to amplify the more variable regions, even across the species boundary. These are known as "universal primers" (Simon *et al.*, 1994; Roehrdanz, 1995; Lunt *et al.*, 1998).

The high mutation rate of certain regions means that they are particularly suitable for studies of variation within and between closely related populations and species. For example, the hypervariable "AT-Rich-Control region" of insect mtDNA has been used in many studies, from analysing genetic differentiation between *Drosophila melanogaster* populations (Hale and Singh 1991), to the elucidation of species relationships within the genus *Jalmenus* (Lepidoptera: *Lycaenidae*) (Taylor *et al.*, 1993). Other mitochondrial regions have a slower mutation rate and are applicable to studies of phylogenetic change; for example, the gene encoding the 16S ribosomal DNA subunit has been used to examine phylogenetic relationships of more than 40 European butterfly species in the family *Satyridae* (Martin *et al.*, 2000). This variation in the mtDNA between individuals, populations or species can be investigated by a number of methods:

(i) Sequence analysis, restriction fragment length polymorphisms (RFLPs) of a PCR amplified product (PCR-RFLP).

(ii) PCR analysis of variable number tandem repeats (VNTR) within the mitochondrial sequence, particularly within insect AT-Rich control regions (Taylor *et al.*, 1993; Lunt *et al.*, 1998).

The principal applications of mtDNA include: (i) phylogenetic analysis and taxonomic investigations, (e.g. Sperling & Harrison, 1994; Brunton & Hurst, 1998; Zimmerman *et al.*, 2000) (ii) the identification of past demographic changes in populations (e.g. genetic bottlenecks and founder events) (e.g. Brower & Boyce, 1991; Brookes *et al.*, 1997; Groman & Pellmyr, 2000), and (iii) the elucidation of the genetical structuring of populations (e.g. Geurgas *et al.*, 2000; Segura, 2000; Taylor *et al.*, 2000)

2.2.4. Randomly Amplified Polymorphic DNA (RAPD)

This technique uses a PCR based approach, although instead of using a pair of site-specific primers, a single oligonucleotide primer of random sequence, generally around 10bp in length, is employed (Welsh & McClelland, 1990; Lynch & Milligan, 1994). Obviously, there are likely to be numerous sites within an entire genome that contain the 10bp of complimentary sequence appropriate for primer annealing, whereas site-specific primers rely on the fact that a *circa* 20bp length of sequence is likely to be unique. The result of this is that the RAPD primer amplifies a collection of DNA products of various sizes up to 3,000bp (Black, 1993; Mitton, 1994); hence the term randomly amplified polymorphic DNA. This arbitrary grouping of amplicons can be visualised on an ethidium bromide stained agarose gel or a silver stained acrylamide gel. The variation in band size and presence between different individuals results from polymorphisms in tandem repeats within the amplified products, or mutations that may alter the primer binding sites (Williams *et al.*, 1990). Thus each individual is represented by a unique banding pattern.

There are many drawbacks. The use of short non-specific primers means that in some cases they may bind poorly to the template DNA, resulting in a lack of reproducibility of the technique. As it is impossible to identify whether the bands originate from the same region of the template DNA, bands of similar size but different sequence, from different individuals, may be erroneously assumed to be homologous (Roderick, 1996). Reproducibility of the results may be hard to maintain, since small variations in methodology, such as variation in the source of *Taq* polymerase enzyme, or even using a different PCR machine, may affect the amplification (Black, 1993; Regner *et al.*, 2000). The application of the technique is questionable where the DNA is likely to be degraded to a significant extent, for example, in aged museum specimens because some primer sites may be lost and the banding pattern erratic (Black, 1993). Another problem could arise from variation in the concentration of the PCR products, if this is low, then some bands may be too faint to visualize and an incorrect genotype may be assigned (Black, 1993). Contamination of the PCR will also have serious effects on the banding pattern,

because the primers are not site-specific. This will cause anomalous and non-reproducible results. For this reason, a negative control should always be used. A large number of individuals from a population are necessary to test for divergence between populations (Yu & Peter Pauls, 1994). The markers are dominant (Hadrys *et al.*, 1992) and therefore no inferences can be made regarding heterozygosity or probable descent of alleles.

Aside from the technical difficulties listed above, the technique has many advantages, these include: (i) The technique can be very economical as the only significant expense is the *Taq* polymerase enzyme. (ii) A high amount of genetic diversity can be uncovered in a relatively small amount of time and effort, in contrast to other techniques such as allozymes, which often lack the necessary variability (Black, 1993; Graur, 1985). (iii) No prior DNA sequence knowledge is needed to design primers (Yu & Peter Pauls, 1994). (iv) Sampling can be carried out in an un-obtrusive manner because only a small amount of tissue is required (Lushai *et al.* 2000).

There are many applications for the technique:

(i) Taxonomy: Markers can be used to identify polymorphism that is present at the family or species level (e.g. Chapco *et al.*, 1992; Hadrys *et al.*, 1992; Wilkerson *et al.*, 1993; Altukhov & Abramova, 2000; Audisio *et al.*, 2000).

(ii) The identification of population structure (e.g. Zhou *et al.*, 2000), gene flow between populations (e.g. Lushai *et al.*, 2000) and detection of hybrid zones (e.g. Arnold *et al.*, 1991).

(iii) Assessment of kinship relationships within populations: The banding patterns for individuals can be used to assess whether two individuals are related (though it may not be possible to elucidate the specific nature of the relationship)(e.g. Hadrys *et al.*, 1992; Gerber *et al.*, 2000).

(iv) The generation of novel specific probes: Polymorphic RAPD bands can be investigated for the presence of microsatellite repeats using sequence analysis. A number of microsatellites markers have been isolated using this method

(Hadrys *et al.*, 1992; Roderick, 1996; Ramser *et al.*, 1997;Osakabe *et al.*, 2000).

(v) The identification of clonal lineages. Clonal lineages can be identified by the specific banding pattern represented by the RAPD-PCR. (e.g. Kreher *et al.*, 2000; Regner *et al.*, 2000).

For a review of the application of RAPDs in evolutionary and applied entomology, see Harry *et al.*, (1998). From the number of studies carried out using RAPDs, it is clear that this technique has been used in a large number of projects and has had a significant impact on the field of population biology.

2.2.5. Restriction Fragment Length Polymorphisms (RFLP)

This technique was widely used during the 1980's to study polymorphisms at the level of the nucleotide, particularly in mtDNA (Burke, 1994). The technique relies on the application of restriction endonuclease enzymes to cut the mitochondrial or nuclear DNA at multiple sites. The enzymes are highly specific for certain sequences, for example, *Msp* I will only cleave the DNA at CCGG sequence motifs. The length of the motif will dictate the frequency at which the enzyme cleaves the DNA. *Msp* I, for example, recognises a 4bp motif, so will cut on average every 256bp throughout the genome, whereas a 6bp cutter, such as *Bam* HI, will cut every 4096bp (Hoelzel & Dover, 1991). The application of frequent cutters such as *Msp* I to genomic DNA would result in such an enormous number of DNA fragments that if the DNA were simply run on an ethidium bromide stained agarose gel, the results would appear as a smear (Hoelzel & Dover, 1991). For this reason it is necessary to use one of two techniques: (i) radio-labelled, sequence specific probes can be applied with Southern blotting (Southern, 1975) and autoradiography to visualise the size of the bands. Or (ii) Specific regions of the genome can be amplified using PCR, (up to 3kb), these can then cut with restriction enzymes, electrophoresed and then ethidium bromide or silver stained to reveal polymorphisms between the cut fragments (PCR-RFLP). The method is similar to RAPDs, because polymorphisms are caused by both sequence variation at restriction sites and differences

in tandem repeat numbers between restriction sites, thus each individual will be represented by a different banding pattern on a gel (Mitton, 1994).

RFLPs can uncover a large amount of genetic information (Roderick, 1996). However, unless an initial PCR amplification is carried out, large amounts of DNA are necessary for good results (Erlich & Arnheim, 1992). If the investigation uses mtDNA, the isolation of mitochondria from the cell may be required. The manufacture and design of RFLP probes can be time consuming, unless information can be obtained about the organism's genome, or probes from a closely related species are available. If not, a suitable probe will need to be created either by screening cloned DNA fragments (Loxdale & Lushai, 1998) or amplifying and labelling specific regions of DNA.

It must be remembered that as with RAPDs, the banding patterns created by PCR-RFLP analysis are dominant and as such it is not possible to distinguish heterozygotes or make inferences about the probable descent of alleles (Avise, 1994).

There are many applications for the technique, including:

(i) Investigating taxonomy, phylogenetics and speciation. Markers can be used to identify polymorphism that is present at the family or species level (e.g. Sperling & Harrison, 1994; Deowanish *et al.,* 1996; Hall, 1990).

(ii) The identification of population structure, gene flow between populations and detection of hybrid zones (e.g. Haag *et al.*, 1993, Brookes *et al.*, 1997).

(iii) The identification of clonal lineages. Clonal lineages can be identified by an individual banding pattern (e,g, Birkle and Douglas, 1999).

The majority of these studies have used RFLP analysis on mitochondrial DNA, whilst a second marker type has also been used in parallel to investigate nuclear variation.

2.2.6. Amplified Fragment Length Polymorphism (AFLP)

A technique that combines elements of RFLP and RAPD analysis is amplified fragment length polymorphisms (AFLP). This method selectively amplifies products from a restriction digest of the genomic DNA, by ligating oligonucleotide adapter molecules to the digested fragments. The sequence of these adapters, combined with the sequence of the restriction enzyme recognition site serve as a primer binding site for PCR, allowing co-amplification of a subset of restriction fragments (see Vos *et al.*, 1995). In order to visualize the fragments produced, one primer is labelled either fluorescently or radioactively, then visualized and sized using autoradiography or an automated sequencer. The result is a specific banding profile for each individual tested.

An important advantage of this method is that it does not require the lengthy development associated with microsatellite loci, and it provides a more reliable banding pattern than RAPD profiling.

The AFLP technique has already been applied to population genetics, mostly of plants. For example, 429 AFLP markers have been used to investigate the partitioning of genetic variation in 11 endangered populations of the heart-of-palm, (*Euterpe edulis*) demonstrating highly significant evidence for differentiation between populations (Cardoso *et al.*, 2000). Another study used AFLP markers to efficiently detect low diversity within, and high variation between, populations of *Withania somnifera* (Negi *et al.*, 2000). An example of the technique being used in entomology is a study where the genetic relationships among biotypes of *Bemisia tabaci* (Hemiptera : Aleyrodidae) was analysed, based on AFLP analysis, resulting in the biotypes being successfully grouped together into four geographic clusters (Cervera *et al.*, 2000).

2.2.7. Minisatellite fingerprinting

Minisatellites are hypervariable DNA regions, consisting of tandem repeats of core sequences made up of between 15 to 60bp units, contiguously repeated up to several hundred times. Minisatellites are also known as variable number tandem repeats (VNTRs), and when non-coding they appear to be non-functional within the genome (Jeffreys *et al.*, 1985a). Allelic polymorphism arises as a result of variation in the copy number of the core sequence, which may arise because of unequal meiotic exchanges, or by replication slippage (Jeffreys *et al.*, 1985a). The polymorphisms can be detected by restricting the DNA with the appropriate endonuclease enzymes, electrophoresing the fragments on an agarose gel, Southern blotting the DNA onto nylon membrane and then hybridising with radioactively labelled probes (Southern, 1975). These probes consist of tandem arrays of the appropriate repeat unit being screened for (Blanchetot, 1992). The probes can be commercially obtained (e.g. A human derived marker: 33.15 (Carvalho *et al.*, 1991; Jeffreys *et al.*, 1985b)) or developed using analogous probes from other taxa (e.g. $(GATA)_4$ (Traut & Epplen, 1992; DeBarro *et al.*, 1994; Loxdale and Lushai, 1998)). This is possible because they are often ubiquitous between species, even those from different kingdoms. For example, sequences from the bacteriophage M13 are homologous to hypervariable regions in many species, including a bee, *Megachile rotunda* (Blanchetot, 1992).

Allelic polymorphism can also be identified directly using PCR. In order to amplify the variable number tandem repeat (VNTR), PCR primer pairs can be designed from published VNTR core sequences. When these are used to amplify the region, the allelic variation will be represented as a change in the charge to mass ratio during gel electrophoresis. Alternatively, these amplified fragments can be used as probes for Southern blotting genomic DNA (Heath *et al.*, 1993). Probes that hybridise to a single site within the genome are known as single locus probes (Hall, 1990), whereas those that bind to multiple sites are known as multiple locus probes (Jeffreys, 1985a). Single locus probes are co-dominant, allowing heterozygotes to be distinguished, but multiple locus

probes are dominant and although they may show more variation in the banding pattern, it is not possible to distinguish heterozygotes (Avise, 1994).

Minisatellites are useful in population genetics because they are highly variable, able to identify recent demographic changes such as genetic bottlenecks (Blanchetot & Packer, 1992), and also to quantify the structure between populations (Blanchetot, 1989 & 1991a; Reeve *et al.*, 1992; DeBarro *et al.*, 1994). Single locus minisatellites are also a powerful tool for identifying the kinship relationships between individuals in a sexually reproducing population (Blanchetot, 1991b & 1992; Chakraborty & Jin, 1993), whereas the multi-locus method is unable to define relationships wider than parental, unless the population is clonal (Carbone *et al.*, 1999).

Minisatellites can show vast amounts of variation compared with allozymes (Reeve *et al.*, 1992). The technique often requires significant amounts of high molecular weight DNA, and thus sampling of small organisms such as insects, from endangered populations without causing damage would be difficult. However, it is possible to use PCR to amplify single loci, using much smaller amounts of DNA. Minisatellites currently seem to have fallen out of favour; this is possibly because of the advent of simpler techniques such as microsatellites, which have become the marker of choice for many studies.

2.2.8. Microsatellite fingerprinting

Microsatellites are short lengths of tandemly repeated DNA sequences that occur throughout the genomes of many organisms, including insects (Roderick, 1996). Their core repeat units are short, ranging from one to six base pairs in length (Tautz & Renz, 1984; Litt & Lutty, 1989; Jarne & Lagoda, 1996), though di-, tri- and tetra-nucleotides are the most common. These repeat units are repeated within an array usually 100bp or less in length (Jarne & Lagoda, 1996). They are also known as simple sequence repeats (SSR) or short tandem repeats (STR) (Tautz & Renz, 1984; Hughes & Queller, 1993). The repeats can be present in exons but there seems to be a length limit here, probably as

a result of selective pressure (Sutherlands & Richards, 1995). The polymorphisms can easily be identified using a combination of PCR amplification (Saiki *et al.*, 1988) and acrylamide gel electrophoresis. The results can be visualised using silver staining, autoradiography or by more modern techniques using fluorescently labelled PCR primers with automated sequencing equipment, such as an ABI-PRISM-377 automated sequencer (PE-Biosystems, UK). The high conservation of sequences surrounding the repeats allows PCR primer design and in a few cases, these may even be conserved between related organisms, e.g. as found by Engel *et al.* (1996) among Artiodactyl species.

There are many advantages to using microsatellites as a genetic marker, and they are rapidly becoming the marker of choice in many population genetics studies (England *et al.*, 1996; Goldstein & Pollock, 1997; Maclean & Iyengar unpublished). They are superior to minisatellites and RAPDs in population studies because they are always co-dominantly inherited in a mendelian fashion, with a normal segregation of alleles, so that a pattern of descent can be attributed to particular alleles. Microsatellites can also detect variation where other methods, such as allozymes, fail (Hughes & Queller 1993; Meglecz *et al.*, 1998). They can be used to sample populations un-obtrusively because of the application of PCR, and can even be used where DNA has become degraded, such as in museum specimens (Queller *et al.*, 1993, Ellegren, 1991). The selective neutrality of microsatellites allows for rapid evolution of sequence variation and length (Roderick, 1996), though levels of variation can be slightly lower than some minisatellites (Edwards *et al.*, 1992; Blouin *et al.*, 1996). This variability makes them very useful in studying small-scale differences in closely related populations, or in studying recent demographic events such as genetic bottlenecks. The heterozygosity of loci in natural outbreeding populations can be well above 50%, peaking at almost 100% (Jarne & Lagoda, 1996).

The main disadvantage of microsatellites is that unless genomic information of the study organism is available, then it will be necessary to construct a genomic library, which is costly in both time and money (Roderick, 1996; McLean & Iyengar, unpublished). Although PCR-based methodologies, such as RAPDs, may overcome this need for library construction (Hadrys *et al.*, 1992; Ender *et* al., 1996; Roderick, 1996; Ramser *et al.*,

1997; Osakabe *et al.*, 2000), this method can be somewhat unreliable (Alex Rogers, personal communication). In some cases primers that are found to work in one species may amplify microsatellite loci in another related species (Traut & Epplen, 1992; Engel *et al.*, 1996). Another potential problem is that some taxa appear to be impoverished of microsatellite loci. A notable example is the order Lepidoptera, where many species seem deficient in microsatellite regions within their genome (Neve & Meglecz, 2000; Meglecz *et al.*, 1998; Meglecz & Solignac, 1998). As a consequence of this, very few lepidopteran studies have employed microsatellites, and even where they have been utilized, the numbers used have been sub-optimal. For example, just four microsatellites were used to uncover the influence of landscape on the population structure of *Parnassius smintheus* (Keyghobadi, 1998). This was the largest number of microsatellites used in any lepidopteran study, prior to the work detailed in this thesis. Ideally a minimum of five or six loci is necessary to increase the analytical power and reduce the inherent biases of studies employing microsatellites (Hughes and Queller, 1993; Ruzzante, 1998). One other proviso in the application of microsatellites to a study is the possibility of null alleles. This is where a mutation in the flanking regions of the microsatellite affects primer binding, resulting in a non-amplifying allele. The result of this is that individuals may be mistakenly assumed to be homozygous (Palo *et al.*, 1995; Pemberton *et al.*, 1995; Lehman *et al.*, 1996; Van Treuren, 1998). The presence of null alleles may be inferred by an excess of homozygotes within the data or by the repeated failure of PCR to amplify any alleles in an individual, particularly in kinship studies if the parental genotypes both appear to be homozygous (i.e. both are actually heterozygous with one null allele each) (Pemberton *et al.*, 1995).

To reduce the inherent costs involved in microsatellite analysis, it is possible to reduce the *Taq* polymerase requirements by multiplexing the amplification of several loci, *i.e.* use more than one primer set in a single PCR, thus simultaneously amplifying different loci within the genome (Wenburg *et al.*, 1996).

Microsatellites have been used in an enormous variety of studies, including those involved with population structure, taxonomy, phylogenetics, speciation, genetic

variability, the identification of clonal lineages, identification of recent demographic changes in populations, and kinship analysis. There are countless numbers of studies within each of these areas that have employed microsatellites, but barely a handful of these have analysed lepidopteran species (Palo *et al.* 1995; Meglecz *et al.*, 1998; Saccheri *et al.*, 1998; Keyghobadi *et al.*, 1999). For detailed reviews of the applications and molecular basis of microsatellites, see Schlötterer (1998).

2.2.8.1. *Statistical methods for the analysis of microsatellite data*

In order to make use of data obtained using microsatellite markers, it is vital that the appropriate statistical method is chosen for the analysis of the data. Many different estimators of genetic distance between populations exist, each taking a slightly different approach to the analysis. Some of these were originally developed for other molecular methodologies, whereas some were developed specifically for microsatellite data. It should be noted that a number of characteristics separate microsatellites from other markers. In particular their mode of mutation differs from that of, for example, allozymes, and a mutation model can be incorporated into the analysis of the data. For example, the Infinite Allele Model (IAM) (Kimura & Crow, 1964) assumes that a single mutational step is all that is necessary for an allele at any locus to mutate to any other allelic state. This mutation model is particularly of relevance to, and was originally devised for allozyme data where this theoretical parameter is correct. However, in the case of microsatellites, it is generally accepted that a major cause of mutation is polymerase slippage during replication (Hughes and Queller, 1993); this will usually result in either the addition or loss of a single repeat unit. A model that adapts itself to this aspect of microsatellites is the Stepwise Mutation Model (SMM) (Ohta & Kimura, 1973; Wehrhahn, 1975; Shriver *et* al., 1993; Valdes *et al.*, 1993); this model assumes that mutation acts by increasing or decreasing the electrophoretic mobility of an allele by one unit. The model also assumes that there is no upper limit to the number of repeats in an allele, a trait that may or may not be appropriate to microsatellites. Another intermediate model that is pertinent to microsatellite data is the Two Phase Model (TPM), this takes

aspects of both the IAM and SMM to produce an intermediate between them, whereby a mutation modifies the current allele size by one unit with probability p and by more than one unit with probability 1-P (Di Rienzo *et al.*, 1994). Evidence has been presented that demonstrates this model as a more appropriate mutation estimator for microsatellites (Di Rienzo *et al.*, 1994).

There are several kinds of estimators used to analyse population structure, but the most commonly used, at least until recently, have been those based on Wright's F-statistics (F_{ST}) (Wright, 1951). F-statistics are based on Wright's inbreeding coefficient F in a subdivided population. Roughly speaking, F is the probability that two randomly chosen alleles are identical by descent in the previous generation. In F-statistics, F is estimated for individuals within a subpopulation (I), between subpopulations (S) and for the total population (T).

F_{IS}, the inbreeding co-efficient, describes the divergence of observed heterozygosity from the expected heterozygosity within subpopulations assuming panmixia. F_{ST}, the fixation index, describes the reduction in heterozygosity within sub-populations relative to the total population as a result of selection or drift. F_{IT}, the overall inbreeding co-efficient, describes the reduction of heterozygosity within individuals relative to the total population as a result of non-random mating within subpopulations (F_{IS}) and population subdivision (F_{ST}). These terms are related (Hartl, 1988):

$$(1-F_{IT}) = (1-F_{ST}) + (1-F_{IS})$$

In other words, the total level of inbreeding equals the inbreeding caused by population subdivision plus the contribution to inbreeding by non-random mating within populations. When looking at spatial population structure, F_{ST} is the most important aspect, and can be estimated by G_{ST} using the following equation:

$$F_{ST} = G_{ST} = (H_T - H_S)/H_T$$

Where H_T is the mean heterozygosity of the metapopulation and H_S is the mean heterozygosity among individuals within subpopulations.

F_{ST} values range from zero to one. Zero indicates that there is no genetic differentiation between subpopulations, whereas 1 indicates that the subpopulations do not share any alleles. However, even at very low F_{ST} values genetic differentiation can still be significant (Wright, 1951). This is particularly pertinent with microsatellites, where the IAM does not account fully for the mutational effects, so that F_{ST} will tend to under-estimate the true level of genetic differentiation (Slatkin, 1995).

F_{ST} and other established methods of genetic differentiation, e.g. G_{ST} (Nei, 1973), θ (Weir and Cockerham, 1984) and measures of genetic distance such as Nei's unbiased distance (Nei, 1978) are all based on the IAM of mutation. As discussed earlier, this model is somewhat inappropriate to the analysis of microsatellite data. Statistics based on models that take into account the mutation pattern of microsatellites have been proposed: R_{ST} (Slatkin, 1995), a measure of genetic differentiation analogous to F_{ST}, and distance measures such as $(\delta\mu)^2$ (Goldstein *et al.* 1997; Feldman, 1997), ASD (Goldstein *et al.*, 1997) and Shriver *et al*'s D_{SW}. These measures, particularly R_{ST}, have been suggested to give more accurate statistical calculations of population parameters (Slatkin, 1995). However, there does seem to be some disagreement over this; Gaggiotti *et al.* (1999) used both F_{ST} and R_{ST} in a study, and consistently found F_{ST}-based estimates fitted the data more accurately where sample sizes were small. Ruzzante (1998) also found F_{ST} fitted data better than R_{ST}, particularly where there were large differences in sample size. However, R_{ST} has been further refined by Goodman (1997) to standardize the data input taking into account various factors such as differences in sample size between populations and the magnitude of variance in allele size between loci.

Another group of distance measures do not assume any particular evolutionary model of mutation. For these a simple Euclidean distance is determined rather than calculations based on population heterozygosity values. These include D_C - Cavalli-Sforza & Edwards Chord distance (1967), D_A - Nei *et al*'s 1983 distance and D_{AS}, an allele sharing

distance developed by Stephens *et al* (1992). Other methods of calculating distance measures have also been developed. It should be noted that both Cavalli-Sforza and Edwards (1967) Chord distance and Nei *et al*'s (1983) D_A distance have been shown to be particularly efficient measures for reconstructing evolutionary relationships when heterozygosities are high, outperforming distances designed for the stepwise mutation model (Takezaki & Nei, 1996). Paetkau *et al.*, (1997) also tested several distance measures on data sets from three Arctic bear populations. The authors found that at finer scales, distance measures developed specifically to accommodate mutational processes performed relatively poorly when compared to conventional distance measures. This is suggested to be because of higher levels of variance associated with the microsatellite-specific distance measures. However, when used to analyse larger evolutionary scales, all distance measures performed poorly; this was attributed to constrained and uneven mutation of the microsatellite loci.

In recent years, a number of other statistical approaches have been devised for the analysis of microsatellite DNA data. Many of these have been used for some time for phylogenetic reconstruction and in the analysis of sequence data, and are now being applied to microsatellite data (Luikart & England, 1999). These include:

(i) Coalescent methods: a mathematical modeling approach for assessing the shape of trees and genealogical processes underlying patterns of shared ancestry (Rogers & Harpending, 1992; Luikart & England, 1999).

(ii) Maximum likelihood (ML): methods whereby the parameters are found that maximize the probability of obtaining the observed data under a certain model (Wilson & Balding, 1998).

(iii) Markov chain Monte Carlo (MCMC) algorithms: A computer intensive stochastic simulation method for solving the mathematical integration necessary to calculate an allele distribution (Ayres and Balding, 1998).

(iv) Bayesian methods: Another modelling method to devise a probability distribution for the parameter of interest (e.g. effective population size) (Stefanini and Feldman, 2000).

In addition to this, a large number of computer programmes are available that can detect changes in the effective population size (N_e). Detecting reductions in N_e is critical in conservation biology where population declines can increase the risk of extinction. Genetic tests are now available that can identify a recently 'bottlenecked' (or declining) population, even when no information is available on the current or historical population size (Luikart & Cornuet, 1998; Luikart *et al.*, 1999; Spencer *et al.*, 2000). Methods are also available for detecting population expansion by modelling of microsatellite data, in particular looking for imbalances between allele size variance and heterozygosity (Kimmel *et al.*, 1998).

The analysis of microsatellite data is a relatively new and rapidly expanding field. A bewildering array of analysis techniques is available, and often they can be applied to the same data and arrive at subtly different conclusions. This is clearly an undesirable situation, for it hampers the comparison of studies where different analytical methods were used. A consensus is urgently needed as to which techniques are most appropriate for particular situations.

2.2.9. DNA sequence allelic variation

There are a variety of basic techniques that can determine allelic variation between individuals. The simplest of these involve differentiating heterozygotes from homozygotes, with no further characterisation of particular alleles. An example is heteroduplex analysis, a method that can identify whether an individual has one or two types of DNA at any particular locus (*i.e.* whether it is homozygous or heterozygous). This is achieved by PCR amplification of a locus, separating the strands by heating and cooling the amplified DNA, then running the results onto a gel. If the locus is heterozygous, there will be two different amplified molecules, which on denaturing become four strands of DNA, each with a unique sequence. On re-annealing, these may form heteroduplex molecules, which contain one or more base-pair mis-matches. When analysed on a gel, these will move at a different speed to the homoduplex molecules, thus

two bands will appear on the gel as opposed to the single band that represents homozygotes. This method relies on a heteroduplex molecule with base pair mis-matches having an altered charge to mass ratio, which may not always be the case.

Another method that can be applied is called single strand conformational polymorphisms (SSCP) (Lessa & Applebaum, 1993). The technique can resolve differences as minute as a single base pair change in DNA fragments up to 300bp long (Orita *et al.*, 1989). It relies on two principles: (i) The molecular arrangement of single stranded DNA results from its nucleotide sequence, and (ii) even small changes in configuration can affect mobility across a gel. The DNA region to be investigated must be isolated either by PCR amplification or restriction digestion, then heated and cooled to separate the two strands, electrophoresed and stained with ethidium bromide. If the individual is homozygous, each complimentary strand will move at slightly different speeds determined by its nucleotide composition, giving two distinct bands per allele. If the individual is heterozygous, there will be four separate bands on the gel. The probability of detecting a single base pair mutation in 300bp is 99% (Hayashi, 1991). In comparison, RFLP analysis can only detect base pair changes if they occur within a restriction site.

A third method that is also capable of separating homoduplex molecules is based on their physical properties and resulting behaviour during electrophoresis on gels containing linear gradients of denaturants (e.g. heat or urea). This technique is known as denaturing gradient gel electrophoresis (DGGE), or where temperature is the gradient: temperature gradient gel electrophoresis (TGGE) (Lessa & Applebaum, 1993; Shchyolkina *et al.*, 2000). The principle of this procedure is that as the DNA moves across the gel gradient, it will reach a point where it will denature. Once this has occurred, any further movement across the gel is severely impeded. The base pair composition of DNA will drastically affect its ability to resist denaturation, for example a mutation from G/C to A/T will increase the resistance. Different mutations will behave differently. Therefore, a locus that is heterozygous will produce two distinct bands because of the sequence variations, whereas a homozygote will produce a single band.

All of the methods described are capable of identifying small changes in a specific region of DNA that does not exceed 500 base pairs in length. The techniques are useful in investigations where specific allelic variation is of importance, but although the techniques are capable of identifying the presence of variation, further characterisation of the region is not possible. Sequencing of the region would be necessary for further investigation, for example, to identify the nature of the variation between the alleles, or to quantify the number of alleles per locus. Therefore its applications to some studies may be limited, and other techniques not involving DNA sequencing, such as microsatellites, could be more informative.

2.2.10. Nuclear Introns (EPIC-PCR)

Introns are non-coding regions within genes. Because these regions are not under the same selective constraints as exons, it is likely that they will have a higher mutation rate. PCR primers can be designed that will anneal to highly conserved nuclear genes, known as exon-primed intron-crossing (EPIC) primers. These provide homologous loci for examining a range of different organisms (Palumbi, 1995). Once a site to be amplified has been located, the introns must be sequenced to obtain information. Once this variation is understood, it is possible to use restriction enzymes to analyse further variation (Palumbi, 1995; Roderick, 1996). It seems likely that this technique will be useful for examining geographic structure within insect species as well as diversity among closely related species (Palumbi, 1995: Roderick, 1996). However, intron sequences are probably too variable to be used to reconstruct relationships between distantly related species. The technique is initially time consuming and costly, as introns must be found and then sequenced. Some introns may not be variable enough and may vary in their location between species.

2.3. SUMMARY

The methods outlined above give an introduction to some of the various techniques that are currently available and are of use in population studies. The techniques covered are not an exhaustive list of the techniques available, rather an overview of the commonly used approaches.

Each technique has advantages and disadvantages and these will affect the decision of when and how to apply it. For instance RFLP or RAPD analysis would not be the best techniques to use for an investigation involving degraded DNA, whereas microsatellites would be better because of their small size. A crucial factor in the choice of technique is the aim of the study, especially as many of the methods are costly. For example, it would be a waste of time and expense to apply a costly technique such as microsatellite analysis when allozyme electrophoresis is likely to be informative enough. An investigation into protein polymorphisms is often a sensible initial study for some projects; if allozymes are unable to resolve enough variation, the appropriate DNA markers can then be used. Not all markers are able to reveal heterozygosity among individuals. Allozymes, single locus minisatellites, and microsatellites are the only co-dominant markers, where estimations can be made of how closely they comply with Hardy-Weinberg expectations, or used in estimations of parameters such as F_{ST} (Wright, 1951; Slatkin, 1995). All other nuclear markers are dominant, i.e. the basis for the presence or absence of DNA fragments is unknown, and therefore cannot be assumed to have an identical cause for all individuals. Markers with a very high mutation rate may not be suitable for historical investigations into genetic structure. For instance, minisatellites and microsatellites both have extremely high mutation rates, thus are better for investigating more recent population variation. DNA sequencing may be the most powerful technique for detecting evolutionary change.

An enormous amount of use has already been made of the techniques available, and a great deal of information uncovered in the field of entomology. Insects are also of special importance as they are convenient for laboratory based studies to imitate

situations in the wild. The number of markers available leaves the way open for research into many diverse and interesting areas of population biology.

3.0 THE FIELD COLLECTION OF *P. BELLARGUS*

3.1. INTRODUCTION

The adonis blue (*Polyommatus bellargus*) is a local species in the UK. It is listed as a priority species in "Biodiversity: The UK Steering Group Report" (Department of the Environment, 1995). The severe decline that the butterfly underwent from 1976 through 1981 caused the number of populations in the UK to drop by over 90% (Emmet & Heath, 1990; Asher *et al.*, 2001) and although the species is making a strong recovery from this, with many former sites occupied by the species having been re-colonised, the butterfly still appears to be vulnerable at the edges of its range (Bourn & Warren, 1998). The sensitivity of this species demands that great care is taken during fieldwork.

In order to found the following generation, a male butterfly must fertilise the eggs produced by the female. Compared with the input from the female, who has to produce and carry the egg-load, this demands a relatively low amount of time and effort from the male, leaving him free to mate with a number of females. However, after fertilization, the female then has to carry the egg-load until a suitable oviposition site has been located. Because each female butterfly produces and carries the egg load, she has a much higher intrinsic value than the male, and thus the removal of one or more female butterflies from a population before oviposition, may have a significant impact on the following generation. However, removal of the male butterflies is unlikely to have such far-reaching consequences, especially if the males are removed after they have mated at least once or if the population is large enough to sustain such sampling without subsequent loss of genetic diversity. It was decided, after discussion with David Sheppard of English Nature, that large populations (total population >200) could withstand the premature removal of up to 30 males and as a rule of thumb, no more than 10 – 15% of any population should be sampled during any single generation. In some cases where the populations were extremely large, this meant that more than 30 males could be removed. The decision over which populations to include and the samples sizes required was therefore controlled by two factors: firstly; the level of sampling that each population

could reasonably sustain; and secondly, the minimum number of specimens that would result in a sufficiently unbiased and powerful statistical analysis – a suggested minimum of 40 individuals (Ruzzante, 1998). If sampling from consecutive generations can be combined (a total of four were available), it was theoretically possible to remove a minimum of 40 butterflies from all but the most impoverished of populations.

The removal of only male butterflies will mean that the samples obtained are not a truly accurate representation of the wild populations, this bias must be considered when deciding which genetic markers to use for the project. Microsatellite loci are unlikely to have any inherent sex bias, because they are generally autosomally inherited in a Mendelian fashion, with no dis-equilibrium between the sexes (Jarne & Lagoda, 1996). The use of microsatellite markers means that in terms of the population genetics, the samples collected can be of either sex provided the sample size is large enough to represent the population and enough markers are used (a minimum of five or six (Ruzzante, 1998)). The other marker that will be used is a region of the mitochondrial genome. Mitochondrial DNA (mtDNA) shows a different inheritence pattern from microsatellites, because it is maternally inherited. This means that the mitochondrial genome carried by any individual, whether male or female, is always inherited from the female parent, with no mixing or recombination of the parental genes occurring. This maternal bias should not cause undue problems because the genome present in both male and female siblings will be identical, therefore collection of male only butterflies will always represent the diversity present in both sexes. The results of all five microsatellite markers and the mtDNA region will be evaluated so that any disequilibrium will clearly show up.

The insects were sampled during the flight period because of the difficulties that would be involved in trying to locate larval specimens. It is unlikely that by attempting to collect individuals during the larval stage, large enough sample sizes could be reached. Other studies, such as that published by Saccheri *et al.* (1998), have used larval communities as a source of study specimens, but these investigations have tended to use species where the larvae are gregarious, and feed in conspicuous groups of 50 – 250.

Contrastingly, the ecology of *P. bellargus* not only means that the larvae are not gregarious, but that the mutualism that occurs with ants means the larvae are often hidden away in ant nests (Thomas *et al.* 1983a).

3.2. POPULATION HISTORIES

A great deal is known about the history of many butterfly species in Britain, this is as a result of organisations such as the Butterfly Conservation Society, who annually compile data received from a network of keen amateur lepidopterists across the country. Other conservation driven bodies such as English Nature, Local Authorities and the National Trust also carefully monitor the species present on their nature reserves. This huge resource of data provides an enormous amount of information across a broad time scale, particularly in threatened species such as *P. bellargus*.

3.2.1. Berkshire

The status of *P. bellargus* in Berkshire hangs very much in the balance. The county represents the most northerly range of the butterfly in Britain, and consequently the species only has a very weak hold in the area. The larger of the two small populations present, at Goring Gap, (Asher, 1994) is the most northerly population in the UK. This would have been an ideal site for sampling, because of its positioning at the edge of the species range, but the numbers supported at the site were considered too small to sustain the level of sampling required for the study.

3.2.2. Dorset

Dorset is generally considered to be the main stay of *P. bellargus* in Britain, with records of at least 80 populations (though this is likely to be an underestimate of the actual figure – many small populations are unrecorded), the majority of which are in the southern part

of the county (Bourn & Warren, 1998; Bourn *et al.*, 1999; Pearman *et al.* 1998; Stewart *et al.*, 2000). The habitat falls into three distinct geographical areas: (i) The coastline including the Isle of Portland and the east Purbeck ridge, (ii) The Dorset downs, and (iii) Cranbourne Chase (Buxton & Connoly, 1973). These areas are not sharply separated, with the underlying chalk areas forming a rough horseshoe shape, running from the north-east to the south-west, then back to the south-east of the county. Here the calcareous rock gives rise to chalk grasslands that are rich and diverse, containing numerous butterfly food-plants, including *H. comosa*. This grassland is semi-natural, having been cleared in Neolithic times and used ever since for agricultural purposes, particularly grazing. Most records of *P. bellargus* are concentrated along the southern calcareous ridges where the butterfly is abundant, forming the only area within Britain where many neighbouring populations become contiguous, forming what could almost be classed as a single continuous population.

The coastal area of the county is one of the only areas in Britain where many of the populations did not undergo a severe decline, although losses were still estimated at 70% (Asher *et al.*, 2001), with many having consistently remained relatively large throughout the late 1970's to the present day. A number of populations were sampled from this area (see table 3.1); this is likely to give a good indication of the level of heterozygosity and allelic diversity present in large, stable populations. Other sites from Dorset were chosen for exactly the reverse reasons; the Isle of Portland was sampled because of its geographical isolation as a fringe population, and because most of the populations underwent a severe decline in the late 1970's (J. Thomas, personal communication). Maiden Castle, another Dorset site, was chosen because the butterfly has only recently recolonised the site (J. Thomas, personal communication).

3.2.3. Gloucestershire

All populations of *P. bellargus* are extinct within this county. Two attempts were made to re-introduce the butterfly during 1990, but by 1995 both these populations had become

extinct. In 1995/6 two more attempts were made to introduce the species, but again these failed (Barker, 1990 & 1991; Meredith, 1992; Joy, 1997; N. Bourn, personal communication).

3.2.4. Hampshire

Hampshire remains a very poor county for *P. bellargus*. There is very little suitable habitat for the butterfly, with the only recorded site adjacent to the Dorset border. This population is currently very healthy in size and amply able to undergo the levels of sampling required. It is also an ideal population to include in the study because it underwent a severe decline in the late 1970's (J. Thomas, personal communication).

3.2.5. Isle of Wight

The Isle of Wight provides habitat for some of the most southerly populations of *P. bellargus* in Britain (Warren, 1987). An extensive ridge of chalk downland spans the island from east to west, providing numerous south-facing escarpments suitable for the butterfly. The mild climate of the area also encourages the species to flourish (C. Pope, personal communication) and one of the largest populations in Britain is located on the island at Brooke and Compton Down (T. Tutton, personal communication). Other sites at Ventnor and Mottistone support medium to large sized populations that should be able to support sampling. The populations from the Isle of Wight are all widely separated from the mainland distribution of the species, not only by unsuitable habitat, but also via an expanse of sea. A sedentary species such as *P. bellargus* is unlikely to attempt such a journey. This geographical isolation will be an important for the study of spatial variation.

3.2.6. Kent

Kent supports all of the most easterly populations of *P. bellargus* in the UK. There are a number of populations across the county, though most of these would be too small to sustain removal of individuals (S. Davis, personal communication). However, there are a few very large populations, one of which is near Folkestone, at Cheriton Hill (Philip, 1993). This population has rapidly increased since 1993, when the population size was negligible, to become one of the largest in the UK (N. Johannsen, personal communication). Another population that was sampled for the project was Wye and Crundale National Nature Reserve, *P. bellargus* has rapidly re-colonised this location over the last few years. Prior to this it had been locally extinct for over 40 years (S. Davis, personal communication). A population that has been suggested to be extremely large and stable was also sampled from the area, at Lydden and Tempel Ewell Downs SSSI. However, in 1999 this population was particularly impoverished of *P. bellargus*, whereas in 2000 there was an abundance of specimens. This may indicate the occurrence of a decline during 1999.

3.2.7. Somerset

The current status of *P. bellargus* in Somerset is very bleak. An unpublished report from 1973 (Buxton & Connolly) suggested the presence of three widely scattered sites in Somerset, all of which are on small hill ranges. However, none of these sightings have been substantiated, and there is no other evidence to suggest the persistence of butterfly at these locations. The last recorded populations were in the late 1980's when the butterfly was re-introduced to two sites. These populations persisted for a few years, but by 1994 were determined to be extinct (N. Bourn, personal communication; Bourn & Warren, 1998). Therefore, it is unlikely that *P. bellargus* persists in the county.

3.2.8. Surrey

Most of Surrey is unsuitable for *P. bellargus*, except where the western-most edge of the North Downs enters Surrey, here there are a few pockets of suitable habitat. There are six known sites where the butterfly has been recorded, although only one of these has a population large enough to withstand sampling. This population, at Ranmore Common (SSSI), is both large and extremely stable (having remained constant through the decline of the late 1970's (J. Thomas, personal communication)) and alongside the other five in Surrey, it is spatially isolated from the rest of the UK range.

3.2.9. East and West Sussex

A dominant feature of both East and West Sussex is the South Downs. This undulating calcareous grassland provides some excellent habitat for *P. bellargus*. It should be noted that although the habitat requirements concerning vegetation are commonly met, only 20% of the South Downs have the southerly aspect necessary for the survival of the butterfly. Nevertheless, the species is far from uncommon across both counties, and it has been suggested that there may be as many as 40 populations present (Gay & Gay, 1996). The information concerning the histories of these populations is less complete than for many other regions of the country, with detailed data only being available as far back as 1989 (J. Gay, personal communication). Populations were therefore chosen based more on their geographical location rather than demographic history.

The eastern most population in Sussex is beside Eastbourne, at the easterly end of the South Downs. This population is large enough to sustain sampling, and is the only suitable habitat towards the south east of England before Folkestone, in Kent, resulting in over 75km of unsuitable habitat separating the sites. The northern-most range of the butterfly in the counties is on the northern limits of the South Downs. This region is represented by samples removed from an extremely large population at Devil's Dyke, north of Brighton. Towards the west of the region, the butterfly was sampled from Mill

Hill, a large population just outside Shoreham. A population has been recorded at this site since late Victorian times, and has undergone declines not only because of the drought of 1976, but also due to the removal of butterflies by over-zealous Victorian butterfly collectors.

3.2.10. Wiltshire

Besides Dorset, the most densely populated region of the UK is Wiltshire, where up to 90 populations of varying size have been recorded (Fuller, 1995). This implies a rather rapid colonisation of Wiltshire; because the first published record of the butterfly in the county was not until 1883, even though reasonably extensive species lists existed before this date (Fuller, 1995). In 1982, when a Butterfly Monitoring Scheme was set up, the only recorded population in Wiltshire was Bratton Castle. This population is an ideal choice for sampling because of the historical data available. Declines have been recorded at the site during 1986, where records suggest a bottleneck of as few as 12 individuals, and again, though less severe, during 1993. Unfortunately, only a few sites were monitored this closely, and none of these were large enough to withstand sampling. Other sites that were sampled from Wiltshire were: a newly created habitat along the A303 road cutting, Prescombe Down Nature Reserve; Cley Hill - an isolated site that has been colonised post 1981; and Cotley Hill - a large and stable population. These populations should give a good indication of the status of the butterfly in Wiltshire.

3.2.11. France

The situation in France for *P. bellargus* is somewhat different from that in the UK. Here the species is consistently one of the more common species (S. Bryant, personal communication). The butterfly is not limited to closely grazed downland, and is commonly found among taller mean sward length grasses (Thomas, 1983a). The eight specimens were collected by Simon Bryant from a location in Southern France (5°22'E 44°45'N).

3.3. FIELD COLLECTION

3.3.1. The 1998 and 1999 Field Seasons

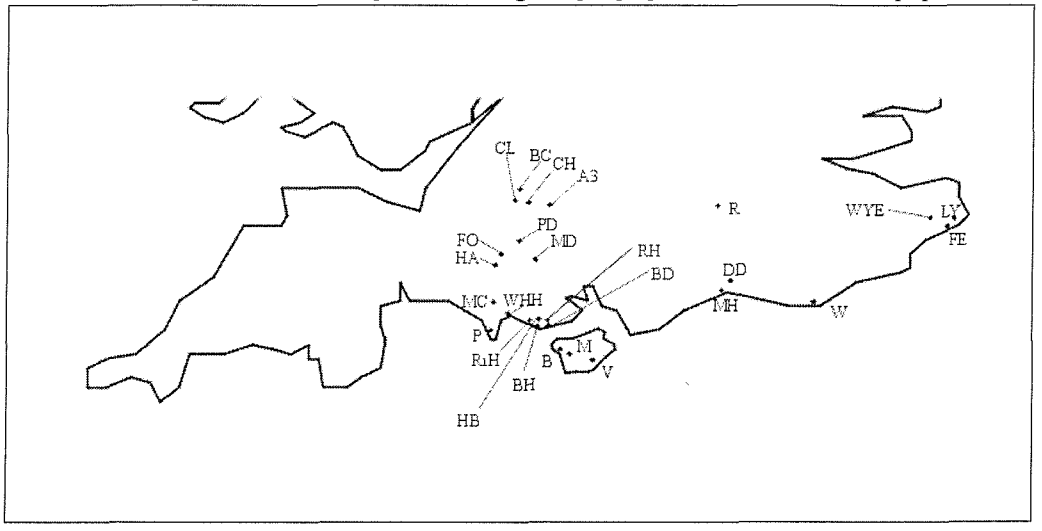
Table 3.1. A summary of the *P. bellargus* specimens collected from across the UK range

COUNTY	LOCATION		NUMBER COLLECTED		HISTORY*	LANDOWNER
			1998	1999		
Berkshire	-	-	-		-	-
Dorset	BH	Blackers Hole	30	21	S	National Trust (NT)
	BD	Ballard Down	50	0	S	NT
	FO	Fontmell Down	0	10	B	EN
	HB	Hill Bottom	16	25	U	Private
	HH	Hod Hill	1	0	B	NT
	MC	Maiden Castle	51	9	R	English Heritage
	HaH	Hambledon Hill	42	0	B	English Nature (EN)
	P	Portland	25	0	U	Local Council
	RiH	Ridgeway Hill	18	0	U	Private
	RH	Rollington Hill	30	22	U	Private
	WHH	White Horse Hill	36	21	S	Private
Hampshire	MD	Martin Down	17	26	B	EN
Isle of Wight	M	Mottistone	29	22	U	NT
	V	Ventnor	43	0	S	NT
	B	Brooke and Compton Down	0	42	S	NT
Kent	FE	Folkestone (Cheriton)	32	42	B	White Cliffs Project
	LY	Lydden	9	40	B	Kent Wildlife Trust
	WYE	Wye	0	36	R	Kent Wildlife Trust
Surrey	R	Ranmore	29	27	S	NT
Sussex	DD	Devil's Dyke	39	38	U	NT
	MH	Mill Hill	28	13	B	Local
	W	Willingdon	5	42	U	Council
Wiltshire	A3	A303 cutting.	7	27	U	Council
	BC	Bratton Castle	16	0	B	Private
	CH	Cotley Hill	27	21	S	Private
	CL	Cley Hill	0	52	R	NT
	Mi	Middleton Down	3	0	U	Wilts. Wildlife Trust
	PD	Prescombe Down	31	25	S	EN
France	FR	5°22'E 44°45'N	0	8	S	Unknown

* The letter denotes population history: U = Unknown, S = Stable, B = Bottlenecked, R = re-colonised

A summary of all the fieldwork carried out is shown above (table 3.1). This includes brief details of the known population histories. Most populations have a representative sample size of at least 40 individuals, though in some cases this number is achieved only by combining specimens collected from different generations. The feasibility of combining populations will be investigated by statistical analysis of the data for each year to determine whether any significant changes in allele frequency have occurred between generations. If no significant difference is found, they will be combined for the analysis of geographically separated populations.

Plate 3.1. A map of southern England showing the geographic location of all 26 populations sampled



3.4. DISCUSSION

To improve the statistical power and reduce the sampling variance of the genetic analysis, a minimum sample size of around 50 individuals is required (Ruzzante, 1998). In the majority of cases, this target has been achieved, provided specimens from sequential generations can be combined.

The sensitivity of *P. bellargus* to environmental and meteorological change means that the populations often show large annual fluctuations in numbers. Therefore, when assessing population size from recently recorded data, it is unavoidable that the actual population sizes vary from those expected. This means that not only will some

populations be larger; some will be much smaller than expected. In a few cases this has resulted in smaller sample sizes than those targeted for the study, with some populations being abandoned from the project. Sub-optimal sample sizes may still be of use, provided that the total number of butterflies represents a notable proportion of the population present. In a few cases, for example the single butterfly collected from Hod Hill, the sample size is so small that it is unlikely to be possible to make any inferences about the genetic structure of the population.

It is important when analysing the data, in particular when attempting to elucidate relationships between genetic diversity and population dynamics, to acknowledge that the majority of the historical data was compiled and collected by amateur entomologists. Much of the data concerning population dynamics being anecdotal rather than scientific, and thus inaccuracies may have occurred.

4.0 EVIDENCE FOR A RECENT COLONISATION OF THE UK

4.1. INTRODUCTION

Alongside the five microsatellites developed for the study, a second DNA marker was isolated. This was a polymorphic region of the mitochondrial DNA (mtDNA). As already discussed in chapter two, mtDNA is a separate entity from the nuclear genomic DNA and is only found within the mitochondria, the organelles of eukaryotic cells responsible for respiration, each mitochondrion carrying its own DNA molecule. In sexually reproducing organisms, the nuclear DNA of an individual is a combination of the genes from both of its parents, whereas in mitochondria, the DNA is only passed on maternally via the egg, therefore each individual inherits its mtDNA from an uninterrupted maternal lineage. This also means that the mtDNA does not undergo meiosis; therefore the sequence variations accumulated are purely as a result of mutation. This makes mtDNA an ideal marker for the elucidation of colonisation events and historical demographic changes in populations.

This chapter details the characterisation of the *P. bellargus* mitochondrial control region, and the application of this marker to expound details of the butterfly's colonisation of Britain. Historical population data for the species suggests that the butterfly was either extremely rare or not present in the UK prior to 1775, when the first confirmed record of the species was made. The collecting of butterflies as a hobby began during the last quarter of the 17th century, and one of its chief proponents was James Pettiver, "the father of British entomology" (Emmet & Heath, 1990). He named and described the majority of the British butterfly species (Pettiver, 1717). Notably he made no description that fits the adonis blue (*P. bellargus*) although this species is a conspicuous butterfly and most of his collecting trips were in the south east, where the species currently occurs. Despite a period of intense entomological activity, no description of this species appears in any work on the British fauna (Ray, 1710; Pettiver, 1717; Albin, 1720; Wilkes, 1741-1742; Wilkes, 1747-1749; Harris, 1775) until 1775 (Harris, 1775). Emmet and Heath (1990) conclude that *P. bellargus* "must have been an extremely rare species".

In the majority of taxa, the mitochondrial genome consists of 2 ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes, 13 protein-coding genes, and an origin of replication (Wolstenholme, 1992). Although many of these mtDNA regions show high sequence variability there are also particularly conserved genes, for example the two ribosomal RNA genes (12SrRNA and 16SrRNA), which are highly conserved across all genomes (Van der Peer *et al.*, 1993). This is probably because the function of the ribosomal genes is to produce a single strand of RNA that can be folded up into a functionally important three-dimensional structure for protein assembly (translation), thus there are strong constraints on sequence variability (Simon *et al.*, 1994). The other group of genes that show considerable conservation of their sequences are transfer RNAs (tRNA); again these are involved in translation, by carrying each amino acid to the appropriate codon on the messenger RNA (mRNA). Once more, it is the functional importance of the molecule that causes the constraints on sequence variation (Wolstenholme & Clary, 1985). However, the mutation rate does vary between species and different tRNA genes (Pashley & Ke, 1992; Ojala *et al.*, 1992). Many of the protein-coding regions contain more variation than either of the two classes of RNA genes, because of the triplet code (codon) used for protein assembly. This degenerate nature of the coding system allows nucleotide variation in the third base pair position of the codon, where these less constrained sites have been shown to evolve at a higher rate then their two counterparts (Nei, 1987).

Mitochondrial DNA has very few spacer or noncoding regions. The most consistently present spacer region is the control region, known as the "AT-Rich control region" in invertebrates (Fauron & Wolstenholme, 1980), or the "D-loop" in vertebrates (Upholt & David, 1977). This is one of the only mtDNA regions that consist of non-coding DNA and a frequent feature is the presence of short tandemly repeated sequences that may be involved in replication (Wilkinson *et al.,* 1997; Zhang & Hewitt, 1997). The region has been designated as the putative origin of replication and transcription for the circular molecule of mtDNA, by virtue of evidence for this in a small number of species and its lack of coding structure (Simon *et al.* 1994; Lunt *et al.*, 1998). Sequence variation can be high in this region, probably as a result of low selective pressure on the non-coding DNA,

inefficient DNA repair systems and the AT-rich nature of the control region, leading to slippage during replication. It is not uncommon to find variable number tandem repeat regions (mtVNTRs) within or adjacent to this control region; in fact over 150 species show evidence for the presence of these mtVNTRs (Zhang & Hewitt, 1997; Lunt *et al.*, 1998).

This assortment of mutation rates among the different regions of the genome lends itself particularly well to PCR amplification, this is because PCR primers can be designed within the highly conserved regions surrounding more variable genes, which can be used to amplify the more variable areas between them. For example, the RNA genes are distributed relatively evenly around the genome amidst the protein coding genes, forming ideal sites for primer design. The result of this is that primers can be developed that are not only conserved within a species, but that are ubiquitous across a broad range of species and taxa. These are known as "universal primers". Many of these universal primers have been developed for mtDNA (for an extensive review of these, see Simon *et al*, 1994) and a set of primers is commercially available (J. Hobbs, Nucleic Acid – Protein service: University of British Columbia).

For the study of population structure and the analysis of recent demographic events, it is important to choose a region of the genome that is likely to have a high mutation rate. An obvious choice for this would be the "A+T-rich control" region, where mutation is rapid and least constrained. Within lepidopteran species, this region has been shown to be the most rapidly evolving sequence (Taylor *et al.*, 1993), although there does seem to be some constraint on the length of the region (~350bp) (Taylor *et al.*, 1993). However, because of its AT-rich nature, care must be taken when analysing this region because there are only two nucleotides available for substitution, thus convergent evolution will be high and difficult to detect (Simon *et al.*, 1994).

A previous study on *Plebejus argus*, a butterfly species closely related *P. bellargus*, investigated the effects of genetic bottlenecks and habitat fragmentation on the species. The "A+T-rich control" region was successfully used to identify levels of variation

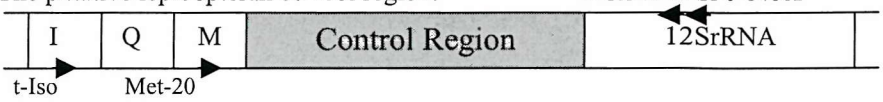
present between populations (Brookes *et al.* 1997). This study applied restriction fragment length polymorphism (RFLP) analysis to the control region rather than direct sequencing, finding 19 haplotypes (Brookes *et al.* 1997). It seems probable that by applying sequence analysis to *P. bellargus*, an even higher level of variation could be uncovered. Therefore the A+T-rich control region seemed to be the most appropriate region for this project and in order to fully elucidate the variation present, sequencing analysis was employed.

In most insects, the control region is flanked on one side by a group of three tRNA genes (methionine (M), glutamine (Q) and isoleucine (I), in that order) and on the other side by the 12SrRNA gene (see figure 4.1). These conserved RNA genes are ideal for primer design, and have been used to create universal primers for amplifying the A+T-rich control region in various insects, including a number of *Drosophila* and lepidopteran species (see figure 4.2) (met20: 5'-TGGGGTATGAACCCAGTAGC; t-Iso - 5'-ATTTACCCTATCAAGGTAA; 12Sr348: 5'-TAGGGTATCTAATCCTAGTT and Sr-J-14612: 5'- AGGGTATCTAATCCTAGTTT) (Clary and Wolstenholme, 1987; Taylor *et al.*, 1993; Simon *et al.*, 1994). However, in many species (including all vertebrates and some invertebrates) this group of three tRNA genes (M, Q, and I) has been transposed to a different site within the genome, so these primers (met20, t-Iso, Sr-J-14612 and 12Sr348) cannot be used to amplify the AT-rich control region or D-loop (Hoffman *et al.*, 1992). These genes have also undergone transposition within some lepidopteran species; here the M-tRNA is directly adjacent to the control region. The locations of the I-tRNA and Q-tRNA are unknown, but may have swapped with M-tRNA within the same cluster (see figure 4.2) (Taylor *et al.*, 1993). This movement is not exceptional, a number of other insects have undergone similar re-arrangements of other tRNA genes (Haucke & Gellissen, 1988).

Figure 4.1. The configuration of a generalised insect control region and flanking genes. Figure 4.2. The configuration of the lepidopteran control region and flanking genes, showing the putative locations for the "I tRNA" and "Q tRNA" genes. Primer binding sites are also shown, see text for appropriate sequences.

Figure 4.1The insect control region:

Figure 4.2. The putative lepidopteran control region:



It is also possible that the I-tRNA and Q-tRNA have transposed to a separate part of the mitochondrial genome, and are no longer flanking the M-tRNA and A+T-rich control region (figure 4.2). If this is the case, then amplification attempts with the primer "t-Iso" will either produce a significantly larger amplicon than expected (>720bp), or no PCR product at all (although failure of PCR amplification could also indicate a mutational event in the I-tRNA gene).

This chapter details the examination of which primer combination is most appropriate for PCR amplification and also the analysis of data obtained by sequencing the control region. Specimens of *P. bellargus* were obtained from throughout the UK range (the Isle of Portland, the Isle of Wight, Kent, South Downs, Sussex, Salisbury Plain, Dorset) plus eight butterflies from southern France were all sequenced. The French specimens originated from three populations, all in fairly close proximity to one another, located at the grid reference 5°22'E 44°45'N. In contrast to the UK, the situation for *P. bellargus* in southern France is rather different; here the butterfly is fairly ubiquitous, being one of the most common species present (S. Bryant, personal communication). This may provide an insight into the level of variation present in large, stable populations close to the centre of a species range, compared to marginal populations, whilst also revealing evidence pertaining to the colonisation of Britain by *P. bellargus*.

4.2. MATERIALS AND METHODS

4.2.1. DNA extraction

Salt extraction technique. (after Sunnucks & Hales, 1996; Aljanabi & Martinez, 1997): Heads from *P. bellargus* specimens were homogenised using UV-cross linked pellet mixers in individual eppendorf tubes in the presence of 300µl TEN (250 mM NaCl, 50 mM Tris HCl, 10 mM EDTA, pH 8.0)/2% SDS and 30µl of 1mg/ml proteinase-K. This was followed by incubation for 3 hours at 55°C. 100µl of 5M NaCl was added to each of the samples, which were then shaken on a vortex prior to centrifugation at 14,000rpm for 10 min. The supernatant was then carefully aspirated and pipetted into clean eppendorf tubes where the DNA was precipitated using 1ml of absolute ethanol at –20°C. After precipitation, the ethanol was decanted off and the DNA pellets washed twice in 1ml of 70% ethanol. After both ethanol washes, all the alcohol was removed and the samples air-dried and re-suspended in 20-40µl of TE.

4.2.2. PCR Screening of Various Universal Primers

The universal PCR primers (t-Iso, Met-20, 12Sr348 and Sr-J-14612) were synthesised (Genosys, UK) and all four possible combinations of forward and reverse primers were assessed to determine the most efficient set for PCR amplification of the A+T-rich control region (see table 4.1).

Table 4.1. The possible combinations of the four universal primers (a forward primer must be paired with a reverse primer).

PRIMER ORIENTATION	PRIMER	1	2	3	4
Forward	t-Iso		X		X
	Met20	X		X	
Reverse	12Sr348			X	X
	Sr-J-14612	X	X		

All 25µl PCR reactions contained 50-100ng of template DNA; 2U *Taq* DNA polymerase (ABgene, UK); 0.2µmoles of each primer; 20mM $(NH_4)SO_4$; 75mM Tris-HCl, pH 8.8; 0.01%(v/v) Tween® 20; 1.5mM $MgCl_2$; 0.25mM dNTPs (ABgene, UK). The reaction

was carried out under the following conditions: 1x 94°C, 4 min; 30x 94°C, 1min, 45°C, 1min, 72°C, 2.5min; 1x 72°C, 7 min. Once completed, the PCRs were assessed for successful amplification on a 1% agarose gel stained with ethidium bromide.

4.2.3. Sequencing the A+T-rich control region

Once the appropriate primer combination for amplification was identified, the next step was to sequence the control region. Initially only six *P. bellargus* individuals were sequenced to ensure that the PCR was correctly targeting the control region. Once this had been verified, the region was sequenced from a subset of specimens across the UK range and also from the group of individuals from France. Initially only six butterflies from each UK region and all eight from France were sequenced, with the intention of scaling this up after preliminary analysis of the results.

The sequencing was carried out using DNA that had been gel purified from a PCR amplification of the region. The PCR amplification was carried out as detailed in section 4.2.2, and visualised on an ethidium bromide stained agarose gel (see figure 4.3). The c. 720bp PCR product was purified by excision of the band from the agarose gel, then isolating the DNA using a Qiaquick gel purification kit (Qiagen, USA). In order to add enough of the purified PCR product to a 10µl sequencing reaction, the concentration of purified DNA was roughly estimated using an ethidium bromide stained agarose gel. The sequencing reaction was carried out via the manufacturers instructions using Big-Dye Terminators (PE-Applied Biosystems, USA) and run on 5% denaturing polyacrylamide gel by vertical electrophoresis at 20-60mA for 2 hours using a Perkin-Elmer ABI 377 automated sequencer. The region was sequenced using both Met20 and 12Sr348, so that both the forward and reverse sequences were obtained.

4.2.4. Statistical Analysis

Once the sequences had been obtained, they were subjected to alignments using the computer programme Clustal-X (Thompson *et al* 1997). This highlighted any sequence variation between the A+T-rich control regions of the individuals studied. The *P. bellargus* control region was compared with other species: *Jalmenus evagoras* (from Ebor, Australia) (Lepidoptera: superfamily Papillionoidae: family Lycaenidae; subfamily Theclinae: tribe Zeziini) (GenBank L16849) and *Strymon melinus* (from North America) from the same subfamily (Theclinae: tribe Eumaeini; classification follows Eliot, 1973) (GenBank L16850). Alignments were also carried out to identify any variation occurring among *P. bellargus* individuals.

Basic statistics (for example the number of haplotypes, transition:transversion ratio (TS:TV), nucleotide composition and mean number of pair-wise differences between haplotypes (Tajima 1983; Nei 1987)) were calculated using *ARLEQUIN* (Schneider *et al.*, 2000). The relationship between populations was also calculated in *ARLEQUIN*, using Tamura's (1992) genetic distance. This distance measure was considered most appropriate because of the high A+T content of the sequences, and also on the basis of the transition:transversion ratio, which was higher than the expected ratio of 1:2 (Tamura, 1992; Oyler-McCance *et al.*, 1999)

A Tree representing the relationship between the haplotypes was constructed in *Phylip 3.57c* (Felsenstein, 1993) using a maximum likelihood method, without the assumption of a molecular clock. Published control region sequences for *J. evagoras* and *S. melinus* were used in the analysis as outgroups. The data were bootstrapped in the subroutine *SEQBOOT*, with 100 iterations, and then 100 distance matrices were created from the bootstrapped data using the subroutine *DNADIST*. These matrices were used to create 100 Neighbour Joining trees, using the subroutine *NEIGHBOUR*, invoking option J to randomise the input order. Finally, a maximum likelihood consensus tree was created

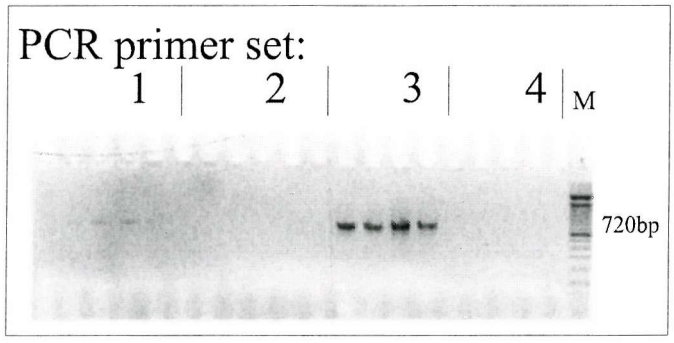
using the subroutine *CONSENSE*. A minimum spanning network between haplotypes was also created using *MINSPNET* (Excoffier, 1993).

4.3. RESULTS

4.3.1. Screening of Various Universal Primers

On screening the four primer groupings (see table 4.1), the combination of Met20 and 12Sr348 gave the clearest and most reproducible results. This combination has also been found to successfully amplify the A+T-rich control region in a selection of other lepidopteran species (Taylor *et al.*, 1993: Brookes *et al.*, 1997). When Sr-J-14612 was combined with Met20, the amplification was also successful, but with reduced reproducibility and lower yields of PCR product, whilst none of the PCR amplifications employing t-Iso were successful.

Figure 4.3. An inverted tone image of the PCR amplified mitochondrial A+T-rich control region using various universal primers (see table 4.1 for details of primer sets 1-4 (Simon *et al.* 1994, Taylor *et al.* 1994)). Lane M: 100bp DNA marker. The samples have been run out on a 1% agarose gel, and stained with Ethidium bromide.



4.3.2. Characterisation of the MtDNA AT-rich Control Region

After optimisation, the AT-rich control region PCR amplified very consistently. However, when attempts were made to sequence the amplicon, problems were encountered. The AT-rich nature of the sequence resulted in extremely short sequences,

in the range of 250-400bp. This was far shorter than would be expected using automated sequencing methods (*circa* >700bp). The only way to overcome this was to combine the forward and reverse sequences for each individual, thus piecing together the entire sequence.

Figure 4.4. The sequence for the *P. bellargus* mitochondrial genome, between the Met tRNA and 12SrRNA genes. The sequences were obtained from individual Ha2, from the population at Hambledon Hill (Dorset). Obtained by combining both forward and reverse sequences.

```
CTTTATTTAGCTTATTTTTAAAAAATAATTTTTTATTTTTATTATAAAAATTATTAAAAA
TGGGTTTAAGAATATAATTATTTTTACCGTTGATTGGGTTTTTCTTTTATTATTTACCGT
GCACCGTATATATACCATATATATATATATTAAATTTTTAATTAATTATTAATTTTAATA
ATTAATTAATAATTTTATTATATTAAAATAAATAATTTATTAATTTAAATTTAATAAATT
TAAATATTTTAATATTTAGTATTTAAATTAGAACATTATAGGCTTATCATGAAATATATT
TTTAAATTTAATATTAGAAAATAGAGAGAGAAAGATTATTAATTGTTTAATATTATTATT
AATTTTTACGTAAAAAAAAAAAAAAAAAAATTCTATTTAAAATAATTTGCATATAGAAAA
AATTTTTATAGTTTTGATATTTTATTTTAAATATAATTTACATATAAATTTTAGTGTTAA
ATTTTATTTATTTTACNCAATAAATAAAATTATTAAGTAGTAATTAACTTTAAATTATTT
AAGAAATTAAGATTTAGTAATTTTGAAATAAATGACTAATTTTGTGCCAGCAGTTGCGGT
TAACAAAAATTTAATTAAAATTTTTCCAGTATATAATAAGTAATAATTAATAATTAAATT
AAAAATTAAATTATTAGGTGAAATTTTAATTTAATAAAAATTTAAAAATTTGTATGAATT
AA
```

722bp
Each line is 60bp.

The sequences for the A+T-rich control region have been published for two other members of the lycaenid family (Lepidoptera) (Taylor *et al*, 1993), these are *Jalmenus evagoras* and *Strymon melinus*. In order to substantiate that the sequence obtained from *P. bellargus* (figure 4.4) represented the control region, a sequence alignment was carried out against the *J. evagoras* and *S. melinus* sequences (see figure 4.5).

The sequences obtained have an extremely high ratio of A+T to G+C and as such are likely to represent the A+T-rich control region. From the alignment (figure 4.5) of the primary sequences against other published sequences for lycaenid butterfly species (*J. evagoras* and *S. melinus*), the homology between them is unmistakable, particularly within the putative 12SrRNA gene, where the only variations are scattered point mutations. The homology between the control region sequences relative to *P. bellargus* amounts to 67% and 75% for *J. evagoras* and *S. melinus* respectively, whereas that for the 12SrRNA region amounts to 86% relative to *P. bellargus* (only the *J. evagoras* 12SrRNA region is available for comparison).

Figure 4.5. Sequences of the mtDNA of an individual of *P. bellargus* (P.b.), from Hambledon Hill (Dorset), between the Met tRNA and 12SrRNA genes, compared with similar sequences for *Strymon melinus* (S.m.) (GenBank L16850) and *Jalmenus evagoras* (J.e.) (GenBank L 16849). A colon (:) indicates identity with *J. evagoras*; and a dash (-) indicates a deduced indel. Indels larger than 12bp relative to the *Jalmenus* sequence are shown at the bottom of the figure. The Met tRNA, 12SrRNA and control region genes, are highlighted, plus also a short microsatellite region.

Within the *P. bellargus* sequences, there are three insertions of note, the most extensive of which is 59bp in length. An insertion is also present within *H. melinus* at the identical site relative to the control region; however, it is shorter in length (31bp) and has poor sequence homology (41%) with the *P. bellargus* insertion (see base of figure 4.5). This

insertion does appear to have notable levels of homology (56%) with an adjacent stretch of the *P. bellargus* control region, and also with the equivalent stretch of *J. evagoras* control region (50%). The other two *P. bellargus* insertions appear to be multiple additions of single nucleotide (adenine), perhaps caused by slippage during replication.

4.3.3. Analysis of the *P. bellargus* control region

After verification that the region being amplified was the control region, the screening for variation within and between UK and French individuals commenced. By sequencing the region in both forward and reverse orientations, the entire region was analysed for six UK and three French specimens and it was determined that, as expected, all of the variation occurred within the control region, rather than the 12SrRNA gene. For this reason, we decided to continue by only sequencing the control region using the Met20 primer.

After sequencing a total of 50 individuals from seven distinct geographic regions of the UK, it was clear that the British populations exhibited remarkably little variation. The only divergence between the sequences was by one or two indels of a (TA) repeat unit in the latter part of the short microsatellite ($(TA)_3C(AT)_7$) within the control region (See figure 4.6). The addition of a single repeat occurred in three individuals and the addition of two repeat units was present in a single individual all four were from Dorset. For all other regions of the UK, there appeared to be a single mtDNA haplotype (represented by 46 sequences) (shown in figure 4.6).

When the sequences from the eight individuals from southern France were aligned, the level of haplotypic variation was found to be much higher than within the UK (see figure 4.6). Not only was there significant divergence between the UK and France, but in contrast to the UK specimens, there also appeared to be much higher levels of variation present among the French individuals. In total, the eight French specimens were represented by six haplotypes.

Figure 4.6. Sequences of the Met20 amplified control region of six French and one UK *P. bellargus* individuals. A colon (:) indicates identity with the predominant UK haplotype; and a dash (-) indicates a deduced indel. The haplotypic variation within the UK is indicated at the bottom of the figure.

```
UK              CTTTATTTAGCTTATTTTTAAAAAATAATTTTTTATTTTTATTATAAAAATTATTAAAAATG
Fr1             :::::::::::::::::::::::::-:::::::::T::::::::::G::::::::::::::-:
Fr2             ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Fr4             :::::::::::::::::::::::::--:::::::::T::::-:::::G::::::::::::::
Fr6             :::::::::::::::::::::::::--:::::::::T:::--:::::G::::::::::::::
Fr3             :::::::::::::::::::::::::-:::::::::T:::::-:::::G::::::::::::::
Fr5             :::::::::::::::::::::::::-:::::::::T:::::-:::::G::::::::::::::

UK              G-TTTAAGAATATAATTATTTTTACCGTTGATTGGGTTTTTCTTTTATTATTTACCGTGCAC
Fr1             :G::::::::::::::::::::::-:::A::::::A:-:::::::C::::::::::::::::
Fr2             :-:::::::::::::::::::::::::A::::::::::::::::::::::::::::::::::
Fr4             :G::::::::::::::::::::::-:::A::::::A:-:::::::C::::::::::::::::
Fr6             :G::::::::::::::::::::::-:::A::::::A:-:::::::C::::::::::::::::
Fr3             :G::::::::::::::::::::::-:::A::::::A:-:::::::C::::::::::::::::
Fr5             :-::::::::::::::::::::::-:::A::::::A:-:::::::C::::::::::::::::

                                          ▼
UK              CGTAT-ATATACATATATATA--TATATTAAATTTTTAATTAATTATTAATTTTAATAATT
Fr1             :::::T:::::T::::::::::--C::::::::::::::::::::::::::::::::::::
Fr2             :::::-:::::::::::::::--::::::::::::::::::::::::::::::::::::
Fr4             :::::T:::::T::::::::::--C::::::::::::::::::::::::::::::::::::
Fr6             :::::T:::::T::::::::::TAC::::::::::::::::::::::::::::::::::::
Fr3             :::::T:::::T::::::::::--C::::::::::::::::::::::::::::::::::::
Fr5             :::::T:::::T::::::::::--C::::::::::::::::::::::::::::::::::::
```

Indels:-
▼ Insertion of either TA or TATA here (UK only).

The sequence alignment (figure 4.6) clearly demonstrates the divergence between the different haplotypes. The French individuals have a number of sequence variations that separate them from the UK haplotypes. The presence or absence of the TA repeat in the short VNTR does not appear to be specific to the geographic origin of the haplotype, it is present in both the French and UK butterflies.

Of the 193 base pairs analysed, eight substitutions and 15 indels characterised a total of nine haplotypes. Six of the substitutions are transitions, and two were transversions, giving a higher ratio than the 2:1 ratio expected (Tamura, 1992; Oyler-McCance *et al.*, 1999). As expected with the AT-rich control region, there was an extremely low G+C content of the sequence data (Clary & Wolstenholme, 1985) and although this varied slightly between individuals, the nucleotide ratios were on average found to be: A 36.45%; C 5.74%; G 6.63%; T 51.47%. Nucleotide diversity (average number of nucleotide differences per site between two sequences) was 0.021 (S.D. = 0.012, n = 58) overall, with the French population at 0.026 (S.D. 0.016, n = 8) and the UK values

ranging from 0.00 (n = 6) (All UK populations except Dorset, n = 43) to 0.01 (S.D. 0.008, n = 7) (Dorset), with the overall mean number of pairwise differences between all haplotypes being 3.252. The value among the French samples was 7.424, whilst the UK values ranged from 0.00 (all populations except Dorset) to 0.295 for Dorset. Gene diversity (the probability that two randomly chosen haplotypes are different) ranged from 0.153 (n = 50) for the UK, to 0.929 (n = 8) for France.

Figure 4.7. An unrooted maximum likelihood consensus tree of control region haplotypes, from UK and French populations of *P. bellargus*. Equivalent published sequences for *Jalmenus evagoras* and *Strymon melinus* have been used as outgroups. Tree created using the *DNADIST* programme in *PHYLIP 3.57c*. Figures in italics indicate bootstrap values after 100 replications.

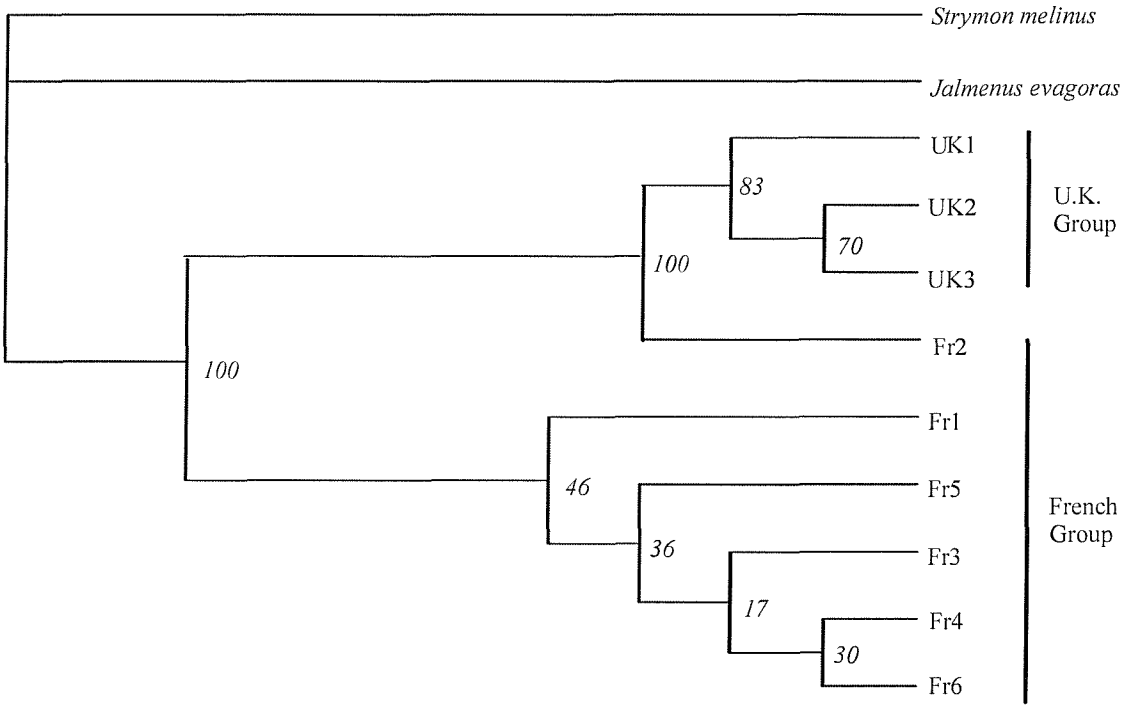
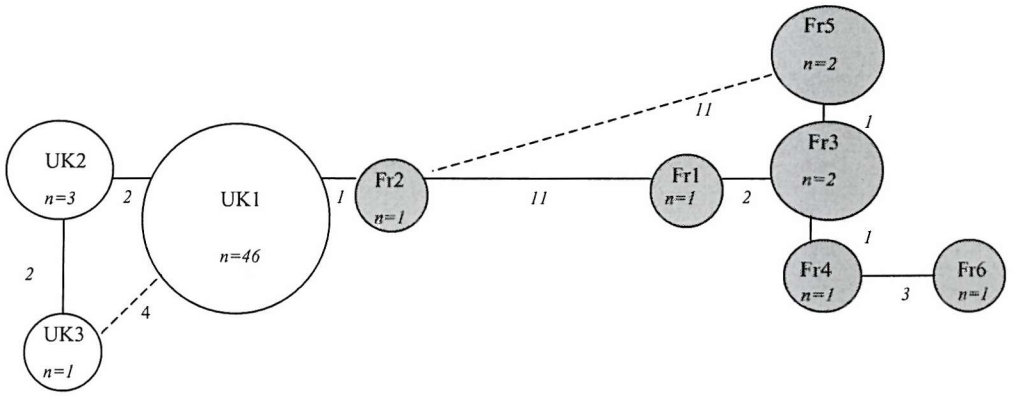
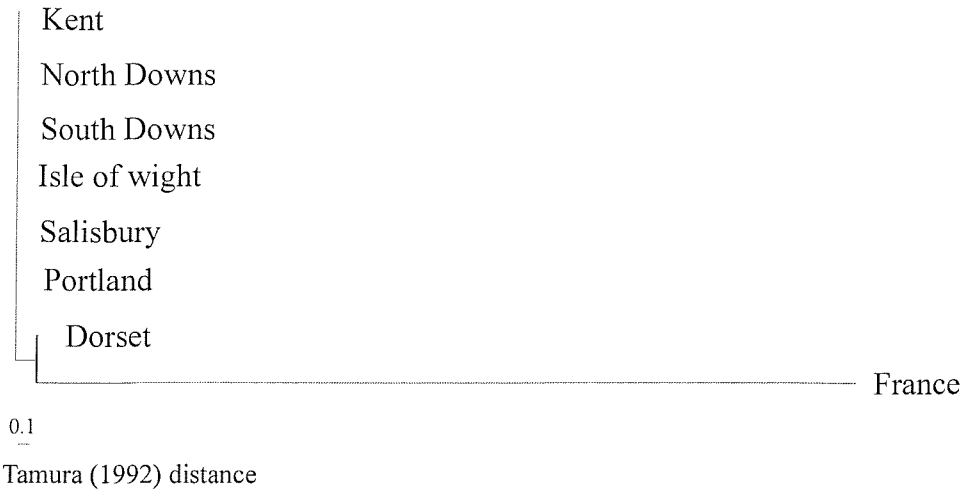


Figure 4.8. A minimum spanning network (Excoffier, 1993) showing the number of base changes between haplotypes. Each haplotypes is represented as a circle with its relationship to the haplotypes most closely related to it (defined by the number of base changes) presented as a line. A dotted line indicates an alternative relationship. The numbers of base pair differences between haplotypes is only indicated where values are >1. Shaded circles are French and non-shaded are UK haplotypes.



The maximum number of differences between any pair of haplotypes is 17 ("Fr6" compared with "UK3"), as shown in the minimum spanning network of haplotypes (figure 4.8). Within the UK there is only variability at a single site within the region, characterised by the addition of two or four nucleotides at a short VNTR, probably representing one or two insertions respectively. The French haplotypes show more variation, characterised by between one and sixteen nucleotide changes between them, although the majority appear to be from one to three mutational steps (see figure 4.8). Most have at least eleven nucleotide differences compared with the predominant UK haplotype, but a single French haplotype ("Fr2") is almost identical to the predominant UK haplotype ("UK1"), with only one nucleotide change separating them (figure 4.8). The maximum likelihood tree (figure 4.7) shows that all but one ("FR2") of the French haplotypes group away from the UK haplotypes (and "Fr2") very robustly, supported by 100% of the bootstraps.

Figure 4.9. Tamura's (1992) genetic distance relationship between the UK and French populations of *P. bellargus*, based on the sequence data for the control region. This tree is not rooted and no pattern of descent can be assumed.



Tamura's (1992) genetic distance is used to show differentiation between the geographic regions (figure 4.9). The unrooted tree shows six of the UK populations grouping together, since their control regions are genetically identical. The only exception to this is the population from Dorset, which is separated from the other UK populations by an extremely short genetic distance (0.148; Tamura's (1992) genetic distance). This results from fixation of haplotype "UK1" within all UK populations except Dorset, where there are two other haplotypes present – although these are only separated from "UK1" by one or two mutations. In contrast, the individuals from southern France are isolated from all of the UK groups by a large genetic distance (6.275; Tamura's (1992) genetic distance). This is as a result of large differences between most of the French haplotypes compared with those from the UK.

4.4. DISCUSSION

The sequences obtained from *P. bellargus* A+T-rich control and 12SrRNA regions showed distinct similarities to those produced for two other lycaenid butterflies (*Jalmenus evagoras* and *Strymon melinus*) (Taylor *et al.*, 1993). When the sequences were aligned, the segments representing the control regions for *J. evagoras* and *S.*

melinus showed 67% and 75% homology with the *P. bellargus* sequence, respectively, whereas that for the 12SrRNA region was much higher, at 86% (*S. melinus* sequences were unavailable). This increased similarity at the 12SrRNA region almost certainly reflects that fact that this ribosomal gene transcribes to a functionally important three-dimensional structure, resulting in increased conservation of the nucleotide content. However, these constraints are not pertinent to the control region, where the nucleotide sequence is not transcribed or translated to any kind of functional secondary molecule, but rather, it represents the origin of replication for the mitochondrial genome. It has been noted by Taylor *et al.* (1993), that the only part of the lepidopteran control region that appears to have any conservation of primary sequence is within, and adjacent to, the $(TA)_n$ VNTR, a short stretch of sequence that also has weak similarities to an equivalent part of the *Drosophila yakuba* control region (Taylor *et al.*, 1993). This area shows considerable sequence conservation relative to other parts of the control region, between *P. bellargus* and the other lycaenid species, and also to some extent among the *P. bellargus* specimens (see figures 4.5 and 4.6), although some sequence variation does occur here. This may infer the region has an increased structural or nucleotide sequence significance, relevant to replication of the mitochondrial genome.

Within the *P. bellargus* sequences, there are two notable insertions relative to the *J. evagoras* sequence, the first of these is a multiple addition of a single nucleotide (adenine); the second is the insertion of ~60bp adjacent to the VNTR region. Remarkably, an insertion is also present within *S. melinus*, at an identical site relative to the aligned control region sequence. However, the two insertions have no obvious sequence homology (41%), suggesting either extremely low constraints on mutation, or that there have been two separate insertion events at the same site, during the evolution of these two lycaenid species. The *P. bellargus* insertion does have sequence homology to ~60bp of adjacent control region sequence, possibly indicating that at some point this section of the control region has undergone duplication, resulting in an adjacent stretch of inserted sequence with low mutational constraints. However, it must be remembered that the A+T-rich nature of the control region, means that unrelated sequences would be expected to have homologies of around 50%, thus the homology of the insert is not much

higher than would be expected by chance. It is also worth noting that because of these insertions, the control region of *P. bellargus* is approximately 80bp longer than that of *J. evagoras*, this contradicts the suggestion by Taylor *et al.* (1993) that the lepidopteran control region is very conserved in length at ~350bp.

Very little indication was obtained regarding the putative locations of either the glutamine (Q) or isoleucine (I) tRNA genes. In accordance with other lycaenid control regions, the sequence data did not indicate their presence between the start of the methionine (M) tRNA and the control region, as found in *Drosophila yakuba* and *Drosophila virilis* (Clary and Wolstenholme, 1985). When attempts were made to amplify the control region using a primer based on the I-tRNA sequence (t-iso), all amplification attempts failed. This indicates that either the I-tRNA gene may have been transposed to a separate part of the genome or that mutations have occurred within the primer binding sequence for "t-iso".

The lack of haplotype diversity and polymorphism within the UK populations of *P. bellargus* is a dramatic departure from expectation based on the results of similar surveys in many other taxa. For example, mtDNA studies of vertebrates (e.g. Avise, 1986; Moritz *et al.*, 1987) and a number of invertebrates generally reflect a much higher degree of differentiation at the population level than those observed within *P. bellargus*. Bark weevils (*Pissodes* species) (Smith and Brown, 1990), the silver studded blue butterfly (*Plejebus argus*) (Brookes *et al.*, 1997) and the marsh fritillary (*Euphydryas aurinia*) (Joyce and Pullin, 2001) all show substantial polymorphism within and among regional populations. Not only this, but the region sequenced for this study, the A+T-rich mitochondrial control region, is generally considered to be hypervariable in insects (Zhang & Hewitt, 1997), including lycaenid butterflies (Taylor *et al.*, 1993; Brookes *et al.*, 1997). For example, Brookes *et al.* (1997) used RFLP analysis of the control region to investigate metapopulation variation in the silver studded blue butterfly (*Plebejus argus*), and in sharp contrast to the three haplotypes found for the entire UK population of *P. bellargus*, they found 19 haplotypes in just a small area of north Wales.

However, the survey of *P. bellargus* in southern France revealed much higher levels of mitochondrial diversity. In contrast to the UK, where 50 individuals displayed three haplotypes, varying by just one or two VNTR insertions, the eight individuals from France were represented by six haplotypes, varying by up to 17 nucleotide differences. The gene diversity values for the two regions are correspondingly disparate, with the UK and France represented by values of 0.153 and 0.929 respectively. This variation was only found within the initial 193bp of control region, the remaining 529bp of the amplicon was found to be monomorphic among all 58 *P. bellargus* individuals that were sequenced. Within this 193bp, there are two insertions specific to *P. bellargus*, relative to other lycaenid species (as compared with *S. melinus* and *Jalmenus* spp.). Of the 17 sequence variations found among the *P. bellargus* specimens, seven are found within these inserted sequences, areas which might be expected to have higher rates of mutation than those displaying homology with the *J. evagoras* sequence.

Another four polymorphic sites found in *P. bellargus* are associated with the $(TA)_n$ VNTR. It has already been discussed that the latter regions of this VNTR, and the adjacent downstream sequences, appear to be somewhat conserved among invertebrate taxa (Taylor *et al.*, 1993). However, the number of (TA) repeat units present in the VNTR is variable among lycaenid and other invertebrate species (Taylor *et al.*, 1993), possibly indicating that there is lower conservation of the initial stretches of the VNTR or that nucleotide content is more important than repeat number. The two substitutions of cytosine for thymine within the VNTR of *P. bellargus* are somewhat unexpected, no such variation has been observed within or between any of the few published lycaenid sequences (Taylor *et al.*, 1993), only the *Drosophila* control region VNTR exhibits similar variation (Clary & Wolstenholme, 1985). However, these two substitutions were never found to occur simultaneously in any of the nine haplotypes found (see figure 4.6), perhaps suggesting a single substitution event that has subsequently undergone rearrangement within the control region of some individuals. The VNTR sequence between the two substitution sites would be unaffected by a reversal of this short region. Most of the French haplotypes exhibited the presence of the cytosine towards the end of the VNTR, whereas all of the UK haplotypes had a cytosine in the middle of the VNTR.

The analysis of the UK and French haplotype variation appears to reveal a separation of the UK haplotypes from the majority of those from France, as demonstrated by the maximum likelihood and minimum spanning network trees (figures 4.7 and 4.8). For the maximum likelihood tree, 100% of the bootstraps separated the two groups. The only exception to this was haplotype "Fr2", which groups strongly into the UK haplotypes. The high bootstrap values (70% to 100%) among the UK haplotypes (and "Fr2") are indicative of the easily resolved relationships among the haplotypes and the small amount of sequence divergence between them. Contrastingly, the values for the French haplotypes (except "Fr2") are all poorly supported (below 50%), indicating the relationships among the haplotypes can only be weakly resolved. The minimum spanning network tree also shows similar haplotype groupings and reveals a putative pattern of descent for the UK haplotypes: with "UK2" and "UK3" both stemming from "UK1", via the sequential insertion of (TA) repeats (or indeed a single insertion of a $(TA)_2$ repeat in "UK3"). With the exception of "Fr2", a minimum of 11 base pair differences can be found between any two UK and French sequences, a large amount of divergence that may result from the limited sampling effort for the French specimens. When it is considered that six haplotypes were obtained from just eight specimens, it is likely that with more comprehensive screening, additional haplotypes would be identified, a proportion of which would probably bridge this 11bp difference. The similarity between "Fr2" and the UK haplotypes also tends to support this conjecture. This means that instead of the phylograms representing two divergent groups of haplotypes, they may correspond to the presence of numerous diverse haplotypes in France as opposed to a small number of closely related haplotypes in the UK. The distinct similarities between "UK1" and "Fr2" suggest that the UK haplotypes may have a recent origin via a rapid colonisation event from France. It has been proposed that in fast colonising events, pioneers rapidly expand to fill new areas, and that the genes of these individuals will subsequently dominate the new population genome (Hewitt, 1999). Thus, the inference of a rapid colonisation of *P. bellargus* in the UK is supported by the evident lack of variation, relative to that present in France.

The maternal inheritance pattern of the mitochondrial genome provides a powerful indicator of colonisation events (Moritz, 1991; Harrison, 1989), and additionally can be used to estimate the numbers of individuals mediating them. In this investigation, where it has been shown that all of the observed UK haplotypes stem from a single predominant version, and that this is closely related to one found in France, where widespread variation is present, the only plausible explanation is that the colonisation of the UK by *P. bellargus* was mediated by one or a few female butterflies. Furthermore, the high level of sequence variation observed in the control region tends to indicate that this colonisation is a recent event, because variation would have accrued among the UK haplotypes by mutation over longer time periods.

The only alternative theory is that a range wide bottleneck reduced the UK variability to just a few closely related haplotypes. This is extremely improbable, because the entire UK population would need to have been reduced to one or a few females (and an unknown number of males) in order to display the variation found. The more probable outcome of this scenario would be the extinction of most UK populations, whilst a few butterflies survived in core areas, resulting in fixation of different haplotypes in separate geographic regions. However, in the majority of cases the UK populations are genetically identical to one another (figure 4.9). The only exception to this is Dorset, where two other haplotypes ("UK2" and "UK3") persist, probably encouraged by the fact that these populations have been observed to be far less prone to declines, events that would tend to remove rare and novel alleles arising from mutation.

If this genetic evidence for a recent UK founder event is combined with the historical population data for *P. bellargus*, the evidence for a recent colonisation becomes even more convincing. The failure of entomologists to describe *P. bellargus* anywhere in the UK until 1775 (Harris, 1775; Emmet and Heath, 1990), despite the publication of numerous natural history guides to insects and butterflies between 1710 and 1775 (Ray, 1710; Pettiver, 1717; Albin, 1720; Wilkes, 1741-1742; Wilkes, 1747-1749; Harris, 1775), indicates that *P. bellargus* must have either been extremely rare before this date, or was not present in the UK. The latter explanation would infer that the colonisation of the UK

occurred as recently as within the last 250 years. This does not rule out the possibility that the butterfly may have previously been native to the UK prior to this date, and subsequently became extinct, but it does indicate that all *P. bellargus* in Britain are the descendants of recent colonists, almost certainly from France. Whether this happened through a chance natural event (perhaps a mated female was blown from France during a storm) or at the hands of man will remain unknown. If the colonisation was anthropogenic, this inevitably raises the issue of whether the species should be considered to be native to Britain, leading to questions about its conservation. There is often debate as to the natural range of species, and those that are deemed to be non-native generally receive much lower conservation efforts. The apparent lack of genetic diversity within the UK also implies that these populations may be less important in the conservation of the species, in comparison to the French populations that appear to be much richer in diversity.

One important proviso of this work is that because of the inheritance pattern of mtDNA, no inferences can be made towards male mediated gene flow. Although there seems to be discrepancies regarding whether dispersal is male or female mediated in butterflies (Kuussaari *et al.*, 1996; Barascud, 1999; Mouson *et al.*, 1999), mark-release-recapture studies of *P. bellargus* have indicated that the male is the main proponent of gene flow (Thomas, 1983a; Emmet & Heath, 1990; Rusterholz, & Erhardt, 2000). This corroborates the implication that just one female may have founded the entire UK population, because females successfully managing to cross the English Channel, would be an extremely rare event. This does not however rule out the slightly more feasible possibility of males occasionally managing the crossing, a possibility that may explain the observation of large amounts of microsatellite variation across the UK (see chapter six). The high mutation rate of microsatellite DNA could also account for this, particularly when is it considered that over ~450 discrete generations of *P. bellargus* have passed since 1775.

5.0 THE ISOLATION OF MICROSATELLITE MARKERS

This work has been published as a primer note (Harper *et al.*, 2000).

5.1. INTRODUCTION

To date, very few studies have employed microsatellites for the study of lepidopteran species. The few investigations that have used microsatellites have tended to apply only a handful of markers (Palo *et al.* 1995; Meglecz *et al.*, 1998; Saccheri *et al.*, 1998; Keyghobadi *et al.*, 1999). One such study used as little as a single microsatellite (though this was used in parallel with several allozyme loci) (Saccheri *et al.* 1998). The largest number of microsatellites used in any published lepidopteran study is just four; used to study the spatial population structure in *Parnassius smintheus* (Keyghobadi *et al.*, 1999). However, with microsatellite data, the power of the analysis increases and inherent biases decrease rapidly with increasing number of markers. Ideally at least five or six markers should be employed in parallel (Hughes & Queller, 1993; Ruzzante, 1998). A recent review on the use of microsatellites in Lepidoptera (Neve & Meglecz, 2000) has showed that their identification is difficult in this group, which may explain the limited use of these markers in this taxon (Meglecz & Solignac, 1998; Meglecz *et al.*, 1998).

All published studies to date have applied a genomic library approach to the isolation of lepidopteran microsatellites, all of which have only had limited success (Keyghobadi *et al.*, 1998; Meglecz & Solignac, 1998). In this study, an approach using an enriched genomic library constructed for CA/GT dinucleotide repeats was used (see Kandpal *et al.*, 1994). This should increase the success of isolating microsatellite repeats, because the library will be enriched with cloned fragments selected for containing repeat regions.

5.2. MATERIALS AND METHODS

5.2.1. Creation of an enriched genomic library

The method was carried out as described by Kandpal *et al.* (1994). It is a simple and rapid technique for constructing small insert genomic libraries that have been highly enriched for dinucleotide repeats, in this case specifically targeting CA/GT repeats.

Genomic DNA was obtained from forty *P. bellargus* individuals by salting out the DNA from the butterfly heads (this extraction method is described in chapter four, section 4.2.1). This resulted in around 40µg of DNA, which was purified away from any contaminating proteins by subjecting it to a phenol chloroform extraction (Sambrooke *et al.*, 1989) and re-suspending it in 100µl of HPLC grade water (Sigma-Aldrich, UK). This purified DNA was then subjected to a restriction enzyme digest using *Mbo* I (Appligene-Oncor, UK), carried out according to the manufacturer's recommendations. The digested DNA was then run on an ethidium bromide stained agarose gel, from which fragments of 300-700bp were recovered using a gel recovery kit (Qiagen, USA). This recovered DNA was then quantified using a DNA Fluorometer (TK0 100, Hoeffer Scientific Instruments, USA), following the manufacturers instructions. Once quantified, 3µg of the recovered DNA was ligated to 9µg of an *Mbo* I adaptor molecule (created by the hybridisation of two oligonucleotides: 5'-GATCGCAGAATTCGCACGAGTACTAC and CGTCTTAAGCGTGCTCATG-ATGC-5' (synthesised by Genosys, UK)) (see Kandpal *et al.*, 1994). The ligation was carried out in a reaction volume of 300µl with 4000 units of T4 DNA ligase, using the manufacturer's recommendations (Promega, UK).

Subjecting the newly ligated DNA fragments to PCR amplification using non-hybridised adaptor oligonucleotides as the primers assessed the efficiency of the ligation. The PCR reaction mixture contained 10ng of template DNA; 2.5 U *Taq* DNA polymerase (ABgene, UK); 1.25mMoles of primer; 20mM $(NH_4)SO_4$; 75mM Tris-HCl, pH 8.8; 0.01%(v/v) Tween® 20; 1.5mM $MgCl_2$; 0.25mM dNTPs (ABgene, UK). The reaction was carried out under the following conditions: 1x 94°C, 4 min; 35x 94°C, 30s, 55°C, 30s, 72°C, 120s; 1x 72°C, 10 min. This reaction can only result in product amplification

if the DNA ligation had worked. The products of this PCR were run on an agarose gel and extracted using a gel purification column (Qiagen, USA), then quantified (using a DNA fluorometer) ready for the next step of the procedure.

These fragments represented the entire *P. bellargus* genome, it was therefore necessary to use an affinity capture method for enrichment, i.e. to selectively retain fragments that contained a $(CA)_n$ repeat motif. This was carried out by binding the purified DNA fragments to a biotin labelled $(CA)_n$ probe (only fragments with a $(GT)_n$ repeat could hybridise), then selectively retaining the fragments with a $(GT)_n$ repeat using streptavidin coated magnetic beads (Promega, UK) and a magnetic particle concentrator (Promega, UK). The hybridized fragments were captured via the biotin-streptavidin bead bond.

The fragments were hybridised to the $(CA)_n$ probe by denaturation in a boiling water bath for 10 minutes, followed by rapid chilling in an ice bath, then subsequent hybridization to a biotin labelled CA probe (Genosys, UK). This was carried out in a reaction volume of 200µl containing 2µg of denatured genomic DNA and 1µg of biotinylated $(CA)_{15}$ probe (Genosys, UK), in 0.5M sodium phosphate, pH 7.4, plus 0.5% SDS at 50°C for 18 h.

The beads (150mg) were prepared for enrichment; this was carried out by incubating them in 100µl of buffer A (100mM Tris, pH 7.5/150mM NaCl) and 100µg/ml of salmon sperm for 30-35 minutes at room temperature. This prevented non-specific binding of the *P. bellargus* DNA. The beads were then washed 3 times in "buffer A" using the magnetic particle concentrator (Promega, UK), thus removing any unbound salmon sperm DNA. Once the beads had been prepared, they were ready to use for enrichment of the biotin labelled DNA. All 100µl of "buffer A" was removed and 200µl of the CA/biotin labelled DNA was added to the beads, this was mixed gently then allowed to stand at room temperature for 30 min. The solution, plus any unbound DNA, was then removed from the beads, using the magnetic concentrator, and replaced with 100µl of "buffer A". The beads were successively washed 6 times with 100µl of "buffer A" (washes A to F), each time with increasing stringency (A: 30 minutes at room temperature; B: 5 minutes at room temperature; C: 5 minutes at 50°C; D: 5 minutes at

50°C in 1/10 dilution of buffer A; E: 5 minutes at 65°C in 1/10 dilution of buffer A; F: 5 minutes at 65°C in sterile water). The wash solutions removed at each stage were purified using PCR purification columns (Qiagen, USA) and retained (see also Mundy & Woodruff, 1996).

An aliquot (0.5µl) from each wash was subsequently PCR amplified using the *Mbo* I adaptor-primers. The amplified products were electrophoresed on a gel and blotted onto a positively charged nylon membrane (Amersham, UK). This membrane was then probed using a ^{32}P labelled $(CA)_n$ polynucleotide (Pharmacia, USA) at 55°C, and the filters exposed for autoradiography. This indicated which wash solution had undergone the optimum stringency conditions for the removal of DNA fragments containing $(CA)_n$ repeat motifs.

After identifying which wash solution was suitable for further work, the fragments from the solution were column purified (Qiagen, USA), quantified and used in the construction of a small insert library. This library was created by ligating the DNA fragments into a pGEM®-T Easy Vector System (Promega, UK), carried out using the manufacturer's recommendations, and then transformation of competent JM109 *E. coli* cells with the ligated vector, again, following the manufacturer's recommendations. The colonies produced were checked for successful insertion of the ligated vector by assessing the colour of the colony, with only white colonies being selected. These positive clones were then transferred onto LB agar plates (See Sambrooke *et al.*, 1989) containing Ampicillin/IPTG/X-Gal and grown overnight at 37°C for 16-24 hours.

5.2.2. Screening the library for clones containing microsatellite loci

The clones that had been selected for successful ligation and transformation were then each screened to see whether the DNA fragments they carried contained a $(CA)_n$ microsatellite repeat. The technique used for this was not according to the method described by Kandpal *et al.*, (1994), where a radioactive protocol was applied. Instead a PCR based approach was used, modified from the "PIMA" method described by Lunt *et*

al., (1999). The $(CA)_n$ repeat motifs were identified using colony PCRs, where the DNA was obtained from positive clones, by suspending a small part of the colony in 10µl of HPLC H_2O, heating it to 95°C for 5 minutes, then centrifugation at 14,000rpm. The DNA was diluted 1:5 with HPLC water, then PCR amplified using three primers; the two *Mbo* I linker specific oligonucleotides from section 5.2.1 plus a third, repeat specific oligonucleotide (5'-TGTGGCGGCCGC$(TG)_8$ V-3'). The 10µl PCR reaction mixture contained 0.15µl of template DNA; 5 U *Taq* DNA polymerase (ABgene, UK); 1.25mM of each primer; 20mM $(NH_4)SO_4$; 75mM Tris-HCl, pH 8.8; 0.01%(v/v) Tween® 20; 1.5mM $MgCl_2$; 0.25mM dNTPs (ABgene, UK) and was amplified under the following conditions: 1x 94°C, 4 min; 30x 94°C, 30s, 55°C, 30s, 72°C, 30s; 1x 72°C, 7 min. The PCR products were electrophoresed on an agarose gel, where reactions that contained an extra amplification product were identified. These represented an amplification where the $(CA)_n$ repeat specific oligonucleotide had bound, resulting in two PCR products, and thus the clone was deemed as containing a putative microsatellite.

5.2.3. Identification of microsatellites markers

In order to further characterise the putative microsatellites, the positive clones were prepared for sequencing by purification of the vector DNA. This involved using the original plated colony to inoculate 5ml of LB medium (Ampicillin (50µg/ml)), that was incubated at 37°C for 16-18 hours. After incubation, a 750µl aliquot of the culture was removed and added to 250µl of autoclaved glycerol, and archived at -20°C. The remaining 4.25mls of bacterial culture was further processed to isolate the plasmid. This growth medium was carefully removed from the cells after centrifuging at 2500 RPM at room temperature; the plasmid was then purified from the bacteria by the method described by Sambrooke *et al.* (1989). The DNA pellet obtained was re-suspended in 30µl of HPLC grade water with 1µl of RNAse added, to remove the RNA, then the DNA was column purified using a PCR purification kit (Qiagen, USA). In order to assess what volume of DNA to use in a sequencing reaction, it was run on to an ethidium bromide-stained agarose gel.

The clones that were positive for a CA/GT repeat were sequenced using big dye terminators on an ABI PRISM 377 automated sequencer (Perkin Elmer), using the forward primer from a set designed to amplify the inserted region within the vector (pGEM®-T Easy (Pharmacia, USA)). The sequencing reactions were carried out as recommended by Perkin Elmer.

5.2.4. Optimisation of the PCR amplification of the microsatellites

Once the sequences had been obtained, they were checked for the presence of putative repeat regions (>6 CA/GT). Where a microsatellite was identified, PCR primers were developed from the flanking sequences using the online software *PRIMER* 3.0 (Rozen & Skaletsky, 1998). These primers were then optimised for PCR by screening them under various annealing temperatures and reaction conditions, using a temperature gradient PCR block (Hybaid, UK). The reaction mixture contained 10-50ng of salt extracted *P. bellargus* DNA; 0.15 U *Taq* DNA polymerase (ABgene, UK); 0.2µMoles of each primer; 20mM $(NH_4)SO_4$; 75mM Tris-HCl, pH 8.8; 0.01%(v/v) Tween®-20; 1.5mM $MgCl_2$; 1.5mM $MgCl_2$; 0.25mM dNTPs (ABgene, UK). The PCR amplifications for each putative microsatellite were carried out in duplicate: in one case in this reaction mixture, in the second case with the addition of 1x Q solution (Qiagen, USA). The following conditions were used for amplification: 4 min at 94°C; followed by 35 cycles of 30 seconds at 94°C, 30 seconds at annealing temp, 30 seconds at 72°C; followed by 30 minutes at 72°C. The PCR product was visualised on a 1% agarose gel, stained with ethidium bromide. In some cases it was necessary to re-design the primer pairs and repeat the subsequent optimisation process.

5.2.5. Determining levels of polymorphism

Once the PCR amplification had been optimised for the microsatellites, it was crucial to assess whether the loci are polymorphic. This was carried out by randomly labelling the microsatellites with ^{32}P dCTP during PCR amplification. The products were then amplified under the following conditions: 10-50ng of salt extracted *P. bellargus* DNA; 0.15 U *Taq* DNA polymerase (ABgene, UK); 0.2µMoles of each primer; 20mM $(NH_4)SO_4$; 75mM Tris-HCl, pH 8.8; 0.01%(v/v) Tween®-20; 1.5mM $MgCl_2$; 1.5mM $MgCl_2$; 0.25mM each of dATPs, dGTP, dTTP (ABgene, UK); 0.25mM of ^{32}P dCTP (Pharmacia, USA); in the case of LbG2, 1x Q solution (Qiagen, USA) also used. The following conditions were used for amplification: 4 min at 94°C; followed by 35 cycles of 30 seconds at 94°C, 30 seconds at annealing temp (see table 5.1), 30 seconds at 72°C; followed by 30 minutes at 72°C. The microsatellites were scored by a four hour run on a 6% acrylamide gel, which after drying was exposed overnight to photographic film (Kodak), then developed using an automated Kodak developer.

The microsatellites were then screened for polymorphism using a more sensitive technique. This time PCR amplifications were carried out using 36 *P. bellargus* specimens under the same conditions as detailed in section 5.2.4, except that the forward PCR primers were 5' end-labelled with a fluorescent phosphoramidite (FAM, HEX or TET) (MWG-Biotech, Germany) and non radioactive dCTP was used. The microsatellites were scored using a 5% denaturing polyacrylamide gel by vertical electrophoresis at 20-60mA for 2h on a Perkin Elmer ABI 377 automated sequencer running *GENESCAN* and *GENOTYPER* software (PE-Applied Biosystems, USA). The data from this analysis was used to assess the microsatellites for their compliance with the Hardy-Weinberg equilibrium. For this analysis, the computer programme *ARLEQUIN* (Schneider *et al.*, 2000) was used to calculate comparisons between the observed and expected heterozygosities (H_O and H_E respectively). The five loci were also analysed for the possible effects of linkage disequilibrium among loci. This was carried out within the programme *GENEPOP*, version 3.2 (Raymond & Rousset, 1995),

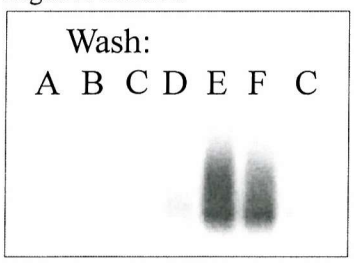
using the data from all 26 UK populations of *P. bellargus* (detailed in chapters three and six).

5.3. RESULTS

5.3.1. Creation of the enriched genomic library

The genomic library from *Mbo* I digested *P. bellargus* DNA was successfully created as described in section 5.2.1. The insert sizes ranged between 200bp and 700bp, with an average size of around 500bp, this size range was dictated by the frequency at which *Mbo* I cleaved the genomic DNA. Once the *Mbo* I adaptor molecule had been ligated to the inserts, they were PCR amplified using the adapter oligonucleotides. The successful amplification of these inserts indicated that the ligation had been successful.

Figure 5.1. Photographic image of an auto-radiograph of a ^{32}P labelled $(CA)_n$ probed Southern blot of the various washes from the Vectrex Avidin beads. A-F denotes the wash sequence; lane C represents the negative control.



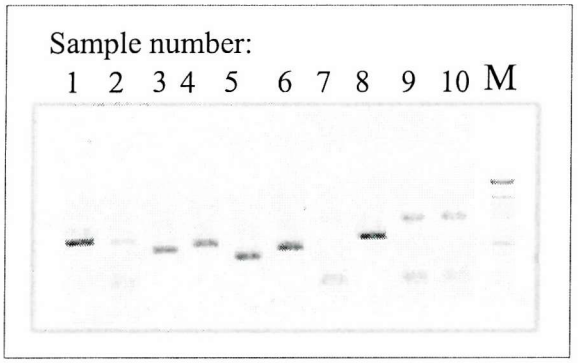
The affinity capture for the enrichment of $(CA)_n$ repeats was successful, as demonstrated by the successful autoradiograph of the membrane that was screened using a ^{32}P labelled $(CA)_n$ probe (figure 5.1). It can be seen from this figure that the majority of the fragments positive for a $(CA)_n$ repeat motif can be found in washes E and F, the two most stringent washes. The longer CA/GT repeat regions are likely to have been released from the magnetic beads in the most stringent wash (F), whereas the fragments in wash E are more likely to be shorter repeats. Therefore, wash solution F was used for all further work. However, the other five were archived at –20°C in case they were required for further analysis.

When the purified fragments from wash F were inserted into the pGEM®-T Easy Vector System (Promega, UK) and used to transform the JM109 *E. coli* cells, 123 colonies were noted to be devoid of any blue colouring. These were deemed as positive for the presence of the vector and insert, and were used for further analysis.

5.3.2. Screening the library for clones containing microsatellite loci

The PIMA technique that was used to screen each of the clones for microsatellite repeats was highly successful. For example, when screening a subset of ten clones from the previous step, four identified as positive in the previous step (figure 5.2).

Figure 5.2. An inverted tone image of an ethidium bromide stained agarose gel, showing the PCR products from a "PIMA" amplification of 10 positive clones (1-10 represent clones: 1/1; 1/2; 1/3; 1/4; 1/5; 1/6; 1/7; 1/8; 1/9; 1/10).



The additional PCR products present in lanes 2, 7, 9 and 10 all indicate the potential likelihood of a microsatellite repeat within the library insert (figure 5.2). However, if the gel is studied more closely, it can be seen that the band appearing at around the 1.1Kbp mark is common to most of the amplifications. It is possible that this band originates from non-specific binding of the primer and hence some kind of spurious PCR product. This band is absent from Lane 2, although there are still two PCR products observable. This clone is a strong candidate for the presence of a putative $(CA)_n$ microsatellite repeat.

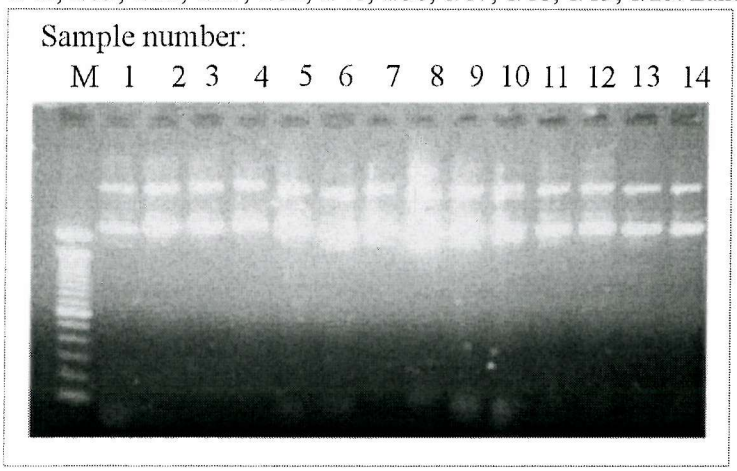
All of the 123 colonies positive for the vector and insert were screened using the PIMA technique (Lunt *et al.*, 1999). Only a subset of these PCR amplifications has been shown.

In total, 43 of the clones screened appeared to be positive for the presence of a $(CA)_n$ microsatellite repeat.

5.3.3. Identification and PCR optimisation of the microsatellites

The 43 positive clones were successfully prepared for sequencing, and aliquots of each bacterial inoculum was prepared and archived at –20°C. The purified plasmid DNA was electrophoresed on an agarose gel in order to assess the concentration of DNA (figure5.3).

Figure 5.3. An ethidium bromide stained agarose gel showing a selection of DNA samples extracted and column purified from the clones found positive for a repeat region. (Lanes 1-14 represent: 1/57; 1/54; 1/47; 1/43; 1/39; 1/36; 1/33; 1/31; 1/41; 1/30; 1/17; 1/18; 1/19; 1/25. Lane M is a 100bp molecular marker)



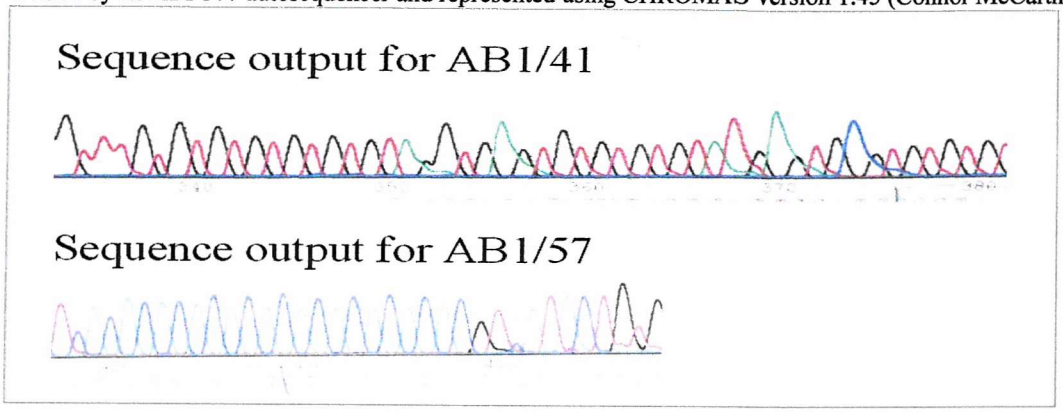
When the DNA is purified from JM109 *E. coli* colonies that have been transformed with the vector, two clear bands can be seen on the gel (figure 5.3). The higher molecular weight band is the bacterial genomic DNA; the second, smaller band is the plasmid DNA. The bacterial genomic DNA will not interfere with the sequencing reaction because of the specificity of the primer used in the PCR.

When the positive clones were sequenced, many did not contain significant repeat regions. Of the 43 positive clones, only 21 were found to contain repeat regions of note (>6 CA/GT), the other 22 contained short stretches of $(CA)_n$, consisting of just two or

three repeats. The frequency of these "false positives" could probably be reduced if the annealing temperature of the PCR used in the "PIMA" technique had been raised slightly.

Two examples of repeat regions from two of the sequence outputs from the ABI PRISM 377 automated sequencer are shown in figure 5.4.

Figure 5.4. The graphical output of the microsatellite sequences for Lb1/41 (above) and Lb1/57 (below), created by an ABI-377 autosequencer and represented using CHROMAS version 1.45 (Connor McCarthy).



The figures both show microsatellite repeat arrays. Lb1/41 shows a highly interrupted CA repeat (figure 5.4), though still a potentially useful microsatellite, whereas Lb1/57 shows an uninterrupted repeat (figure 5.4): $(CA)_{13}$. There were several cases where the sequences obtained from positive clones showed an extensive repeat region, but with the repeat situated directly adjacent to the *Mbo* I linker molecule, with no available sequence for primer design. This is unfortunate because some of these repeat regions were extensive, and likely to be extremely polymorphic.

A number of the sequences containing microsatellites were duplicated in different clones; with on average each microsatellite being represented twice. This reduced the number of unique repeat regions identified to 11, of which only 9 were suitable for primer design on both sides of the microsatellite.

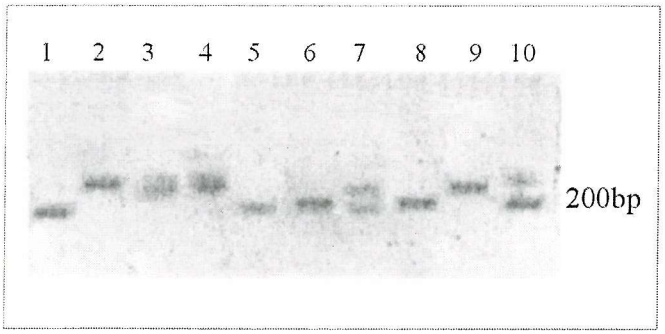
The initial round of primer design from the sequences was carried out using *PRIMER* 3.0 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html). When these primers were used to PCR amplify *P. bellargus* DNA, only four of the microsatellites amplified successfully (Lb1/41, Lb4/18, Lb1/57 and Lb3/41). After further optimisation (addition of 1x Q solution (Qiagen, USA)), two more primer sets resulted in amplification (LbG2 and Lb3/6), though one of these consistently amplified a number of spurious bands alongside the probable microsatellite (Lb3/6).

The three microsatellites that did not amplify (Lb3/27, Lb2/17, and Lb4/19 were again subjected to primer design both using *PRIMER* 3.0 and by eye. The new primers were screened and resulted in the successful amplification of Lb4/19, but neither Lb3/27 nor Lb2/17 could be optimised.

5.3.4. Determining levels of polymorphism

Once the PCR conditions for all of the microsatellites had been optimised, they were assessed for polymorphism. This was carried out via an agarose gel for Lb4/19, this microsatellite was variable enough that polymorphism could easily be identified using this method. For all others, autoradiographic screening of an acrylamide gel was used.

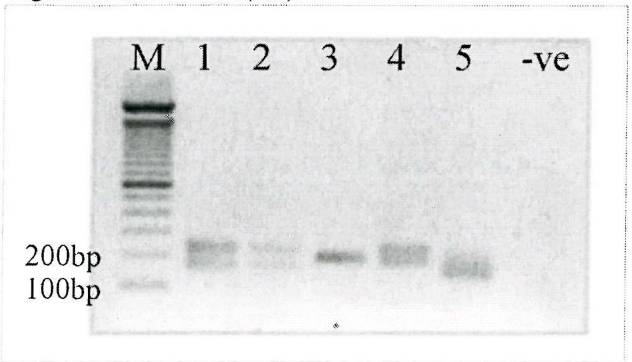
Figure 5.5. An autoradiograph of PCR amplified Lb1/41 from a group of ten *P. bellargus* individuals, demonstrating the levels of polymorphism present.



The microsatellite (Lb1/41) is clearly polymorphic among the 10 randomly chosen *P. bellargus* individuals (figure 5.5). Similarly, Lb4/19 shows high levels of polymorphism

between the five *P. bellargus* screened (figure 5.6). Similar results were also obtained for LbG2, Lb1/57 and Lb4/18. The repeat region, Lb3/41, was found to be monomorphic.

Figure 5.6. An inverted tone image of an ethidium stained 1% agarose gel, showing polymorphism at Lb4/19 between a set of five *P. bellargus* individuals. Also shown are a 100bp molecular marker (M) and a negative PCR control (-ve).



When the microsatellites were screened for their compliance with the Hardy Weinberg equilibrium using the computer programme *ARLEQUIN* (Schneider *et al.*, 2000), significant discrepancies were identified between the observed (H_O) and expected (H_E) heterozygosities (table 5.1). When the loci were tested for linkage disequilibrium, only nine out of 260 combinations of loci and populations were found to be significant. When a Bonferonni sequential correction was performed on the results of this analysis, all but one result of the previously significant results were found to be non-significant. This was between loci LbG2 and Lb4/19 in the population at Devil's Dyke. However, no linkage disequilibrium was identified for this locus combination in any of the other 25 populations tested.

A summary of the characteristics of each of the polymorphic microsatellites isolated is shown in table 5.1.

Table 5.1. Characteristics of five polymorphic microsatellite loci from *P. bellargus* with core repeat, primer sequences, size range of PCR product, optimal PCR annealing temperature (T_A), observed (H_O) and expected (H_E) heterozygosity. GenBank accession numbers: AF276048-AF276052.

LOCUS	MICROSATELLITE	PRIMER SEQUENCE (5'-3')	T_A (°C)	SIZE RANGE (BP)	H_O	H_E
Lb1/57	$(CA)_{13}$	F:TGTATCAGCAACAGCTCGGT R:GGAAGCGTTCATCGGTAAAA	46	172-224	0.281	0.487
Lb4/18	$(CA)_8$GGCCCC-GCCGC$(T)_9$	F:GATACCTATGCCAGGCTCCA R:CGAATGTCATACAGGTTGCG	50	146-153	0.500	0.742
Lb4/19	$(GACGGT)_{15}$	F:AGATTAGGCATTGGCGTGTC R:CTGCGATCCATTTCCGTTA	50	180-260	0.778	0.854
Lb1/41	$(GT)_7$AGGTGA$(GT)_5$ ATGAGTGC$(GT)_4$	F:CGTGTCTGTCACCCCCTTTA R:ATGACGGGTAGGGATTAGGG	50	170-236	0.485	0.653
LbG2	$(CA)_{10}$GTC$(CGCA)_3$ ATGC$(CAGC)_6(GC)_3$ TCT$(CA)_{12}(GA)_4$	F:ATCAAGGTCCGCACAGCA R:CGCTCGCTTATCAGACAACA	50	164-182	0.622	0.681

5.4. DISCUSSION

The enriched genomic library approach used in this study was extremely successful for the isolation and characterisation of lepidopteran microsatellites. Previous studies using non-enriched approaches have only had limited success, with no published study yielding more than four microsatellites, even after targeting more than one repeat motif (Keyghobadi *et al* 1999). As with other published studies (Meglecz & Solignac, 1998), a large proportion of the microsatellite repeats found were either compound or interrupted.

In total only 11 putative microsatellites were revealed, and even though the library was targeted only at CA/GT repeats, other repeat motifs were found; for example, Lb4/19 is an array of $(GACGGT)_{15}$. It is hard to explain why repeats such as this should have been isolated when the probes used were specific to $(CA)_n$ repeat motifs. The only possible explanation is that the GT within the GACGGT motif bound to the $(CA)_{15}$ probe. Another non-$(CA)_n$ repeat that was isolated was $(CACAT)_n$, though it is easier to see why

this repeat motif would be isolated. Many of the repeat arrays consisted of huge numbers of tandem repeats (>400bp), though it is unfortunate that all of these were unsuitable for primer design because of their close proximity to the linker. Perhaps if a restriction enzyme that cuts less frequently than *Mbo* I had been used, then this problem of the repeat motif being adjacent to the linker region would have been avoided. The presence of these long tandem arrays does tend to contradict claims previously made in literature, that infer repeat arrays in Lepidoptera are generally short (Meglecz *et al.* 1998).

Of the 11 putative microsatellites, only nine were suitable for subsequent primer design, and of these nine, only five polymorphic microsatellite loci were identified. This success rate of 45% is high when compared to other studies where the success rate ranged from 0 to 16% with an average success rate of 3.3% (taken from 7 unpublished studies, summarised by Meglecz & Solignac, 1998).

When the loci were tested for linkage, no significant effect was found. Whereas when observed heterozygosities (H_O) were calculated for each microsatellite locus, they were found to be significantly lower than the expected heterozygosities (H_E). This homozygote excess may result from a number of factors, such as inbreeding or the population tested failing to comply with one or more conditions of the Hardy-Weinberg equilibrium. For example, the populations tested were relatively small and possibly inbred (although even with increased mating between related individuals, random mating should eliminate any dis-equilibrium between the values (Wright, 1951)) or there may be some sub-structuring within the populations, inhibiting random mating (the Wahlund effect) (Palo *et al.*, 1995). However, if these are factors are having an impact, all five markers would be expected to show a similar level of disparity. The most likely possibility is that there may be "null alleles" present. This is where the primer binding site for a microsatellite has mutated, resulting in a non-amplifying allele, and hence heterozygotes may be wrongly assumed to be homozygotes (Lehman *et al.*, 1996; Van Treuren, 1998). This has been shown to be a relatively common occurrence in human microsatellites (Callen *et al.*, 1993). Obviously this will result in an apparent excess of homozygotes. This possibility will be further investigated in chapter six.

6.0 THE MICROSATELLITE ANALYSIS

6.1. INTRODUCTION

Southern England represents the most northerly range of *P. bellargus* in Europe, where its distribution is restricted to south facing escarpments of calcareous grassland. This reliance on various habitat features results in an extremely local distribution (Thomas, 1983a), where populations rarely exceed 1000 individuals (Emmet & Heath, 1990; Bourn *et al.*, 1999; Stewart *et al.*, 2000). The sedentary nature of the species means that even where populations are in close proximity to one another, transfer of individuals is rare, with dispersal distances rarely exceeding 25 metres (Thomas, 1983a). This means that there is apparently only limited opportunity for gene flow among colonies, a situation aggravated by the geographical partitioning of calcareous grassland sites across the UK, where distances are often in excess of 50 kilometres. These habitat patches show immense variation in size and proximity to one another, with some areas being represented by a network of metapopulations; however, the extent of this population connectivity varies hugely among geographic regions. Typically, populations are separated from their nearest neighbour by tens of metres to a number of kilometres (e.g. Thomas, 1983a; Emmet & Heath, 1990).

The species underwent a chronic decline between the 1950's and 1970's, followed by an acute decline in the late 1970's, resulting in more than 90% of the UK populations becoming extinct by 1981 (Thomas, 1983a; Emmet & Heath, 1990; Asher *et al.*, 2001). The species has slowly recovered from this, with over twice the number of populations currently present relative to that of 1978-1981 (Emmet & Heath, 1990). This chapter aims to test the prediction that the sedentary nature of *P. bellargus* adults and their patchy distribution have combined to constrain gene flow over all but short distances. Herein, patterns of microsatellite DNA variation will be analysed within and among populations from throughout the UK range, to assess the patterns of gene flow and the relationships among populations separated by much greater distances than covered by dispersing adults. This may elucidate the patterns of recolonisation for populations that had become

extinct, whilst also identifying whether the genetic bottleneck experienced by the species has had lasting effects on the population structure and genetic diversity. Levels of heterozygosity will also be investigated relative to various population characteristics, possibly identifying putative causes of inbreeding.

Five polymorphic microsatellite loci have been isolated and characterised for *P. bellargus* (See Chapter five and Harper *et al.*, 2000). This chapter details the application of these markers to screening the levels of genetic diversity within and between the UK populations of the butterfly, and the analysis of the data obtained. All 26 populations sampled, comprising a total of 1173 specimens, were subjected to microsatellite analysis using all five markers. A summary of the field collection and historical data associated with each of these 26 populations is detailed in chapter three. A mixture of population histories was sampled (*i.e.* populations that have undergone demographic changes such as decline, recolonisation, expansion alongside populations that have had no severe population turbulence), this should give an idea of the effects that a decline has on a discrete population.

6.1.1. Dispersal and gene flow between populations

Dispersal can be defined as an event where an individual permanently leaves its home range, or site, for an area where they will reproduce or would have reproduced if it had survived and found a mate (Howard, 1960). Most organisms will undergo some form of dispersal during their life cycle. The most important genetic consequence of dispersal is that it generally results in gene flow, defined as "the movement of genes and their subsequent incorporation into new gene pools" (Endler, 1977; Gaines and McClenaghan, 1980). Obviously, dispersing individuals must breed at their new site for gene flow to occur and their offspring must not be removed from the population via selection (Barton, 1992). Unless the levels of gene flow are high, other stochastic events such as genetic drift, and natural selection will produce fluctuations in allele frequencies, encouraging differentiation and structuring among populations (Peterson, 1996).

The benefits of dispersal are clear: the association of an individual with different conspecific organisms (thus avoiding breeding with related individuals), or an improvement in the quality of its habitat, will tend to improve its fitness and reproductive potential. The theoretical rationales behind the prevalence of dispersal are the avoidance of inbreeding (Pusey, 1980; Clutton-Brock, 1989) and to reduce competition for resources (Greenwood, 1980; Dobson, 1982). These grounds are particularly relevant for mammals and other large animals, especially where there is competition for resources such as mates. For smaller organisms with low vagility, in particular some sedentary butterfly species, transfer may also have a more accidental basis such as wind dispersal (Boughton, 2000). Additionally, certain environmental factors can affect the rate of dispersal (and hence gene flow) between populations. For example, butterfly populations connected by "corridors" of suitable habitat, or open areas, are more likely to experience population exchange than those separated by barriers such as unsuitable habitat (e.g. for grassland butterflies woodland will act as a barrier) or geographic distance (Haddad, 1999; Roland, 2000). Population and habitat characteristics also seem to affect dispersal, for example both high densities of butterflies and an abundance of flowers appear to reduce emigration, although this will not necessarily increase immigration (Kuussaari *et al.,* 1996). It is also possible that for species that exist in isolated populations, natural selection will encourage sedentary behaviour. This is because dispersal from the natal patch is likely to result in lower reproductive potential as a result of the lack of cospecifics, possibly even mortality.

6.1.2. The implications of reduced gene flow, habitat fragmentation and inbreeding

Environmental factors can have a severe impact on dispersal and gene flow between populations. Habitat destruction caused by intensive agriculture and urban development is fragmenting the distributions of many species (Mader, 1984). This is particularly the case in the United Kingdom, where many species are confined to small pockets of

suitable habitat (Hanski, 1994), resulting in highly subdivided populations with reduced gene flow between them.

In this situation, sedentary butterfly species usually exist in closed populations, implying that most individuals will spend their lifetimes in a single habitat patch (Thomas, 1984; Warren, 1992; New *et al.*, 1995). Generally these patches exist within a metapopulation – a network of small, extinction-prone local populations, interconnected by occasional dispersal events (Hanski & Thomas, 1994; New *et al.*, 1995). A large amount of research has been carried out regarding the dispersal of butterfly species that exist in metapopulations (e.g. Debinski, 1994; Hanski & Thomas, 1994; Hanski *et al.*, 1994; Hill *et al.*, 1996; Neve *et al.*, 1996; Lewis *et al.*, 1997; Sutcliffe *et al.*, 1997; Saccherri *et al.*, 1998). In this situation, where organisms have a short dispersal range, isolation by distance can have a considerable impact on genetic differentiation between populations and long distance gene flow can only take place via intermediate populations and stepping stone habitats (Neve *et al.*, 1996; Peterson, 1996). Many studies have demonstrated that gene flow among insect populations declines with increasing geographic distance (Archie *et al.*, 1985; Loxdale *et al.*, 1985; Liebherr, 1986; King, 1987; Kukuk *et al.*, 1987; McCauley & Eanes, 1987; Sokal *et al.*, 1987; Rosenburg, 1989; Costa & Ross, 1993; Britten *et al.*, 1995). In extreme cases, the isolation of populations can lead to changes in morphological traits such as those associated with reduced flight (Dempster *et al.*, 1976; Dempster, 1991; Thomas, 1998).

This reduction in gene flow will result in genetic drift between populations, and this may be exacerbated by frequent demographic events such as founder effects and population bottlenecks. This is because of poor temporal stability, particularly in small populations, resulting in repeated cycles of both extinction and recolonisation, and decline and recovery. The resulting population turnover can be rapid in isolated habitat patches, where it has been proposed that small populations rarely persist longer than 100 years, whereas in larger habitat areas, where the population sizes are greater, the rate is much reduced (Thomas & Harrison, 1992; Hanski & Thomas, 1994). Therefore the core populations will continually be acting as a source for recolonising the smaller, peripheral

sink populations (Boughton, 2000). However, Hanski and Thomas (1994) stress that for many temperate butterflies, the size and distribution of habitat patches is such that even the core sites would not be safe from extinction, consequently survival of all populations must be seen in a metapopulation context.

Not only are the populations within a metapopulation at risk of extinction; other effects associated with metapopulation dynamics will also have an impact. In particular, population bottlenecks will frequently occur, which will strongly influence the genetics of remnant populations (Brookes *et al.*, 1997). This genetic change is important in terms of fitness, where simple genetic arguments have generated serious concerns about the loss of genetic variation in small, isolated populations and consequently their long term viability (Frankel & Soule, 1981; Soule, 1987; Gilpin, 1991). This increased frequency of bottlenecks and inbreeding may be the most important genetic consequence of the population fragmentation involved in metapopulation structure. Numerous studies have inferred a disadvantage to inbreeding, for reviews of this, see Frankham (1995a) and chapter one, section 1.2 of this thesis, however, there is currently little evidence at present to support this (Wauters *et al.*, 1994). It would therefore be prudent to investigate this possibility in *P. bellargus*, where the species not only exists as a metapopulation, but has also undergone severe population bottlenecks across its UK range.

6.1.3. The analysis of gene flow and population structure

Approaches to the analysis of dispersal and population structure can be grouped into two categories: direct and indirect methods. Direct methods are those that use actual observations of movements of individuals, such as mark-release-recapture (Morton, 1982; Roland, 2000), whereas indirect methods use genetic data to infer movement (Slatkin, 1985). The direct methods of studying gene flow have the advantage that they are based on actual movement of individuals. However, they have several disadvantages:

1. Casual observations can be misleading because the dispersal is likely to extend further than observational methods can assess. Therefore, their ultimate fate will remain unknown (Slatkin, 1985).
2. The reproductive success of immigrants will be assumed equal to that of the residents. This will not necessarily be the case (Kaufman & Wool, 1992).
3. Direct measurements from mark and recapture studies are both temporally and spatially limited – enough individuals must be marked and subsequently recaptured in order to obtain robust data (Slatkin, 1987).
4. Marking and monitoring individuals for more than a few migration events can be extremely impractical (Dobzhanskky & Wright, 1943; Roderick, 1996).
5. No information can be obtained regarding the genetic status of the population.

These difficulties in using direct measures of dispersal and gene flow have lead to the common use of indirect measures, extrapolated from genetic frequency data (Whitlock, 1999). Here the population structure can be inferred from data obtained using molecular markers (see chapter two for a review of techniques available). Data such as allele frequencies for defined population units can be used to calculate levels of similarity and subdivision among the populations, and hence build up a picture of the genetic relationships and levels of gene flow between them (Wright, 1931; Slatkin, 1985; Roderick, 1996; Whitlock, 1999). Besides the comparative ease of data generation (Avise, 1994), molecular data can theoretically provide a temporal aspect to population structure (Koenig *et al.*, 1996), and will also only reflect dispersal that has lead to successful reproduction.

Numerous molecular techniques are available for this type of study (see section chapter two), and although many researchers have used protein polymorphisms to estimate dispersal rates and nuclear gene flow in an enormous number of species, including butterflies (Brookes *et al.*, 1997; Meglecz *et al.*, 1997 & 1999; Saccheri *et al.*, 1998; Clarke & O'Dwyer, 2000; Raijmann & Menken, 2000; Sula & Spitzer, 2000), microsatellite markers have largely supplanted their use (Goldstein & Pollock, 1997). Microsatellite loci are tandemly repeated DNA sequences, with a repeat unit length of

between one and six base pairs (see Tautz & Renz, 1984; Litt & Lutty, 1989; Jarne & Lagoda, 1996). They are commonly found in non-coding regions of the genome, and are generally considered to be selectively neutral (Ashley & Dow, 1994; Avise, 1994). This lack of selective pressure alongside their abundance and high levels of polymorphism make them a better choice of marker for population genetic studies than allozymes, which are functional enzymes and consequently subject to selective pressures (Lagercrantz *et al.*, 1993; Avise, 1994). Nuclear microsatellites have been used in numerous studies on a wide range of organisms with apparent success e.g. *Drosophila melanogaster* (England *et al.*, 1996), red grouse (*Lapogus lapogus scoticus*) (Piertney *et al.*, 1999), the northern hairy nosed wombat (*Lasiorhinus krefftii*) (Taylor *et al.*, 1994), water buffalo (*Bubalus bubalis*) (Barker *et al.*, 1997) etc.

Although it is relatively easy to obtain large amounts of genotype frequency data using microsatellite markers, there are numerous approaches to the statistical analysis of these data. Many of these calculate an estimation of genetic distance between populations using the frequency data, each with a slightly different approach to the analysis. Some of these were originally developed for other molecular methodologies, whereas some were developed specifically for microsatellite data (see chapter two, section 2.2.8.1 for a review of these). Almost all of these approaches are based on the assumption that gene frequency distributions are dictated by gene flow (Cockerham & Weir, 1993; Gaggiotti *et al.*, 1999), and that gene flow is representative of dispersal. However, Bossart & Prowell (1998) warn that gene flow is only one of a number of contributory factors shaping population genetic structure. They suggest that current statistical approaches may make too many assumptions about temporal homogeneity and break down under the influence of stochastic environmental events. Therefore caution should be taken in this approach.

Many of the genetic distance estimators that can be applied to microsatellites allow for a mutation model. These range from the "infinite allele model" (IAM), where it is assumed that a single mutational event can transform an allele to any other allelic state (Kimura & Crowe, 1964), to the "stepwise mutation model" (SMM), where a single mutational event can only alter an allele by a single repeat unit (Shriver *et al.*, 1993; Valdes *et al.*, 1993).

Combining aspects of these two extremes (IAM and SMM) can create other models, these are known as two-phase models (TPM) (DiRienzo *et al.*, 1994). Indices such as F_{ST}, R_{ST}, G_{ST} and θ are all measurements based upon one of these models. Another group of estimators is also available; these take a more simplistic approach, by calculating a Euclidean distance measure without the inclusion of a mutation model. A number of studies have found these approaches are particularly efficient for reconstructing evolutionary relationships with microsatellite loci (Takezaki & Nei, 1996; Lugon-Moulin *et al.*, 1999). These include measures such as Nei *et al.*'s (1983) D_A and Cavalli-Sforza and Edwards (1967) chord distance. Other analytical measures are also available for use with microsatellite data; these do not involve simple calculations based on gene frequency data. Here a more sophisticated modelling approach is taken to deduce genealogical processes, inter-population relationships, effective population size, etc.

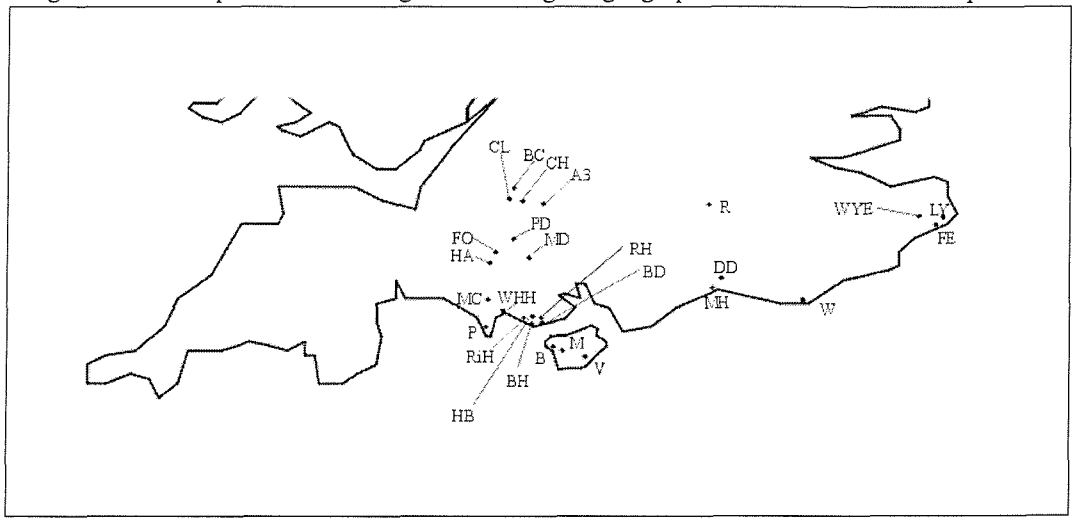
Not only can these analyses be used to determine genetic relationships and patterns of gene flow between populations, but past demographic events can also be elucidated. For example, statistical software packages are available that can detect recent changes in effective population size (N_e), such as bottlenecks, or population expansion. These methods assess the data for signatures of these changes such as imbalances between allele size and heterozygosity (Kimmel *et al.*, 1998). For example, populations exhibiting significant heterozygote excess would be considered as having experienced a recent genetic bottleneck. This heterozygosity excess should not be confused with an excess of heterozygotes, the former compares the observed and expected heterozygosities in the sense of Nei's (1978: p177) gene diversities, whereas the latter compares the number of heterozygotes with the Hardy-Weinberg expectation (Luikart & Cornuet, 1998). Obviously with such a wide range of techniques available for data analysis, it is difficult to judge which of these would be most appropriate to this project. Ideally the data should be subject to multiple analyses applying different statistical approaches, bearing the appropriate hypotheses in mind.

6.2. MATERIALS AND METHODS

6.2.1. Sample Collection

Details of sample collection over two years, comprising four consecutive field seasons, from populations across southern England, as detailed in chapter three, section 3.3 (Table 3.1). The populations represented the entire UK range of *P. bellargus*, with butterflies sampled from at least one site within each region (see figure 6.1).

Figure 6.1. A map of Southern England showing the geographic location of all 26 sampled.



6.2.2. DNA extraction

The DNA was extracted from the *P. bellargus* individuals following the salting out procedure described in chapter four, section 4.2.1. Extracted DNA was stored in 20-40µl of T.E. buffer (pH8) at –20°C.

6.2.3. Analysis of the microsatellite DNA

The microsatellites developed in chapter five (Harper *et* al., 2000) were employed for this analysis. The reaction mixture contained 10-50ng of salt extracted *P. bellargus* DNA;

0.15 U *Taq* DNA polymerase (ABgene, UK); 0.2µMoles of each primer (the forward primer of each pair was 5' end labelled with fluorescent phosphoramidites FAM, HEX or TET) (MWG-Biotech, Germany); 20mM $(NH_4)SO_4$; 75mM Tris-HCl, pH 8.8; 0.01%(v/v) Tween®-20; 1.5mM $MgCl_2$; 1.5mM $MgCl_2$; 0.25mM dNTPs (ABgene, UK); in the case of LbG2, 1x Q solution (Qiagen, USA) was used. The following conditions were used for amplification on a "Touchdown" thermal-cycler (Hybaid, UK): 4 min at 94°C; followed by 35 cycles of 30 seconds at 94°C, 30 seconds at annealing temp (50°C for Lb1/41, 1/G2, 4/19, 4/18 and 46°C for Lb1/57), 30 seconds at 72°C; followed by 30 minutes at 72°C. The microsatellites were scored using a 5% denaturing polyacrylamide gel by vertical electrophoresis at 20-60mA for 2h on a Perkin Elmer ABI 377 automated sequencer running *GENESCAN* and *GENOTYPER* 2.5 software (PE-Applied Biosystems, USA).

6.2.4. Analysis of the data

6.2.4.1. *Hardy-Weinberg analysis*

Before using the microsatellite data to make inferences regarding the population structure, they were analysed to assess their independence and for any possible bias. In a randomly mating population, observed heterozygosity calculated from allele frequencies would be expected to conform to the expected value calculated within the Hardy-Weinberg equilibrium.

The presence of non-amplifying or "null" alleles at microsatellite loci can complicate the interpretation of genotypic data, resulting in a reduced level of observed heterozygosity (Lehmann *et al.*, 1996). This is because a heterozygote possessing a null allele will mistakenly be assumed to be homozygous. Theoretically, the presence of these null alleles can be detected by testing each population for a departure from the Hardy-Weinberg equilibrium. If null alleles are present, then there will be a heterozygote deficit relative to the expected values of the equilibrium (Lehmann *et al.*, 1996; Van Treuren, 1998).

Microsatellite genotypes were tested for departure from Hardy-Weinberg equilibrium for all 26 populations at each locus, using the computer programme *ARLEQUIN* (Schneider *et al.*, 2000). *ARLEQUIN* employs a Markov-chain random walk algorithm (Guo & Thompson, 1992), which is analogous to Fisher's (1935) exact test but extends it to an arbitrary sized contingency table. In order to reduce the likelihood of type I errors, a Bonferroni correction was applied to the data. Where a significant heterozygote deficit was found, estimations were made towards the frequency of null alleles (n) by the method of Brookefield (1996). The method involved applying one of the following equations:

(1) $$n = \frac{A + \sqrt{(A^2 + B)}}{2\,(1 + H_e)}$$

or:

(2) $$n = \frac{(H_e - H_o)\mathrm{s}}{(1 + H_e)}$$

Where: n is the frequency of null alleles, H_e is the expected heterozygosity, H_o is the observed heterozygosity, N = the proportion of the sample represented by null homozygotes, $A = [H_e(1+N)-H_o]$ and $B = [4N(1-H_e^2)]$.

The decision over which estimator for n should be used depends on whether there are individuals within the population in which no bands have been observed. If so, then equation 1 should be applied. Alternatively, for populations where all individuals have given one or more bands, then equation 2 should be applied (Brookefield, 1996). Where all loci failed to amplify within an individual, it was designated as a PCR failure rather than a null allele homozygote.

6.2.4.2. *Temporal Variability*

Because many of the samples were collected in more than one year, it was important to assess whether there was any genetic structuring over time for the same collecting site (e.g. genetic drift, gene flow, extinction and recolonisation; Viard *et al.*, 1997; Lugon-Moulin *et al.*, 1999). If this is not the case, then samples from different generations at the same site can be pooled together to increase the sample size.

Therefore, temporal variation in allele frequencies for populations where samples were obtained in different years (sites: BH, HB, MC, RH, WHH, MD, M, FE, LY, R, DD, MH, W, A3, CH, PD) was investigated using the exact G-test advocated by Goudet *et al.*, (1996). When carried out on genic tables, this was found to be the most powerful test for genetic differentiation, particularly when sample sizes are unbalanced (Goudet *et al.*, 1996), as is the case here. A total of 5000 permutations of genotypes among samples were performed to assess the significance of the temporal differentiation. Computations were performed using *FSTAT* version 2.9.1. (Goudet, 2000) updated from Goudet (1995).

6.2.4.3. *Genetic evidence for a population bottleneck*

To test for evidence of a recent reduction in effective population size (N_e), the microsatellite data was subjected to analysis using the programme *BOTTLENECK* 1.2.02 (Cornuet & Luikart, 1996; Piry *et al.*, 1999). This programme works on the assumption that allele number (k) is generally reduced faster than observed heterozygosity (H_O) (Hedrick *et al.*, 1986), thus recently bottlenecked populations will display an excess of heterozygosity relative to allele number.

The program Bottleneck computes for each population sample and for each locus the distribution of the gene diversity expected from the observed number of alleles (k), given the sample size (N) under the assumption of mutation-drift equilibrium. This distribution is obtained through simulating the coalescent process of the loci number (5) under three possible mutation models, the IAM, the TPM and the SMM. This process enables the computation of the average H_{EXP}, which is compared to the observed gene diversity (H_{HW}, or Hardy-Weinberg heterozygosity; see Nei 1987) to establish whether there is a gene diversity excess or deficit at each locus. In addition, the standard deviation (SD) of the mutation-drift equilibrium distribution of the gene diversity (H_{EXP}) is used to compute the standardized difference for each locus (H_E) [$(H_{HW} - H_{EXP})/\mathrm{SD} = H_E$]. The distribution obtained through simulation also enables the computation of a p-value for the observed gene diversity (Cornuet & Luikart, 1996; Luikart & Cornuet, 1998; Luikart *et al.*, 1998; Luikart, 1999; Spencer *et al.*, 2000).

For the analysis, statistics for all three models of mutation were calculated. This included two-phase model of mutation (TPM), as recommended by Luikart *et al.*, (1998) to analyse microsatellite data. This model is based on a Stepwise Mutation Model (SMM) with 10% multi-step mutations. A Wilcoxon signed rank test was performed to test the hypothesis that the average standardised difference across loci for each population is not significantly different from zero (Spencer *et al.*, 2000). A second indication of a bottleneck is a shift away from an L-shaped distribution of allele frequencies, to one with fewer alleles in the low frequency categories. This was also assessed using the program *BOTTLENECK* (Cornuet & Luikart, 1996; Piry *et al.*, 1999).

6.2.4.4. *The effects of population size on genetic diversity*

Recent genetic theory predicts that large population sizes are necessary for maintaining genetic diversity, and that small populations are more prone to result in the random fixation of mildly deleterious mutations (Gilpin 1991; Lynch *et al.*, 1995; Lande, 1994). This speculation is supported by evidence for small isolated populations having an

increased risk of extinction (Saccherri *et* al., 1998; Nieminem *et al.*, 2001), but to date there is little evidence for their likelihood to have a lowering of genetic diversity.

Estimates of gene diversity (Nei, 1987) were made for each of the five loci and 26 populations using the programme *FSTAT* (version 2.9.1.), and a mean value across loci was calculated. This estimator is analogous to H_E, and will give a measure of population genetic diversity that is unbiased by sample size, the effects of null alleles, breeding system or the Wahlund effect *etc.* For both of these reasons it would be inappropriate to simply use the observed level of heterozygosity (H_O). The algorithm in *FSTAT* is based on the following equation (see also Nei, 1987: equation 7.39, p164):

$$H_{sk} = \frac{n_k}{n_k\text{-}1}\left(1 - \sum p_{ik}^2 - H_{ok}/2n_k\right)$$

Where n_k is the sample size, p_{ik} is the frequency of allele A_i in sample k and H_{ok} is the observed proportion of heterozygotes in sample k.

Gene diversity was analysed with respect to three potential explanatory factors: (a) Population size (assigned as small, medium or large), (b) Whether the population bottlenecked, was the result of a recent colonisation event, or had a stable history, (c) Isolation from other populations, measured as the number of other sampled populations located within 37.5km. The data were analysed in GLIM with normal errors (equivalent to a three-factor analysis of variance) (see Crawley 1993). In order to ensure correct assignment of population size and demographic change, the categorisations were based on a various unpublished population surveys from Butterfly Conservation and English Nature. Both these organisations have a network of volunteers that assess colony size each year by walking transects and although this results in a somewhat crude measurement, it is a suitable approach to roughly evaluate colony size and history.

6.2.4.5. *Genetic Distance Estimators*

Estimates of genetic distance were calculated from allele frequency data using three different methodologies. The Cavalli-Sforza and Edwards (1967) chord distance (D_C) was calculated using the software *MICROSAT* version 1.5b (Minch, 1997). This Euclidean measure is a purely geometric distance based measure, and as such takes no account of any particular mutation model.

The second genetic distance measure applied was θ, Weir and Cockerham's F_{ST} analogue (Weir and Cockerham, 1984), calculated using *FSTAT* version 2.9.1. Statistical significance was obtained using an exact G-test by randomising genotypes among samples. F_{ST} and many of its analogues, including θ, are based on the assumptions of the IAM, where a single mutational event can theoretically result in any other allelic state. This means that no inferences can be drawn to the alleles former state based on its current state. The algorithm in *FSTAT* is based on the following equation:

$$\hat{\theta} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2 + \sigma_w^2}$$

Where σ_a^2 is the among sample variance component, σ_b^2 is the within sample variance component, and σ_w^2 is the within individual variance component.

The third genetic distance measure was Slatkin's R_{ST} (Slatkin, 1995), calculated using *RST Calc.* Version 2.2 (Goodman, 1997). This is an analogue of F_{ST} that is based on the SMM rather than the IAM. This means that a single mutational event can only alter the allelic state by the removal or addition of a single repeat unit, consequently a pattern of descent can be assumed for each allele. R_{ST} can be defined as:

$$R_{ST} = \frac{(\overline{S} - S_W)}{\overline{S}}$$

Where S_W is the sum over all loci of twice the weighted mean of the within-population variances σ_a^2 and σ_b^2, and $\overline{S}$ is the sum over all loci of twice the variance $\sigma_{(a+b)}^2$ of the combined population.

It should be noted that due to the high number of null alleles detected within the data, it is possible that estimators based on heterozygosities, such as F_{ST} and R_{ST}, will not truly represent the populations.

6.2.4.6. *Phylogenetic Relationships between populations*

Phylogenetic reconstructions were carried out to assess evolutionary relationships between populations based on the microsatellite data. Neighbour-joining (NJ) trees were constructed using Cavalli-Sforza and Edwards chord distance D_C (Cavalli-Sforza & Edwards, 1967). This Euclidean measure was shown to be the most relevant method to calculate genetic distance between closely related groups where heterozygosities are high (Takezaki & Nei, 1996), as will be the case with population analyses with highly polymorphic markers such as microsatellites. In particular, this distance outperformed distances designed for the infinite allele and stepwise mutation model (Takezaki & Nei, 1996).

Genetic distances (D_C) were calculated from the microsatellite allele frequencies using *MICROSAT* (Minch, 1997), creating 1000 distance matrices from resampled datasets bootstrapped over loci. The resulting bootstrapped distance matrices from *MICROSAT* were then used as an input file for the NJ tree subroutine in the program *NEIGHBOUR* within *PHYLIP* version 3.57c (Felsenstein, 1995). The input order of populations was randomised to ensure that the final tree topology was not dependent on the order the samples were entered (option J in *NEIGHBOUR*). The *CONSENSE* subroutine within *PHYLIP* was then used to produce a consensus NJ tree that provided estimates of robustness at each node based on the bootstrapping of the gene frequencies.

6.2.4.7. *Mantel testing for isolation by distance*

In order to test for isolation by distance a mantel test was carried out (Mantel, 1967). A standard Mantel is a correlation between a matrix of geographical distances and a matrix of genetic distances (Manly, 1991 & 1997), providing the coefficient of correlation, *r*. A percentage *p*-value is derived by ranking the observed *r*-value against recalculated *r*-values following randomisation, by column or row, of one of the matrices.

The genetic distance measure used was Cavalli-Sforza & Edwards (1967) chord distance, calculated in the program *MICROSAT* (see section 6.2.4.6). Geographic distance was the straight-line distance between the centres of the various populations. The program used for the mantel analysis was a Pascal translation of the FORTRAN code found in Manly (1991). It is available free from J. Goudet (Université de Lausanne). Significance tests were performed over a total of 10,000 randomisations.

The relationship between genetic and geographic distance was further investigated using a regression analysis. The relationship between the two variables (genetic and geographic distance), r^2, was repeatedly recalculated after increasing numbers of pairwise chord distances were removed from the analysis, depending on the geographical distance between the two appropriate populations (increasing increments of 7.5km). For the first calculation, all pairwise chord distances were included. For the second all pairwise chord distances of at least 7.5km. For the third all pairwise comparisons of at least 15km; and so on, until a critical point was reached where no further significant relationship could be found between the two variables. This should give an indication of the critical geographic distance between populations after which gene flow no longer has a significant effect.

6.2.4.8. *Analysis of Molecular Variance (AMOVA)*

The data were subjected to a hierarchical analysis of molecular variance (AMOVA), as described by Excoffier *et al.* (1992), that provides estimates of variance components to reflect genetic diversity at different levels of a hierarchy. The variation between geographic regions was level one of the hierarchy, variation among populations within each of the geographic regions a second level, and the variation among individuals in a population as the third level. The analysis was performed using the program *ARLEQUIN* version 2.000 (Schneider *et al*, 2000). Significance values for the covariance components, σ_a^2, σ_b^2, σ_c^2 and for the fixation index (F_{ST} analogue) Φ_{ST}, were calculated using 10,000 permutations (of individual genotypes) to produce a null distribution to which the values were compared.

The populations were grouped into the following geographical structure (refer to figure 6.1 for the geographical location of each of these populations):

1) East Dorset: (BD, BH, HB, RH, WHH, MC, RiH)

2) West Dorset: (P)

3) The Isle of Wight: (M,B,V)

4) Surrey: (R)

5) South Wiltshire: (HA, MD, FO, PD)

6) North Wiltshire: (A3, BC, CL, CH)

7) Kent: (FE, LY, WYE)

8) West Sussex: (DD, MH)

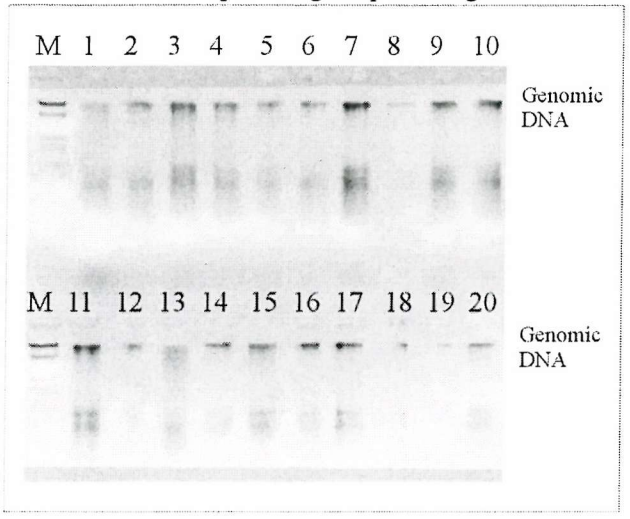
9) East Sussex (W)

6.3. RESULTS

6.3.1.1. *Sample Collection*

DNA was successfully extracted from all 1173 specimens using the salt extraction technique (figure 6.2).

Figure 6.2. The salt extracted DNA from a subset of twenty of the 1173 *P. bellargus* specimens. A clear band towards the top of the gel represents genomic DNA.



6.3.1.2. *Hardy-Weinberg Analysis*

Of the five microsatellites analysed, all showed significant deviations from the Hardy-Weinberg equilibrium in at least a subset of the 26 populations (75% of locus/population combinations) (appendix II), even when the total error rate was controlled by a sequential Bonferroni adjustment, 61 tests remained significant. Deviations were all heterozygote deficiencies, occurring at all five loci to differing extents. This was lowest for LbG2, where significance was identified in just six populations. For Lb4/18, there was significant deviation in nine populations, and for Lb4/19 there were eight populations showing a significant deviation. For Lb1/41 and Lb1/57, a significant deviation was initially found in all of the populations analysed, although this was reduced to 19 and 18, respectively, after Bonferroni correction.

Genetic data obtained from natural populations of outbreeding organisms usually indicate agreement of observed and expected genotype frequencies according to Hardy-Weinberg equilibrium (Nei, 1987). In the case of consistency across loci, heterozygote deficiencies may be ascribed to a variety of factors including inbreeding and population substructuring (Lade *et al.*, 1996; Paxton *et al.*, 1996) because, by these effects, all loci are affected simultaneously. Because this consistency is not present in this data, it is unlikely that these processes are the sole cause for the disequilibrium. Another explanation for the heterozygote deficiencies found is the possible presence of null alleles within the populations. This is where there is nucleotide sequence variation at the primer annealing sites. The fact that for each locus there was a subset of individuals that failed to amplify also tends to indicate the likelihood of null alleles. When the frequencies of null alleles within the data were calculated (see appendix III) and averaged across all sites, the frequencies were 0.2 ($\pm$0.025 standard error (SEM)) at locus Lb1/41, 0.18 ($\pm$0.026 SEM) at Lb1/57, 0.03 ($\pm$0.013 SEM) at LbG2, 0.09 ($\pm$0.023 SEM) at Lb4/18, and 0.07 ($\pm$0.022 SEM) for Lb4/19. Most populations that deviated from the Hardy-Weinberg equilibrium displayed frequencies of around 0.15 to 0.3, this is compliant with other published studies where null allele frequencies have been calculated (Neumann & Wetton, 1996; Van Treuren, 1998).

6.3.1.3. *Temporal Variability*

Out of the sixteen populations tested for temporal variation in allele frequencies, none showed any significant overall loci differentiation between generations (exact G-test, $p>0.05$). Therefore all further analyses were performed on the total number of individuals collected from a given site over the duration of the project.

6.3.1.4. *Genetic Evidence for a population bottleneck*

Out of the 26 populations analysed for evidence of a bottleneck, the two-phase mutation model (TPM) indicated that one population showed evidence of a heterozygote excess.

However, under the infinite allele model (IAM), estimates of heterozygosity excess were significant for nine populations. Under the stepwise mutation model (SMM), no populations indicated heterozygosity excess. None of the allele distributions showed a departure from the standard L-shape in the mode-shift test.

Table 6.1. Wilcoxon signed rank tests for excess heterozygosity for 26 *P. bellargus* populations across the UK. Details of numbers of loci exhibiting an excess of heterozygosity; calculated *p*-value; population histories (R = recolonised; B = bottlenecked: S = stable; ? = unknown); significance (p <0.05) are all shown.

POPULATION	% LOCI WITH AN EXCESS		*P*-VALUE		HISTORY	SIGNIFICANCE	
	IAM	TPM	IAM	TPM		IAM	TPM
A3	4	1	0.31	0.98	?		
BD	5	1	0.02	0.98	S	*	
BH	3	1	0.59	0.98	S		
BC	3	0	0.69	1.0	B		
B	3	1	0.41	0.97	S		
CL	5	5	0.02	0.02	R	*	*
CH	3	2	0.5	0.89	S		
DD	4	1	0.03	0.95	?	*	
FE	5	2	0.02	0.92	B	*	
FO	5	4	0.02	0.07	B	*	
HA	4	2	0.07	0.92	B		
HB	4	0	0.07	1.0	?		
LY	4	3	0.03	0.5	B	*	
MC	4	3	0.04	0.67	R	*	
MD	4	2	0.31	0.95	B		
MH	3	2	0.11	0.92	B		
M	3	1	0.41	0.97	S		
P	2	1	0.59	0.97	?		
PD	4	3	0.04	0.69	?	*	
R	4	2	0.07	0.89	S		
RIH	5	2	0.02	0.92	?	*	
RH	2	1	0.68	0.98	?		
V	3	2	0.31	0.95	S		
WHH	3	1	0.31	0.95	S		
W	4	2	0.11	0.59	?		
WYE	4	3	0.31	0.59	R		

Of the populations known to have undergone a founder event or population bottleneck, 60% showed a significant (p <0.05) excess of heterozygosity under the IAM, indicating the occurrence of a recent population bottleneck. Under the TPM, this is reduced to 10%. Of the populations where no bottleneck has been recorded, 12.5% showed a significant heterozygosity excess under the IAM, with none under the TPM (see table 6.1).

However, after the application of a Bonferroni correction, the critical *p*-value is reduced to, 0.002, thus resulting in no significant detection of heterozygosity excess.

6.3.1.5. *Population parameters and their effects of on genetic diversity*

The ANOVA analysis of population gene diversity with respect to various explanatory factors indicated that there was no significant effect on gene diversity from either observed demographic population change or the effects of isolation. However, a significant effect was found for population size (see table 6.2).

Table 6.2. Summarisation of the ANOVA results.

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	*F*
Population size	2	0.02121	0.011	6.34**
Demographic history	2	0.00205	0.00102	0.563
Isolation	1	0.00021	0.00207	0.114
Interaction within samples	23	0.0385	0.00167	

** significant at $p<0.01$

The ANOVA results infer a highly significant effect of population size on genetic diversity of the population, whereby reductions in genetic diversity correspond with small population size. The mean values for population gene diversity in each of the three categories were: small: 0.64 (Standard Deviation (SD) = 0.02), medium: 0.67 (SD: 0.01), and large: 0.72 (SD (0.01).

6.3.1.6. *Genetic Distance estimators*

Alongside Cavalli-Sforza and Edwards (1967) chord distance, pairwise population values for θ, (Weir and Cockerhams F_{ST} analogue (Weir and Cockerham, 1984)) and R_{ST} were calculated for the data, alongside their associated *p*-values (derived from 30,000 permutations of the data). After a Bonferroni correction, the critical *p-value* was adjusted to 0.000154. The average pairwise population F_{ST} value was 0.118, whilst the lowest

value recorded (0.0109, $p = 0.0087$) was found among the Isle of Wight populations (B and M) and the highest recorded value, 0.286 ($p = 0.00003$), was found between Kent (Wye) and Wiltshire (BC). The average pairwise population R_{ST} value was 0.142, whilst the highest (0.441, $p = 0.316$) was found between Kent (WYE) and Wiltshire (A3), the lowest was found between Sussex (MH) and Dorset (RiH) (-0.0168; effectively an R_{ST} value of 0.00, $p = 0.001$). The sample size from RiH was particularly small (<16 individuals), possibly encouraging biases in the results. If all sample sizes less than 20 are discounted from the analysis, then the lowest R_{ST} value is –0.005 (effectively 0.00, $p =$ 0.81), found among the west Dorset samples (MC and WHH).

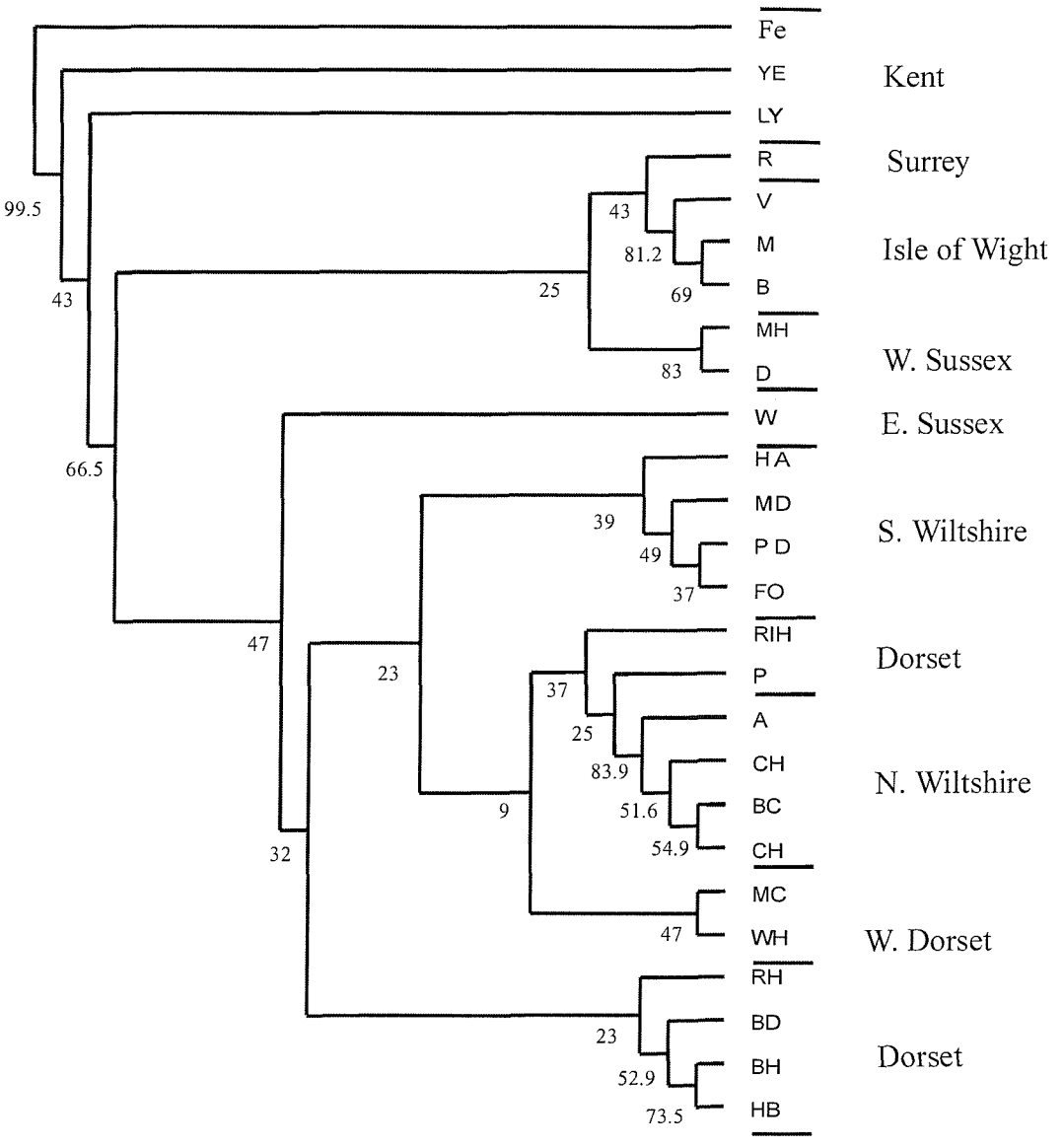
Approximately 92% of the F_{ST} comparisons and 72% of the R_{ST} comparisons were deemed to be significant following a Bonferroni correction (Sokal & Rohlf, 1995). Of the F_{ST} values from comparisons between geographic regions that were separated by expanses of unsuitable habitat, 94.8% of the pairwise comparisons were significant, as opposed to 71.1% in comparisons within a geographic region. The equivalent values for R_{ST} were 74.9% and 52.7% respectively. The average F_{ST} value for comparisons across geographic regions separated by unsuitable habitat was 0.128, whereas the average value for those within a geographic area was 0.045. The equivalent values for R_{ST} were 0.152 and 0.065 respectively.

6.3.1.7. *Phylogenetic relationships between populations*

The phylogenetic reconstructions, using Cavalli-Sforza and Edwards (1967) chord distance, resulted in a consensus Neighbour Joining (NJ) tree that was only weakly supported at a large number of nodes (<50%). However, the general trends apparent in this dendogram (figure 6.3) comply well with the geographical distribution of the populations sampled. Populations from within a geographic region were generally linked within the same clustering on the NJ tree, although a number of these were rather tenuous. There were exceptions to this though; for example the positioning of R, a population from Surrey, in the Isle of Wight clustering seems a highly improbable

grouping. Similarly, two populations from southern Dorset, RiH and P are grouped with populations from northern Wiltshire.

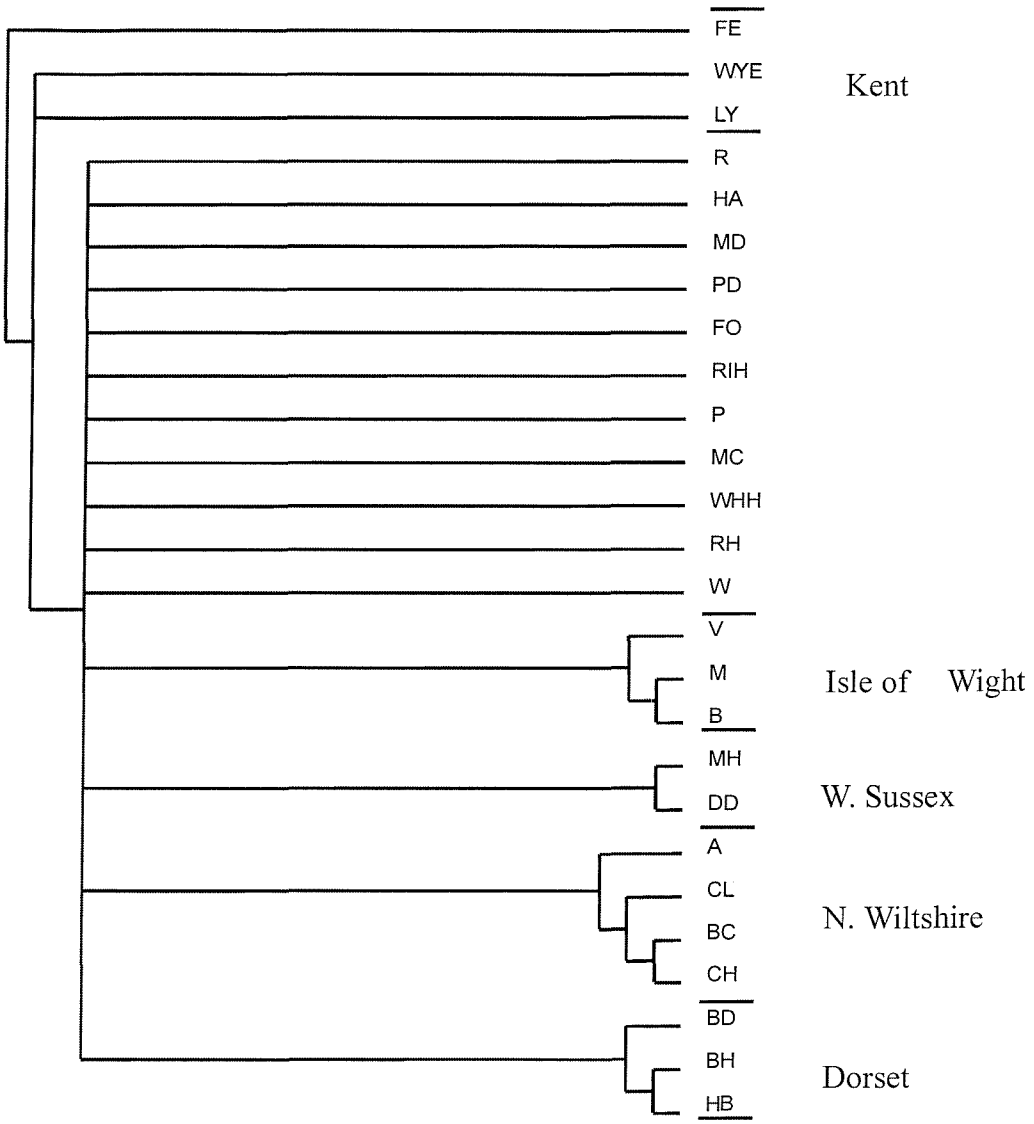
Figure 6.3. Consensus tree based on Cavalli-Sforza and Edwards (1967) chord distances derived from allele frequencies at five microsatellite loci (Neighbour-joining method of tree construction). Bootstrap values are indicated at each node.



After the tree was made more robust, by collapsing nodes supported by less than 50% of bootstraps (see figure 6.4), all unexpected relationships were broken down and five major groupings of populations were revealed. These were supported by high bootstrap values

(>80%; figure 6.3). The first assemblage consists of the three populations from the Isle of Wight (bootstrap value: 81.2%), M and B being joined in 69% of the resampled trees.

Figure 6.4. Neighbour joining consensus tree based on Cavalli-Sforza and Edwards (1967) chord distance. Branches not supported by a bootstrap value of at least 50% have been collapsed.



All of these populations are in close proximity to one another (<30 kilometres), particularly M and B, which are separated by less than 10 kilometres. The two populations sampled from Sussex also group together very tightly (bootstrap value: 83%); these populations are also separated by no more than 20 kilometres. Four of the populations from Wiltshire also form a distinct cluster (A3, BC, CH and CL), supported

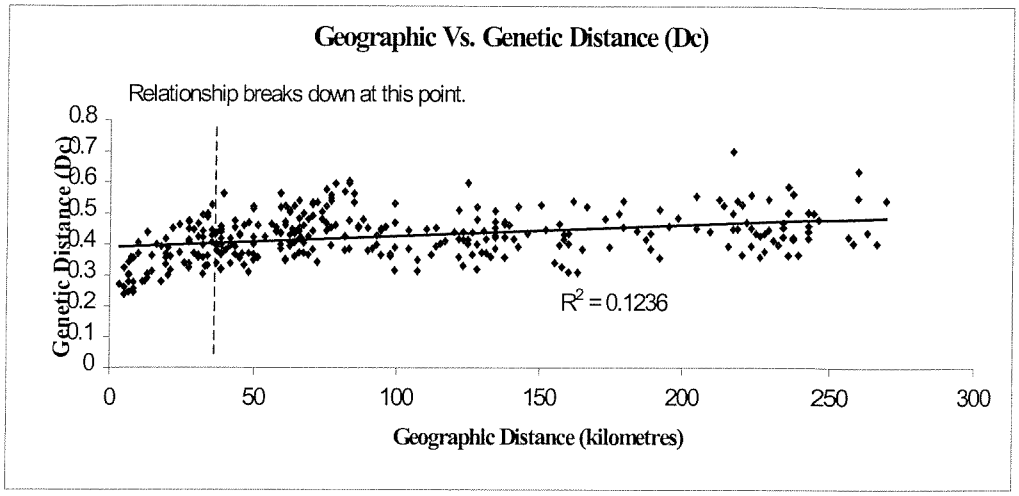
by 83.9% of bootstraps; also grouping where the populations are all geographically close to one another. All three of the Kent populations (LY, WYE and FE) also form a cluster, with FE and LY grouping together more tightly (99.5% of bootstraps); this group of three populations is extremely isolated from other populations within the UK. Three of the populations in Dorset (BD, BH and HB) are weakly clustered together, supported by 52.9% of the bootstraps. One other notable grouping is among the southern Wiltshire populations (MD, P, and FO) and although is broken down, is grouped together in almost 50% of the bootstraps (49%). The relationships between the remaining samples on the dendogram remain poorly resolved, showing low bootstrap support (<50%). This may be attributable to the fact that many of these populations are not in the immediate vicinity of any other population included in this study.

6.3.1.8. *Mantel Testing for isolation by distance*

In order to establish whether any isolation by distance effect has occurred, matrices of genetic distance data (Cavalli-Sforza and Edwards (1967) chord distance, D_c) and the logarithms of geographical distance data (in kilometres) between all the UK sample sites were constructed. These matrices were analysed for their degree of correlation using a Mantel test (figure 6.5) (Mantel, 1967; Manly, 1991 & 1997).

The Mantel test results gave a statistically significant *r-value* of 0.3516 ($p = 0.0007$, for 20,000 randomisations), indicating that there is a strong isolation by distance effect present among the populations. However, this relationship is clearly not of a linear nature (see figure 6.5 below) and when investigated further, by fitting various models under a regression analysis, a curved line showed a more robust fit to the data. When a logarithmic line is fitted to the data, the r^2 increases to a value of 0.231 ($p < 0.0001$).

Figure 6.5. Cavalli-Sforza and Edwards (1967) chord distance vs. total geographic distance, for all pairs of sites. Value for r^2 is shown, as is the point where the relationship between the two variables becomes non-significant.



After pairwise population indices representing sites separated by geographic distances less than 22.5km were removed from the analysis, the relationship between genetic and geographic distance became increasingly unstable (see table 6.3). This reached a critical point when all populations separated by less than 37.5km were removed from the analysis, where no further significant relationship could be found (p <0.05) (see table 6.3). This seems to indicate that a significant amount of gene flow between populations is only occurring where populations are separated by less than 37.5km. However, it should be borne in mind that by removing pairwise values from the Mantel test, the power of the analysis may be reduced, and it is possible that this effect causes the eventual non-significant result for populations separated by less than 37.5km.

Table 6.3. The calculated r^2 values, associated probability values and significance rating (p <0.05) for a regression analysis of the data, after successive removal of pairwise populations dependant on the geographic distance between them.

DISTANCE BETWEEN POPULATION PAIRS (KM)	R^2	P-VALUE	SIGNIFICANCE ($P < 0.05$)
>0	0.124	0.0007	***
>7.5	0.073	<0.0000	***
>15	0.054	<0.0000	***
>22.5	0.035	0.0013	**
>30	0.024	0.011	*
>37.5	0.015	0.0613	NS

6.3.1.9. *Analysis of Molecular Variance (AMOVA)*

The data was analysed by an AMOVA (Excoffier *et al.*, 1992) to investigate the hierarchical structuring of the UK populations. Each population was designated as belonging to a specific geographic region, defined as a discrete area containing an isolated subset of the UK populations. The analysis examined variance components among these geographic regions, plus among the populations within each of these geographic regions, and also among individuals within each population. Over eighty-seven percent of the variance was explained by within population variation, (see table 6.4), the remainder is partitioned between (i) variation among populations within regions (4.5%), and (ii) variation among regions (8.18%). There is almost twice the amount of variance between regions as there is among populations within them, an indicator that there is limited gene flow between regions, and hence there is population differentiation. The assigned value for the fixation index Φ_{ST} (F_{ST} analogue) is 0.127. All variance components were found to be significantly different from zero, when compared against a calculated null distribution ($p = 0.05$) (see table 6.4).

Table 6.4. The calculated AMOVA values for the variance components and hierarchical assignment of variation.

ASSIGNMENT OF VARIATION	**VARIANCE COMPONENT**	**Variance %**	***P*-VALUE (AGAINST NULL DISTRIBUTION)**
Among geographic Regions	0.122Va	8.18	<0.0001
Among populations within regions	0.067Vb	4.5	<0.0001
Among individuals within populations	1.299Vc	87.32	-

When the populations were analysed phylogenetically to produce a Neighbour-Joining tree (figure 6.4, section 6.3.1.7), the results indicated that the Kent populations strongly grouped away from the other UK populations. In order to substantiate whether this grouping was causing a bias in the results of the AMOVA, the analysis was repeated with all three Kent populations (FE, LY and WYE) removed from the data. The resulting analysis demonstrated very little change compared to the original analysis that included

the Kent populations. The results showed a small increase in the "among regions" variance component, from 8.18% to 8.81% and a slight reduction in the "among populations within regions" variance component, from 4.5% to 4.04%. All values remained significant (p <0.0001).

In order to substantiate that the populations "R" "P" and "W" being designated as a geographical grouping of their own did not seriously bias the "among regions" variation, the analysis was performed after "P" & "W" were grouped into the next most appropriate geographical region ("P" with the east Dorset group and "W" with West Sussex group) and with "R" removed from the analysis (there was no other obvious choice for a geographical grouping). The results showed a small increase in the "among populations within regions" variance component, from 4.5% to 4.84% and a slight reduction in the "among regions" variance component, from 8.18% to 7.54%. However, all values remained highly significant (p <0.001).

6.4. DISCUSSION

6.4.1. The indication of excess homozygosity and null alleles

The development of microsatellite markers in butterfly species has consistently been found to be more difficult than in many other species (Meglecz & Solignac, 1998; Keyghobadi *et al.*, 1999). This difficulty may result from their low frequency within the genome (see chapter five), or some other factor that impedes amplification. Observations from other similar studies (Palo *et al.*, 1995; Meglecz *et* al., 1998; Meglecz & Solignac, 1998; Keyghobadi *et al.*, 1999), where significant homozygote excesses are consistently found, suggest a high mutation rate in the regions immediately surrounding microsatellite repeats in butterflies, resulting in large numbers of null alleles. The effects of population structuring alone cannot explain the large deviations from Hardy-Weinberg proportions found in all of these studies (Palo *et al.*, 1995). Similar observations are found in the microsatellite data from *P. bellargus*, where all five loci revealed homozygote excesses to varying extents. In some cases, the Hardy-Weinberg proportions deviated so greatly that

they could only be attributed to elevated levels of null alleles. For example, at Lb1/41 in the Isle of Wight populations (B, M, and V), the observed differences between the observed (H_O) and expected (H_E) heterozygosities almost reached an order of magnitude. However, when null allele frequencies were calculated for all loci and population combinations presenting significant homozygote excesses, most were found to have frequencies comparable to those found in other published data (Neumann & Wetton, 1996; Van Treuren, 1998; Keyghobadi *et al.*, 1999), although in a few instances these levels were slightly higher (for example at Lb1/41 in the Isle of Wight populations).

This widespread presence of null alleles within the dataset presents a number of analytical challenges. For example, many of the genetic distance estimators applied to microsatellite data, such as F_{ST} and R_{ST}, are calculated from population heterozygosity values, thus the presence of null alleles will tend to confound the calculation of these indices. In this situation, the application of a Euclidean genetic distance estimator, such as Cavalli-Sforza and Edwards (1967) chord distance (D_C) or Nei *et al*'s (1983) D_A, may be more appropriate. This should eliminate biases within the data, resulting from erroneous calculations of heterozygosity. As a result of this, I have not relied on F_{ST} based indices in this study, and where possible applied a Euclidean distance measure in their place. Aside from their benefits to datasets exhibiting null alleles, these measures (D_C and D_A) have both been shown to be more efficient at reconstructing evolutionary relationships using microsatellite data (Takezaki & Nei, 1996; Lugon-Moulin *et al.*, 1999).

6.4.2. Evidence for recent genetic bottlenecks and inbreeding

The application of tests to detect recent reductions in effective population size (genetic bottlenecks) from selectively neutral marker data should indicate which populations have had a recent decline. This can then be compared to the chronicled historical data available for many *P. bellargus* populations. The results showed that there are still an appreciable number of rare alleles in all of the populations examined, because none showed a significant mode-shift in allele frequencies as is observed in other bottlenecked

populations (Hartl and Hell, 1994; Houlden *et al.*, 1996). Other effects of a bottleneck, defined as an excess of heterozygosity, were detected under the IAM in 60% of the populations where documented evidence suggested a recent decline, whereas only 12.5% of populations documented as having remained stable displayed these effects. However, only 10% of declined populations displayed evidence for bottlenecks under the TPM and none under the SMM. If this data is submitted to a Bonferroni correction, then none of the tests remain significant. However, as discussed by Cabin & Mitchell (2000) the blanket application of this kind of statistical correction will needlessly lead to increased type II error rates, therefore I have presented both corrected and uncorrected results. The argument that repeated tests on the hypothesis and dataset causes increased type I errors is in itself rather a arbitrary distinction, because by this logic the test could be applied to all of the many tests throughout this thesis where each runs the same 1 in 20 chance of a type I error.

There are a variety of possible explanations for why the bottleneck is not detected under either the SMM or TPM. The presence of null alleles in the population may have a confounding effect on the analysis (Cornuet & Luikart, 1996; Le Page *et al.*, 2000), or the disjunctive patterns of allele size distributions present in all populations may indicate a departure from the SMM, possibly making this test uninformative. Alternatively, five loci may be too few to allow the test to achieve sufficient analytical power (Cornuet & Luikart, 1996). Also, the historical data is somewhat anecdotal, as a result of having not been collected under any methodological consensus by amateur entomologists, hence may be inaccurate. Nevertheless, the test appears to at least show trends under the IAM, and the implications are that the effects of declines are still traceable in the majority of putative bottlenecked populations.

When gene diversity values were calculated from the microsatellite data and analysed with respect to various aspects of the chronicled population data, no significant effects from population isolation or declines and extinction/recolonisations were found. This is somewhat surprising, because theoretical predictions suggest that these kinds of demographic changes will strongly affect the genetics of remnant populations (Wright,

1931; Gilpin, 1991; Lande, 1995; Lynche *et al.*, 1995; Brookes *et al.*, 1997), and studies of wild populations have also indicated an increased likelihood for isolated butterfly metapopulations to have an increased extinction rate (Saccheri *et al*, 1998; Nieminem *et al.*, 2001).

However, when genetic diversity was analysed with respect to approximate population size, a highly significant effect was uncovered. The results clearly demonstrated that genetic diversity was influenced by population size, whereby smaller populations appear to be predisposed to inbreeding (as inferred from the heterozygosity levels (Frankham, 1995a)). This effect has only been hinted at in other studies, for example in the silver studded blue butterfly (*Plebejus argus*), where allozyme and mtDNA data implied the loss of diversity from small populations with rapid turnover rates and also for *Maculinea teleius*, where a relationship was also inferred between small effective population size and low heterozygosity (Figurny-Puchalska, 2000). Not only this, but many endangered populations and species have been demonstrated to show reduced diversity (Frankham, 1995a). However, whilst it may seem obvious that there will be a relationship between population size and diversity, whereby the smaller number of individuals increases the likelihood of mating among relatives (inbreeding), very few studies to date have conclusively found such a relationship between these variables within a species.

Gene diversity (e.g. H_{sk} (section 6.2.4.4) gives a measure of population expected heterozygosity (hence also homozygosity), which will also be related to levels of inbreeding (Frankham, 1995a). This is because observed heterozygosity is predicted to decrease with increased mating among relatives (Wright, 1950). Therefore, the observation that reduced gene diversity is associated with small population size, rather than isolation, may indicate that conservation efforts should concentrate more on the size of suitable habitats than their connectivity to neighbouring sites.

6.4.3. Evidence for population sub-structuring, gene flow, and isolation by distance

Although, as already discussed, F_{ST} and R_{ST} may be subject to the confounding effects of null alleles, pairwise comparisons using these indices did provide evidence for geographic sub-structuring among *P. bellargus* populations. Over 90% of F_{ST}, and 70% of R_{ST}, pairwise comparisons across geographic regions indicated significant differences, and the average pairwise values for F_{ST} and R_{ST} were 0.128 and 0.152 respectively. For both estimators, the highest recorded values were found between the Kent and Wiltshire populations, a comparison representing the furthest separation of regions. Correspondingly, the lowest values for both estimators were found within regions, among two Isle of Wight populations (for F_{ST}) and two Dorset populations (for R_{ST}), and the average values for F_{ST} and R_{ST} within geographic regions were 0.045 and 0.065 respectively. Overall, these values suggest that there are barriers to gene flow, particularly between populations separated by unsuitable habitat.

The indication of geographical structuring from the F_{ST} and R_{ST} values is also well supported by the dendograms, Mantel and AMOVA analyses. These all indicate a pattern of connectivity among populations that relates to the geographic distance separating them, in accordance with Wright's (1943) model of isolation by distance. Whilst the initial phylogram reconstruction resulted in a topology that reflected the geographic distribution of populations, in a few places it was only weakly supported by the bootstrap data. When this tree was made more robust, by collapsing the nodes with low bootstrap values, many of the relationships broke down, leaving only a few very strong population clusters. These groupings consistently represented 'local' populations, separated by short geographic distances (<20 kilometres). This suggestion of 'isolation by distance' appeared to be confirmed by the Mantel analysis, where genetic and geographic distances were significantly correlated. After the systematic removal of populations separated by less than 37.5 kilometres from the analysis, this relationship appeared to break down, possibly indicating that only populations separated by short distances are related to one another, probably representing intra-regional groupings.

The tendency for populations within a geographic region to be more closely related to one another, than populations among geographic regions, was further illustrated by the AMOVA analysis. Here twice the amount of population variance was identified between regions than was present among populations within a region. Both of these variance components were highly significant, as was the assigned fixation index, clearly suggesting significant genetic structuring among populations. This overall estimate of F_{ST} (0.127) corresponds to the genetic estimates achieved from other lepidopteran species via allozyme electrophoresis, such as the sedentary fritillaries *Euphydryas chalcedona* $F_{ST} = 0.090$ and *E. editha* $F_{ST} = 0.120$ (McKechnie *et al.*, 1975), *Melitaea didyma* $F_{ST} = 0.10$ (Johannsen, 1996), the mountain species *Oeneis chryxa* $F_{ST} = 0.081$ (Porter & Shapiro, 1991), Polish populations of *Maculinea nausithaus* $F_{ST} = 0.153$ (Figurny-Puchalska, *et al.*, 2000) and French populations of *Parnassius mnemosyne* $F_{ST} = 0.135$ (Napolitano & Descimon, 1994). Other studies using microsatellites have also demonstrated similar values: Hungarian populations of *P. mnemosyne* $F_{ST} = 0.070$ (Meglecz *et* al., 1998) and *Militaea cinxia* $F_{ST} = 0.2$ (Palo *et al.*, 1995). Population F_{ST} values derived from microsatellite and allozyme data have been demonstrated to be comparable, because of the similarity of values assigned by each method using the same sample sets (Meglecz *et* al., 1998). This is because F_{ST} measures the differentiation among (sub)populations relative to the limiting amount under complete fixation (Wright, 1978), and is therefore unlikely to be affected by the increased heterozygosity associated with microsatellites.

All of the analytical approaches to the microsatellite data have unequivocally implicated geographic distance as an important 'barrier' to gene flow, contributing to the genetic structuring among populations. Thus populations that are in close proximity to one another, or linked by stepping-stone populations, are typically more genetically similar to one another than populations from separate, unlinked geographic regions. Such genetic differentiation generally results from stochastic events, such as genetic drift and local selection, exacerbated by the diminished exchange of individuals among populations (gene flow) (Wright, 1931; Leberg, 1991). This implication of reduced regional gene flow corresponds with the ecology of *P. bellargus*, where mark-release-recapture studies

have inferred that the exchange of individuals among their discrete colonies is extremely rare (Thomas, 1983a). Even in Dorset, where populations are in close proximity to one another, no direct transfer of individuals was observed, and although 2-8% of adults were identified outside their main flight areas, they were generally within 25 metres and on the same side of any putative barriers (such as unsuitable habitat) (Thomas, 1983a). This kind of closed population structure, taken alongside the butterfly's specific requirement for south-facing calcareous grassland, results in clusters of colonies that are often separated from others by large distances, inhibiting dispersal of the butterflies. This may explain the sharp increase in genetic differentiation across short geographic scales, beyond which no relationship remains between the two variables. The slight contrast in the inferences from the dispersal studies and microsatellite data, where no transfer of individuals was observed, but short distance gene flow was indicated, can be attributed to the failure of most dispersal studies to detect either rare but consistent long-distance dispersal or infrequent mass dispersal events (Slatkin, 1987). It has been postulated that even as few as one migrant every other generation can be enough to prevent genetic drift between populations (Slatkin, 1995), events which are extremely unlikely to be directly observed. The fact that the butterfly has been able to recolonise sites in Dorset that are 10-15km from the nearest population certainly indicates a higher dispersal ability that suggested by Thomas (1983a).

In a review of gene flow and isolation by distance in a variety of other insects, similar trends have been identified for many sedentary species, where isolation by distance relationships become weak for distances greater than tens of kilometres, as a result of limited gene flow allowing nearly all populations to diverge (Peterson & Denno, 1998). Similar effects were found in *Melitaea didyma*, where regional habitat partitioning prevented gene flow among regions, whilst local population exchange remained high (Johannesen *et al.*, 1996). More complex patterns have been found in other lepidopteran species, for example a sedentary lycaenid butterfly, *Euphilotes enoptes*, where gene flow appeared to be occurring over spatial scales much greater than possible for dispersing individuals; a discrepancy the author attributed to stepping-stone gene flow, via intermediate populations (Peterson, 1996). This observation has distinct parallels with

the data for *P. bellargus*, where gene flow has been inferred over distances of up to 37.5km, yet the butterfly has never been observed to stray much more than 25 metres from its home range (Thomas, 1983a). It is likely that stepping-stone gene flow is also playing a part here, but that the structure of the habitat in the UK limits this to ~40km, because even stepping-stone gene flow cannot traverse areas where no populations are present.

Although the Mantel and AMOVA analyses revealed regional structuring of populations, the dendogram analysis was less conclusive. The Neighbour Joining tree indicated only weak relationships among some of the local populations, even where short distances were involved. For example, FO and HA, are only linked by 39% of the bootstraps, even though the two populations are separated by less than 5 miles. This lack of accordance between the genetic and geographical data may be attributable to the susceptibility of *P. bellargus* to declines and founder events, demographic changes that encourage rapid genetic drift to increase among-population genetic variance (Wright, 1940; Slatkin, 1977; Wade & McCauley, 1988; Whitlock & McCauley, 1990). Significantly, HA is recorded as having undergone a severe decline in the late 1970's, whilst FO has been recolonised since extinction around 1978-1981 (J. Thomas, personal communication), perhaps explaining the poorly resolved relationship.

An emerging pattern from this study, and a variety of others, is that genetic isolation by distance is typically evident in sedentary insects, though often at different geographic scales. It would appear that gene flow can often traverse distances much greater than indicated by dispersal studies, and consequently population isolation is frequently lower than would be expected. However, this observation must not encourage complacency towards the conservation of sedentary species, such as *P. bellargus*, because the critical factor in this stepping-stone gene flow is a connective network of populations in close proximity to one another. The effect of population isolation is already evident at the higher geographic scale, where regional populations have diverged from one another, and this effect is also creeping in at the more local scale.

Whilst the numbers of *P. bellargus* populations have increased since the decline of 1978-1981, the areas of suitable habitat have frequently shrunk, resulting in smaller population sizes. The work in this thesis has indicated population size as an important factor in preventing inbreeding, which in turn has already been shown to have detrimental effects on population survival (Saccheri *et al*, 1998; Niemeinem *et al.*, 2001). Therefore, if the fragmentation of calcareous grassland continues and grazing levels decline, then populations of *P. bellargus* will become progressively smaller and more isolated from one another, leading to the eventual cessation of population interconnectivity, increased inbreeding and an increased risk of colony extinction.

Whilst this work has obvious relevance to the conservation of *P. bellargus*, the results are also relevant to other arthropods, particularly sedentary species with a closed population structure. A large amount of conservation effort has been concentrated on the maintenance of corridors between suitable habitats and population connectivity, but clearly the preservation of habitats able to sustain larger populations is also an important consideration. Therefore rather than allowing areas that are rich in species biodiversity to become small patches, perhaps connected by corridors, it would be prudent to ensure that large habitat areas are also safeguarded.

7.0 LEPIDOPTERAN MUSEUM SPECIMENS AS A SOURCE OF DNA

7.1. INTRODUCTION

One of the aims of this thesis was to assess the current level of genetic variation present in *Polyommatus bellargus* populations, and compare this to the variation present in museum specimens of the species collected prior to the genetic bottleneck of 1978. Because only a limited amount of work has previously been carried out to determine the efficiency of preserved lepidopteran material as a source of DNA (e.g. Hammond *et al.*, 1996; Zakharov *et al.*, 2000), it was necessary to assess the quality of the DNA extracted, particularly with reference to its efficacy in PCR based experiments. It was integral to this aim to identify what method should be used to extract the DNA from the specimens and which part of the butterfly would yield the highest quality DNA. When considering which body part was to be used, it was important to keep in mind that unobtrusive methods must be applied, because the specimens used will be from entomological collections. This chapter details the technical work carried out to optimise the approach taken for this work.

Entomological collections have always been particularly valued as a source for study of morphological variation on which most classical taxonomy has been based. In particular, butterflies have always been popular with collectors. Consequently there are lepidopteran collections around the world that amount to hundreds, even thousands of individual butterfly specimens. Moreover, detailed data are usually available for each butterfly pertaining to the location and date of its capture. For example, the London Museum of Natural History has tens of thousands of individual stored butterflies, the majority of which are accompanied by specific details of their capture. With the advent of increasingly sophisticated molecular techniques, particularly those based on the polymerase chain reaction (PCR) (Saiki *et al.*, 1988), renewed importance and relevance has been brought to this vast resource of preserved biological material. This has exciting implications for phylogenetic studies of extinct species, for species whose sampling is difficult or near impossible in the wild, and crucially it also opens up a whole new range

of possibilities for studying temporal genetic changes in populations. To date, only preserved lepidopteran specimens aged ten years, or less, have been used for this kind of molecular analysis, yet the vast majority of museum specimens are much older than this. In order to effectively use these specimens for molecular studies, preliminary work is needed to assess the suitability of long-term stored lepidopteran material.

In accordance with other studies using preserved animal specimens from a wide range of vertebrate and invertebrate species (Paabo, 1989; Thomas *et al.*, 1989; Bulat & Zhakarov, 1992; Desalle, 1994; Lister, 1994; Taylor *et* al., 1994; Roy *et al.*, 1994; Hammond *et al.*, 1996; Bouzat *et al.*, 1998a; Groombridge *et al*, 2000; Zakharov *et al.*, 2000), it should be possible to use DNA from specimens as old as 100 years, and much more, provided the choice of molecular marker takes into account that the DNA will have undergone oxidative damage and will be extremely fragmented (in the range of a few hundred base pairs) (Thomas *et al.*, 1989; Ellegren, 1991; Cano & Poinar, 1993; Mitton, 1994; Hammond *et al.*, 1996). This means that a PCR-based approach will be essential. Microsatellites are likely to be the most useful type of marker in this situation, because of their small size (100-300bp), high variability, and ability to elucidate detailed genetic data. Short regions of mitochondrial DNA will be useful because they can be highly variable and because of the high copy number of the genome in each cell, the likelihood of undamaged appropriate regions of template DNA is increased. Most population studies use these two techniques in parallel.

A problem with using old insect collections may be the willingness of the owner to allow specimens to be sampled. The delicacy and diminutive size of most insect specimens means that the balance between minimising damage and obtaining enough DNA from the specimen is very fine. However, if enough DNA for PCR amplification can be extracted from legs or antennae, then archive material, including type specimens, can be tested in a minimally destructive way. Hopefully this would encourage curators and amateur collectors to be more willing to allow their specimens to be used in scientific study. Another consideration that is of paramount importance when using old stored samples is that they may be contaminated with DNA from micro-organisms, mites, or necrophagous

beetles. Such contaminants in pilot studies employing universal primers could seriously affect PCR results by giving spurious PCR products (Lister 1994). The chances of this kind of contamination affecting studies using microsatellites are somewhat lower because these markers tend to be highly species specific.

During the work detailed in this chapter, I will investigate the potential for applying molecular markers to obtain genetic data from preserved lepidopteran specimens aged from 6 months to 105 years. To achieve this aim, a number of individual parameters were individually investigated, these include:

(i) Identify the most suitable extraction method for obtaining DNA from preserved specimens. Both crude and high quality techniques will be investigated

(ii) Identify the most appropriate body part of lepidopteran specimens for molecular analysis. Attempts will be made to find a balance between minimum damage to the specimen and optimum amounts of DNA.

(iii) Investigate the effect of specimen age on the quality of DNA that can be extracted (i.e. what level of degradation it has been exposed to).

(iv) Investigate the effect of specimen age on the efficacy of PCR amplification.

(v) Investigate how amplicon size affects the efficacy of PCR amplification.

(vi) Assess the potential for preserved specimens as a source for isolating DNA markers.

7.2. MATERIALS & METHODS

7.2.1. Sample collection

Tissue samples were collected from fresh killed butterflies and a selection of preserved lepidopteran specimens of different ages, none of which showed any obvious degradative damage. These were obtained from a variety of sources:

Fresh Samples: Specimens of speckled wood butterflies (*Pararge aegeria*) were cultured in Southampton in Light:Dark 16:8, 18°C. Twenty-five of these were used as a source of fresh material for DNA extraction (heads, legs, thorax, antennae and wings) against which stored samples could be compared.

Stored Samples: Tissue samples (heads, legs, thorax, antennae and wings) were obtained from preserved lepidopteran specimens ranging from one month to 105 years old (see table 7.1). Because of the destructive nature of this research, specimens were obtained from sources where it was impossible to ensure a single species across an age range, therefore a variety of species were used.

Table 7.1. Details of the biological material used for this study.

SPECIES	AGE OF SPECIMEN	LOCALITY OF SAMPLING
Pararge aegeria	Fresh killed	Southampton Common
Pararge aegeria	One month	Southampton Common
Pararge aegeria	Six months	Southampton Common
Pararge aegeria	2.5 Years	Madeira
Callimorpha dominula	7 Years	Oxford
Pararge aegeria	11 Years	Madeira
Peribatodes rhomboidaria	103 years	Unknown
Polyommatus bellargus	105 Years	Folkestone.

7.2.2. Assessment of the efficiency of various DNA extraction techniques

A pilot study was undertaken to investigate which DNA extraction technique could give the most efficient results from stored lepidopteran material. Five different crude and high

quality extraction techniques were tested for the isolation of total genomic DNA from each specimen. Four of these methods were assessed on wing fragments from some of the oldest preserved samples (103 year old *P. rhomboidaria*) and also with several more recently stored samples (2.5 year old *P. aegeria*). The methods applied were: salting out (Sunnucks and Hales, 1996; Aljanabi & Martinez, 1997), guanidium thiocyanate (Hammond *et al.*, 1996), TE release method (Rose *et al.*, 1994), and a Chelex® 100 (Biorad, UK) release based extraction (Cano *et al.*, 1993). A crude extraction (direct PCR) (Grevelding *et al.*, 1996) was also used, however, this was only applied to legs removed from six-month-old *P. aegeria*. Once extracted, the DNA was assessed qualitatively by visualising the DNA after gel electrophoresis and quantitatively by directly measuring the DNA concentration.

7.2.2.1. *Crude extractions*

A. Direct-PCR (after Grevelding *et al.*, 1996).
Eppendorfs containing leg samples in 1x*Taq*-polymerase reaction buffer (Appligene-Oncor, UK) (10.8µl of H_2O:1.2µl of 10x PCR reaction buffer, respectively) were placed in a water bath at 100°C for 5 minutes prior to use in PCR amplification.

B. TE extractions (after Rose *et al.*, 1994):
Samples were gently homogenised in 40µl of TE (Tris/HCl 10mM, pH8.0; EDTA 1mM) using UV-cross linked pellet mixers (Scotlab, UK), and left at 4 °C for ten days to allow release of the DNA from the tissue into solution. The wing tissue was not removed from the mixture.

C. Chelex release method (after, Cano *et al.*, 1993):
The tissue was treated with 500µl of 5% Chelex® 100 (Bio-Rad Ltd, UK) and incubated at 56°C (4 h) in a water bath. This procedure was followed by boiling the solutions in a heating block at 100°C for 15 min. The supernatant was centrifuged briefly and stored at -20°C until required for DNA assessment or PCR.

7.2.2.2. *High quality extractions*

A. Salting out technique The DNA was extracted from the samples following the salting out procedure described in chapter four, section 4.2.1. Extracted DNA was stored in 20-40µl of T.E. buffer (pH8) at –20°C.

B. Guanidium thiocyanate extraction (after, Hammond *et al.,* 1996):
The tissue samples were homogenised in 500µl of guanidium thiocyanate (0.5M:EDTA 0.1M), mixed with 250µl of ammonium acetate (7.5M) at –20°C, and then allowed to stand on ice for 10 minutes. 500µl of chloroform:IAA, 24:1 was added to remove the organic waste and followed by the addition of 1ml of Ice-cold absolute ethanol to precipitate the DNA in the supernatant. The ethanol was decanted off and the DNA pellets washed twice with 1ml of 70% ethanol, and when dry re-suspended in 20-40µl TE.

7.2.2.3. *Qualitative and Quantitative analysis of the DNA isolated*

After extraction from the various specimens and body parts, the DNA was qualitatively assessed. This involved checking whether the extract was present as a single, clear, high molecular weight band on either an agarose or an acrylamide gel, or whether it was present as a smeary low molecular weight band - indicating degradative damage from hydrolytic enzymes.

1. **Agarose gel visualisation**
The DNA was visualised using a 1% agarose gel, run in 1xTBE buffer at 100v and 320 mA for 1 hour, stained with 0.5µg of ethidium bromide. Gels were examined under UV light and photographed. By this method the DNA could be calibrated against commercial molecular weight markers (100bp ladder, Lambda/*Hind*III and Lambda/*Eco* R1 (Gibco BRL, USA)) and a serial dilution of lambda stock DNA (536µg/ml), run on the gel alongside the DNA.

2. **Acrylamide gel visualisation**

To calibrate very low concentrations of DNA, samples were run on 6% 15cm polyacrylamide gels in TBE buffer, run at 180v and 400mA for 3 hours, which were subsequently silver stained (for the silver staining protocol see appendix I).

3. **Quantification using a DNA spectrophotometer.**

The DNA concentration was measured in TE buffer, using a DNA spectrophotometer (SP6-550 UV/VIS, Pye Unicam, UK). To calculate the concentration of DNA, readings were taken at a wavelength of 260nm, whereby an OD of 1 corresponds to approximately 50µg/ml. A reading was also taken at a wavelength of 280nm, which when taken as a ratio of the OD at 280 (OD_{260}/OD_{280}) provides an estimate of the purity of the nucleic acid. Pure DNA has a ratio of 1.8, whereas that contaminated with protein will be significantly less, and indicates that the quantification of the DNA will be inaccurate.

4. **Quantification using a DNA Fluorometer.**

More accurate calibrations of DNA concentration were obtained by fluorometric analysis. The DNA fluorometer (TK0 100, Hoeffer Scientific Instruments, USA) was allowed to warm up for 1 h prior to use and then gauge calibrated following the manufacturer's recommendations. Then 2ml of TNE (10mM Tris, 1mM EDTA, 0.1M NaCl)/Hoechst stain (100ng/ml) (Bio-Rad Ltd., UK) was placed in a crystal cuvette and the gauge set to zero. For sample reading, 2µl of sample DNA was mixed with the TNE in the cuvette by gentle pipette action. The immediate reading on the gauge was recorded and DNA concentration was estimated by reference to a calibration curve set up using concentration standards (Lambda DNA (536µg/ml)). Between readings, the cuvette was rinsed with 2ml of TNE and blotted dry on a clean tissue.

7.2.3. Identifying the most appropriate body part for molecular analysis

Ideally, when using preserved insects for molecular research, the tissue removed for analysis should cause a minimal amount of damage to the set specimen. This is the only way that collectors and museum curators will be encouraged to allow butterflies from their collections to be sampled. Therefore, it was prudent to conduct a study to find a compromise between minimising the damage to the specimen and maximising the likelihood of extracting good quality DNA. A number of body parts were tested, including: heads, thorax, wings, wing base (the muscle at the base of the wing), legs and antennae.

7.2.3.1. *DNA extraction*

Twenty-five freshly killed *P. aegeria* were submitted for DNA extraction from their heads, antennae, legs, wing base and thorax. The DNA was salt extracted, via the method already described in chapter four, section 4.2.1.

7.2.3.2. *Qualitative and Quantitative analysis of the DNA isolated*

The DNA extracted was calibrated for quality and quantity using a selection of the methods already described. These were: ethidium bromide stained agarose gels and DNA fluorometry (See section 7.2.2.3).

7.2.4. The effect of specimen age on the quality of DNA extracted

To investigate the feasibility of obtaining DNA from preserved lepidopteran tissue, a basic study was undertaken. This was to assess the effect of specimen age on the quality of the DNA isolated. This should indicate the level of degradation the DNA has been

exposed to, and provide a rough idea of the amount of DNA that is recoverable from these sources. The samples are detailed in table 7.1.

7.2.4.1. *DNA extraction*

The heads and legs of the specimens (table 7.1) were removed and submitted for DNA extraction. The DNA was isolated using the salting out method already described in chapter four, section 4.2.1. The DNA was suspended in 20µl of TE buffer.

7.2.4.2. *Qualitative and Quantitative analysis of the DNA*

The DNA isolated was then visualised on an agarose gel (see section 7.2.2.3). This not only demonstrated the presence of DNA, but more importantly, it also indicated the level of degradation each sample had undergone. The concentration of DNA samples was also roughly assessed from the intensity of the ethidium bromide staining of the samples. This was deemed as more appropriate than fluorometric analysis, where no information about the quality of the DNA is revealed.

7.2.5. Investigating the effects of specimen age and amplicon size on PCR efficiency

It was crucial to investigate whether DNA extracted from various specimens of different ages was PCR amplifiable, and if so, to identify the size range within which this was possible. The legs from a range of specimens were used for this study (see table 7.1). Initially two amplicons spanning the mitochondrial COI–COII genes were used to assess the feasibility of amplifying DNA from preserved specimens. One of these was a 1.3Kbp region spanning both the COI and COII genes; the other was a short fragment of the COI gene (180bp). Five polymorphic microsatellites specific for *P. bellargus* were also investigated, using the 105-year-old *P. bellargus* specimens from Folkestone. The PCR

efficiency for each of these targeted regions (ranging from 180bp to 1300bp) should provide clear evidence for the amplifiable size range appropriate to degraded DNA.

7.2.5.1. *PCR amplification of the COI-COII region*

Universal PCR primers specific for the mitochondrial COI-COII region [BI-COI (GATACCCGAGCTTATTTAC) (Powers *et al*, 1999) and COII-croz (CCACAAATTTCTGAACATTGACC (Crozier *et al*., 1989)] were used to amplify salt extracted DNA from a range of specimens (aged 11 years or less (table 7.1)), plus a DNA sample that had been obtained by boiling a single, 6-month-old *P. aegeria* leg (see section 7.3.2.1). Each 25µl reaction mixture contained 10-50ng of template DNA, 2.0 units of *Taq* DNA polymerase (Appligene-Oncor, UK), 0.2µM of each primer, 20mM $(NH_4)SO_4$, 75mM Tris-HCl, pH8.8, 0.1% (v/v) Tween®-20; 1.5mM $MgCl_2$ and 0.25mM dNTPs (Pharmacia, USA). A negative control was also included with each PCR; this consisted of the above reagents without the addition of DNA. The following conditions were used for the amplification: 2 minutes at 94°C, followed by 30 cycles of 1 minute at 94°C; 1 minute at 37°C and 1 minute at 72°C, this was followed by a final extension step of 7 minutes at 72°C. This was carried out on a "Touchdown" thermal cycler (Hybaid, UK). After amplification, reactions were visualised on a 1% agarose gel and the size of the products was scored against a 100bp ladder (Gibco-BRL, USA) as a standard molecular marker.

7.2.5.2. *PCR amplification of the COI region*

Universal primers designed for use with invertebrates (See Simon *et al*, 1994) were used to amplify a sub-region of the COI region of the mitochondrial genome (Primer sequence: CI-J-2196: TTGATTTTTTGGTCATCCAGAAGT and C1-N-2329: ACTGTAAATATATGATGAGCTCA). This region was chosen because its small size (*c.* 180bp) meant it should be ideal for amplification in the range of specimens used (all specimens detailed in table 7.1, except the 2.5 year *P. aegeria* and 105 year old *P. bellargus*). Each 25µl reaction mixture contained 10-50ng of template DNA, 2.0 units of

Taq DNA polymerase (Appligene-Oncor, UK), 0.2µM of each primer, 20mM $(NH_4)SO_4$, 75mM Tris-HCl, pH8.8, 0.1% (v/v) Tween®-20; 1.5mM $MgCl_2$ and 0.25mM dNTPs (Pharmacia, USA). A negative control was also included with each PCR; this consisted of the above reagents without the addition of DNA. The following conditions were used for the amplification: 2 minutes at 94°C, followed by 30 cycles of 1 minute at 94°C; 1 minute at 37°C and 1 minute at 72°C, this was followed by a final extension step of 7 minutes at 72°C. This was carried out on a "Touchdown" thermal cycler (Hybaid, UK). After amplification, reactions were visualised on a 1% agarose gel and the size of the products was scored against a 100bp ladder as a standard molecular marker.

7.2.5.3. *PCR amplification of the microsatellites*

The microsatellites developed in chapter five, Lb1/41, Lb1/57, LbG2, Lb4/18, and Lb4/19 (Harper *et* al., 2000), were employed for this analysis. Each 12.5µl reaction mixture contained 50-100ng of salt extracted (chapter four, section 4.2.1.) *P. bellargus* DNA (obtained from the legs of 20 specimens captured at Folkestone in 1896); 0.15 U *Taq* DNA polymerase (ABgene, UK); 0.4µMoles of each primer (the forward primer of each pair was 5' end labelled with fluorescent phosphoramidites FAM, HEX or TET (MWG-Biotech, Germany)); 20mM $(NH_4)SO_4$; 75mM Tris-HCl, pH 8.8; 0.01%(v/v) Tween®-20; 1.5mM $MgCl_2$; 0.25mM dNTPs (ABgene, UK); in the case of LbG2, Ix Q solution (Qiagen, USA) was used. A negative control was also included with each PCR; this consisted of the above reagents without the addition of DNA. The following conditions were used for amplification: 4 min at 94°C; followed by 35 cycles of 30 seconds at 94°C, 30 seconds at annealing temp, 30 seconds at 72°C; followed by 30 minutes at 72°C. This was carried out on a "Touchdown" thermal cycler (Hybaid, UK). The microsatellites were scored using a 5% denaturing polyacrylamide gel by vertical electrophoresis at 20-60mA for 2h on a Perkin Elmer ABI 377 automated sequencer running *GENESCAN* and *GENOTYPER* 2.5 software (PE-Applied Biosystems, USA).

7.2.6. Developing microsatellite markers using DNA from preserved specimens

Because of the likelihood that microsatellites would make one of the most efficient DNA markers to analyse preserved specimens, a simple study was carried out to investigate the feasibility of using preserved specimens for the isolation of these markers. The specimens used were 6-month-old *P. aegeria*; these were young enough that the DNA would not have undergone severe degradation.

The majority of attempts to isolate microsatellite markers (including the method already described in this thesis) have involved the creation of genomic libraries – costly in both time and money. This work investigates the feasibility of using a PCR based approach to the isolation and characterisation of microsatellite repeats (see Ender *et al.*, 1996) whilst using dried lepidopteran specimens as a source of DNA. The method involves "randomly amplified polymorphic DNA" (RAPD) fingerprinting (Williams *et al.*, 1990; Hadrys *et al.*, 1992) to amplify arbitrary fragments of genomic DNA. The RAPD profile obtained can then be Southern blotted and screened for microsatellite repeats by hybridisation with synthetic, radio-labelled $(CA)_n$ oligonucleotides (Genosys, UK). Fragments deemed as containing microsatellite loci could then undergo direct cloning and sequencing.

Forty RAPD-PCR primers from the Primer kit-B and F (Operon Technologies, USA) were tested. Total genomic DNA was isolated from two six-month-old *P. aegeria* specimens, by means of the salt extraction method described in chapter four, section 4.2.1. RAPD amplification was performed, each 25µl reaction consisting of: 10-50ng of template DNA, 2.0 units of *Taq* DNA polymerase (Appligene-Oncor, UK), 0.2µM of primer, 20mM $(NH_4)SO_4$, 75mM Tris-HCl, pH8.8, 0.1% (v/v) Tween®-20; 1.5mM $MgCl_2$ and 0.2mM dNTPs (Pharmacia, USA). Amplifications were carried out on an Omnigene thermal cycler (Hybaid, UK) under the following conditions: 2 minutes at 94 ^oC for one cycle, followed by 40 cycles of 1 minute at 92 ^oC; 1 minute at 35 ^oC; 1 minute at 72 ^oC plus a final extension step of 6 minutes at 72 ^oC. Negative and positive

controls for the PCR were routinely included. After amplification, 20 µl of the reaction mix was run on a 1.5% agarose gel (run at 100v and 320mA for 1 hour) stained with 0.5µg of ethidium bromide. The gel was examined under UV light. A 100 bp ladder (Gibco-BRL, USA) was used as a standard molecular length marker.

RAPD-DNA profiles that were found to give reproducibly polymorphic profiles during the initial screening were Southern blotted to search for potential microsatellites (after Ender *et al.*, 1996). The RAPD-PCR reactions were run on a 0.8% agarose gel at 45v/320mA for 4 h in the presence of a control lane containing 100pg of the purified unlabelled $(CA)_n$ oligonucleotides and two molecular size markers (Lambda/*Hind* III and Lambda/*Eco* R1, Appligene-Oncor, UK). The samples were Southern blotted (Southern 1975) onto a positively charged nylon membrane (Boehringer-Manheim, Germany). The $(CA)_n$ oligonucleotide (40ng) was used in a random labelling reaction with $[\alpha^{32}P]$ dCTP (3000 Ci/mM). The subsequent radio-labelled probe was hybridised to the blotted membrane under stringent reaction conditions: overnight at 65°C in a Hybaid (UK) oven. The membrane was washed once for two minutes at 65°C in 2xSSC; 0.1%SDS, followed by a 30 minute wash in 2xSSC; 0.1%SDS at room temperature and finally 0.5xSSC; 0.1%SDS for 15 minutes at room temperature (this wash was repeated if radioactive levels remained high). Radioactivity was checked between each wash and if necessary, subsequent washes were discarded. Positive signals were detected by exposure of the probed and washed membrane to X-ray film (Kodak, UK). The duration of exposure depended on the level of signal detected from the membrane using a Geiger counter.

7.3. RESULTS

7.3.1. The efficiency of various DNA extraction techniques

Comparisons of DNA obtained by various extraction methods indicated that salt extraction was the most suitable technique for obtaining DNA from preserved specimens (see table 7.2). DNA isolated by this method was clearly shown to be suitable for further processing, by the high PCR success rate of four microsatellites and a short region of the mitochondrial COI gene (over 90%; see results section 7.2.4.2). No images of the extracted DNA have been included in this section, because the low amounts of DNA isolated from wingtips are extremely hard to visualise. The only reliable method for quantification of the DNA was fluorometry analysis (Table 7.3).

Table 7.2. The mean and standard errors of DNA amounts (ng/µl) released from the wingtips of lepidopteran specimens by four different DNA extraction methods. Each mean and standard error was calculated using data from five *P. aegeria* specimens. The results were obtained using DNA fluorometry.

AGE	EXTRACTION TECHNIQUE			
	Salt	Guanidium Thiocyanate	TE	Chelex
2.5 Years	13.6 (4.6)	5.6 (2.9)	3.4 (1.3)	21.8 (7.4)
103 Years	6.4 (1.7)	1.2 (0.9)	2.2 (0.7)	25.2 (4.3)

Although the TE release method was very simple to use, and previous studies have successfully applied this technique (Lushai *et al.* 2000), it recovered consistently lower and more variable amounts of DNA than any of the other methods tested. It is possible that this method is less efficient with increasing age of specimen. The guanidium thiocyanate extraction was comparable in success to the TE release method, but isolated marginally less DNA and had a slightly lower PCR success rate (20% of reactions worked compared to 30%).

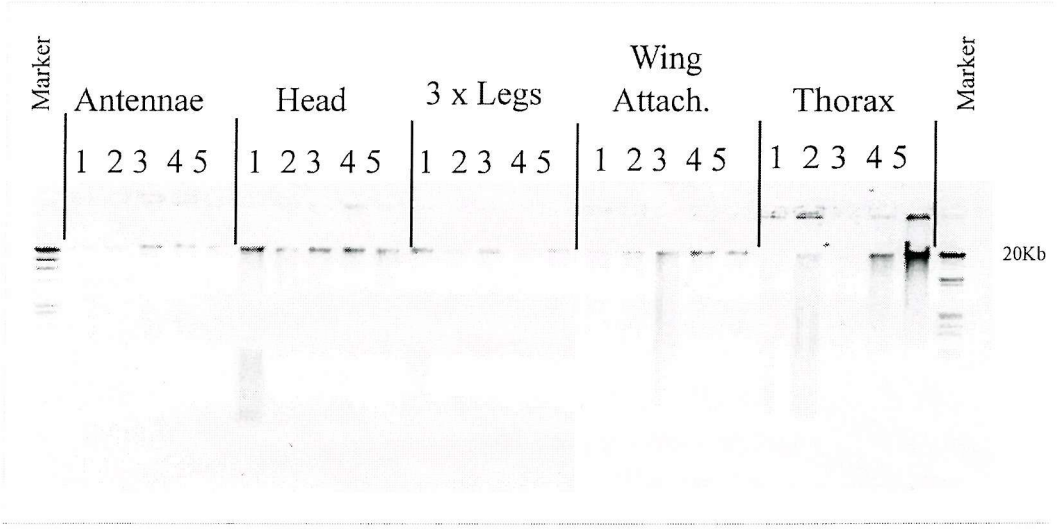
The Chelex method proved inadequate for further study. The inferred DNA concentration was extremely inconsistent, depending on the calibration method used. For example, on neither ethidium bromide stained agarose gel nor silver stained acrylamide gel was there was there any visible indication of DNA, yet fluorometric analysis

indicated much higher levels. This suggested an anomalous detection of a fluorescent moiety unrelated to the presence of DNA, a theory that was confirmed by analysing a solution of the Chelex-100 with no DNA present.

7.3.2. The most appropriate body part for molecular analysis

The DNA extracted from various body parts of fresh killed *P. aegeria* butterflies was examined both qualitatively (figure7.1) and quantitatively (table 7.3). The results indicate which tissues provide the most suitable source of DNA for molecular analysis.

Figure 7.1: An inverted tone image of a subset of genomic DNA extractions, obtained by salting out DNA from various body-parts of freshly killed *P. aegeria* specimens. 1µl of each extraction has been run on a 1% agarose gel stained with Ethidium Bromide. Molecular markers are Lambda/*Hind* III/*Eco* R1, and Lambda/Hind III.



The qualitative analysis of the DNA extracted from the various body-parts shows that the DNA is good quality, being represented by a single, high molecular weight band on an agarose gel (figure 7.1). However, there does appear to be smearing of the DNA in some instances, particularly where there are large amounts of DNA present.

Table 7.3. The mean and standard errors for the quantity of DNA (ng/µl) of released via salt extraction of five different body-parts from twenty-five *P. aegeria*. This was carried out using two different methods of analysis (DNA spectrophotometry and DNA fluorometery).

SPECIMEN	SPECTROPHOTOMETER READING (NG/µL)	FLUOROMETRY READING (NG/µL)
Head	564 (44.6)	121.2 (28.9)
Antennae	404 (73.8)	21.8 (5.5)
Legs (x3)	325.2 (54.9)	30.4 (4.9)
Wing base	222.2 (31.0)	56.2 (9.8)
Thorax	682 (100.1)	263 (111.0)

The mean and standard errors for the DNA quantity isolated from each body-part has been calculated for twenty-five *P. aegeria* (table 7.3). Two methods of DNA quantification have been used, these are: DNA spectrophotometry and DNA fluorometry. The results from these methods gave very different results for the concentration of DNA extracted; in some cases this was more than an order of magnitude. However, the fluorescence method is generally accepted as the more reliable technique of the two, because it only measures DNA, whereas spectrophotometry cannot differentiate between DNA and RNA, plus protein in the sample will interfere with the reading. Therefore, the fluorescence results have been referred to where necessary. Nevertheless, the general trends for both methods are broadly parallel, with both indicating that the head and the thorax produce the highest amounts of DNA.

The results of the study indicate that the head and legs are able to provide the most consistent source of high quality DNA (see table 7.3). Whilst other body-parts, such as the thorax and wing base, produce reasonable amounts of DNA, the extractions from these parts are less consistent. The DNA extracted from the legs is very consistent between individuals, but it is somewhat lower than the amounts obtained from the head, nevertheless the amounts isolated should be more than adequate for PCR amplification. The DNA obtained from the antennae also seems to be high quality, showing no sign of degradation. The only proviso here is that the antennae are less reliable as a source of DNA because they have often been lost from preserved specimens.

7.3.3. The effect of specimen age on the quality of DNA extracted

The initial investigations into the effects of specimen age concentrated on analysing the quality of DNA that could be extracted. This was demonstrated via ethidium bromide stained gels showing the salt extracted DNA from heads of a range of different aged specimens (figure 7.2) and that extracted from single legs of the 105 year old *P. bellargus* (figure 7.3).

Figure 7.2: An inverted tone image showing genomic DNA that was salt extracted from the heads of different aged lepidopteran specimens. The samples have been run on a 1% agarose gel stained with ethidium bromide. Molecular markers are Lambda/*Hind*III/*Eco*RI, and Lambda/*Hind*III.

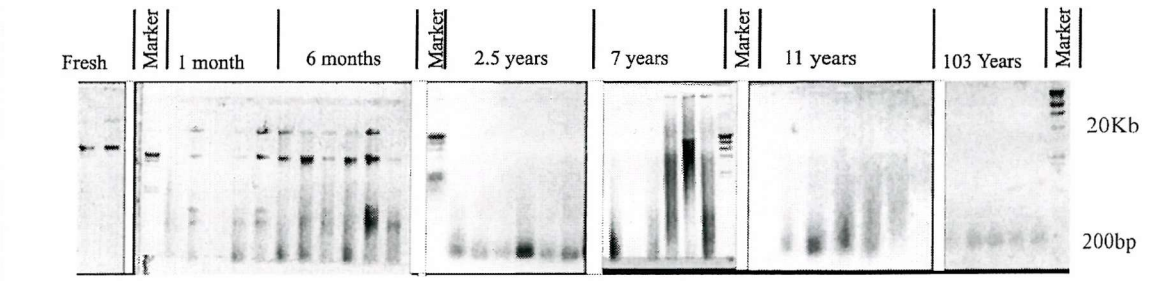
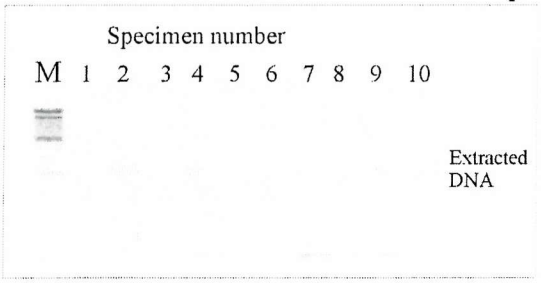


Figure 7.3. An inverted tone image of salt extracted DNA (100-300bp) from the legs of ten *P. bellargus* individuals from Folkestone, captured in 1896. The samples have been run on a 1% agarose gel stained with ethidium bromide. Lane "M" denotes the presence of a 100bp ladder molecular marker.



It is very clear that the quality of the DNA is reduced with increasing specimen age. In the 103-year-old specimens, the DNA extracted from heads (figure 7.2) has been reduced to fragments of a few hundred base pairs, observable as a vague smear at the base of the gel, this can also seen in the extracts from the 105 year old *P. bellargus* legs (figure 7.3). In contrast to this, the level of degradation after six months is minimal and even a small

amount of RNA remains in a few samples (RNA is generally more sensitive to degradation). The RNA can be seen as one or two faint bands towards the lower-middle part of the gel. The consistency among samples decreases with increasing age; for example, the 7-year-old specimens show a vast amount of variation of DNA quality among them, even though they were stored under identical conditions. The same trends were present in the DNA extracted from the legs of specimens, though obviously lower concentrations of DNA were isolated (figure 7.3).

A further observation that can be made from this experiment concerns the way in which storage conditions can affect the DNA. This is clearly demonstrated by a comparison between the 2.5-year-old samples and the 7 and 11-year-old specimens. The younger butterflies were stored in wax-coated envelopes, prolonging the time taken for them to dehydrate, resulting in extensive DNA degradation. The older samples were pinned and dried immediately after death and have consequently suffered less DNA degradation.

This qualitative analysis of the DNA extracted from preserved specimens is useful, but may not infer much about the suitability of the DNA for molecular based studies. Any study of this kind, involving preserved specimens, will undoubtedly require a PCR approach. It is therefore imperative to investigate how the DNA degradation associated with sample ageing affects the feasibility of PCR amplification.

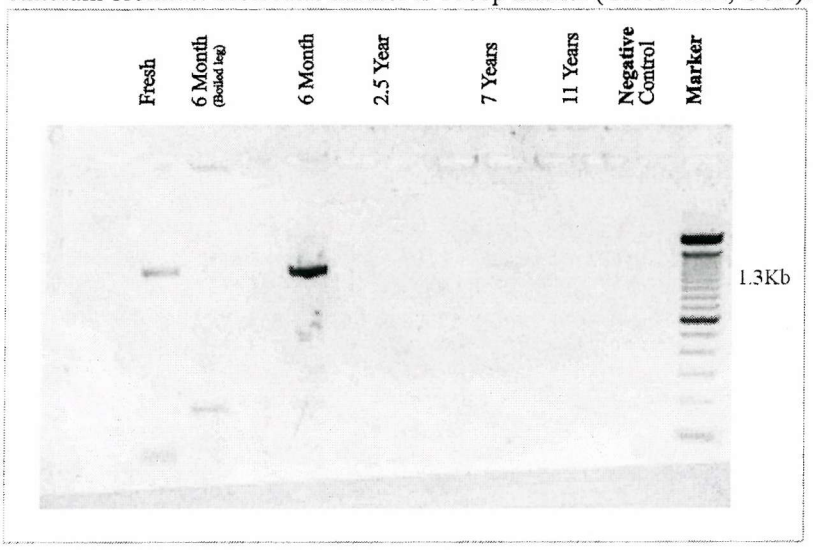
7.3.4. The effects of specimen age and amplicon size on PCR efficiency

7.3.4.1. *PCR amplification of the COI-COII region*

The results indicate successful amplification of the 1.3Kbp COI-COII region in a number of instances, even from the legs of specimens as old as 7 years (figure 7.4). DNA isolated from both fresh-killed and 6-month old *P. aegeria* specimens consistently resulted in amplification of the region, identified as a single, clear band on an agarose gel. This correlates well with the quality of the DNA that was visually assessed in the

previous section (section 7.3.3), where DNA from the 7 year old butterflies was shown to be far less degraded than any of the other preserved specimens over one year old.

Figure 7.4: A inverted tone image showing a subset of a PCR amplification of the COI-COII region of the mitochondrial genome, from DNA extracted from the heads of an age range of stored lepidopteran material. Two amplifications are shown for each aged specimen, each with different DNA concentrations; a 1:5 and a 1:0 dilution of stock DNA respectively. The samples have been run on a 1% agarose gel stained with ethidium bromide. Molecular marker is 100bp marker (Gibco BRL, USA).

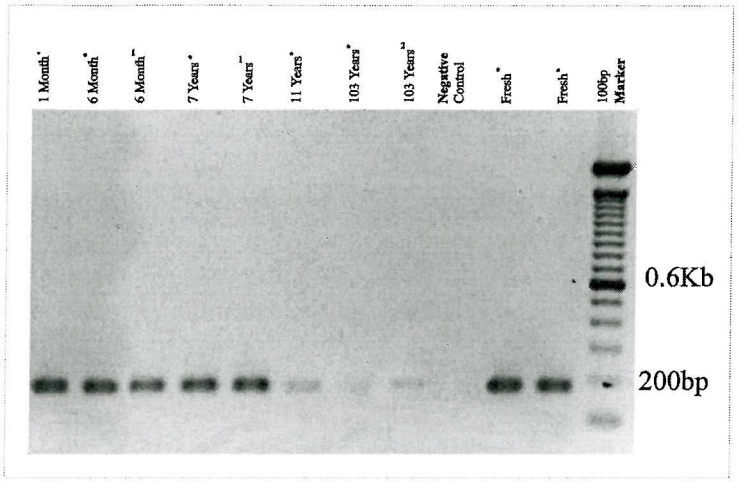


The amplification of DNA released by boiling a 6-month-old *P. aegeria* leg, has produced an extra spurious band of around 200-300 base pairs. The origin of this is unknown, but it is likely to be the result of non-specific binding of the primers to regions of DNA alongside the targeted 1.3Kbp of DNA.

7.3.4.2. *PCR amplification of the COI region*

The partial success of the 1.3Kbp amplification implies that a smaller target region would be more appropriate. The smaller COI region was amplified, using universal mitochondrial DNA primers (Simon *et al.* 1994). This region was *c.* 180bp, and as such was much smaller and closer to the size range of microsatellite markers.

Figure 7.5: An inverted tone image of the PCR amplified mitochondrial COI region, using Universal primers (CI-J-2196 and C1-N-2329) (Simon *et al.* 1994). The template DNA was salt extracted from various aged lepidopteran specimens. The samples have been run out on a 1% agarose gel, and stained with ethidium bromide.



PCR amplification has clearly been successful using DNA isolated from the legs of specimens of all ages, identified as a single band of just under 200bp in size (see figure 7.5). The quality of the amplification was extremely consistent with DNA from specimens aged 7-years or less. However, after this point, the PCR becomes less robust, with levels of amplified DNA decreasing. This can be roughly judged by the intensity of the band produced on the gel (figure 7.5). This reduction in PCR efficiency correlates very closely with the quality of the DNA extracted (section 7.3.3), demonstrating that the more degraded the DNA, the lower its efficiency as a template for PCR amplification.

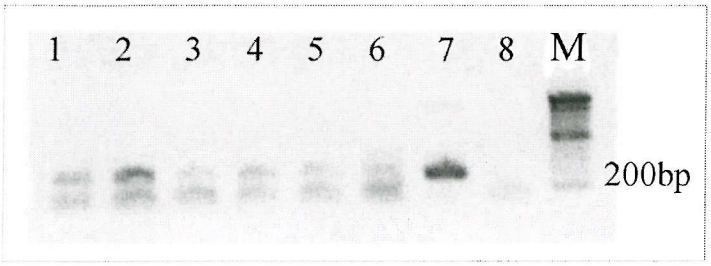
Crucially, this shorter region was successfully amplified from the 103-year-old *P. rhomboidaria*; the oldest specimen analysed. The highly fragmented nature of the DNA extracted from this specimen has not prevented amplification of this short region of the COI gene, with over 90% of PCR amplifications being successful. This has exciting and far-reaching implications for the feasibility of microsatellite markers as an approach for the genetic analysis of preserved lepidopteran specimens.

7.3.4.3. *PCR amplification of the microsatellites*

The PCR amplification of four of the five *P. bellargus* microsatellites (Harper *et al.*, 2000) (Lb1/57, LbG2, Lb4/18, and Lb4/19), demonstrated the applicability of these markers to DNA from museum preserved specimens. The template DNA had been salt extracted from legs removed from a collection of twenty *P. bellargus* specimens caught at Folkestone during 1896 (105 years old), and although this represented the oldest group of specimens analysed, the PCR resulted in clear and reproducible amplification from all but one of the primer sets (Lb1/41) (see figure 7.6 and 7.7). Only Lb1/41 failed to amplify from the museum specimens, even after repeated attempts to optimise the PCR. This failure is difficult to explain, because Lb1/41 is not significantly larger in size to the other four microsatellites (c. 200bp), theoretically ruling out DNA degradation as the cause of failure. One possible theory is that a null allele has reached fixation within this population, a high incidence of null alleles has previously been reported for this locus (see chapter five). However, this argument also seems improbable because the other four loci all demonstrate reasonable levels of heterozygosity (0.3 to 0.8) (see section 6.3.1.2, chapter six), therefore there is no reason to assume such low levels of heterozygosity in Lb1/41.

Figure 7.6. An inverted tone image of a PCR amplified microsatellite (top section: Lb1/57; bottom section: Lb4/19). The template DNA was salt extracted from 105 year old *P. bellargus* specimens (labelled 1-16). The amplifications have been run out onto a 1% agarose gel and stained with ethidium bromide. "M" denotes a lane containing a 100 base pair DNA ladder.

Figure 7.7. An inverted tone image of a PCR amplified microsatellite (top section: LbG2). The template DNA was salt extracted from 105-year-old *P. bellargus* specimens (labelled 1-8). The amplifications have been run out onto a 1% agarose gel and stained with ethidium bromide. "M" denotes a lane containing a 100 base pair DNA ladder.



The remaining four loci (Lb1/57, LbG2, Lb4/18, and Lb4/19) all resulted in good quality microsatellite data (see chapter six for details of this). When assessed using ethidium stained agarose gel, the amplification results, although clearly present, are hard to interpret because of their small size and the low levels of amplified DNA (see figures 7.6 and 7.7). However, when the same amplifications were assessed using a Perkin Elmer ABI 377 automated sequencer running *Genescan* software (PE-Applied Biosystems, USA), confirmation of their amplification was irrefutable.

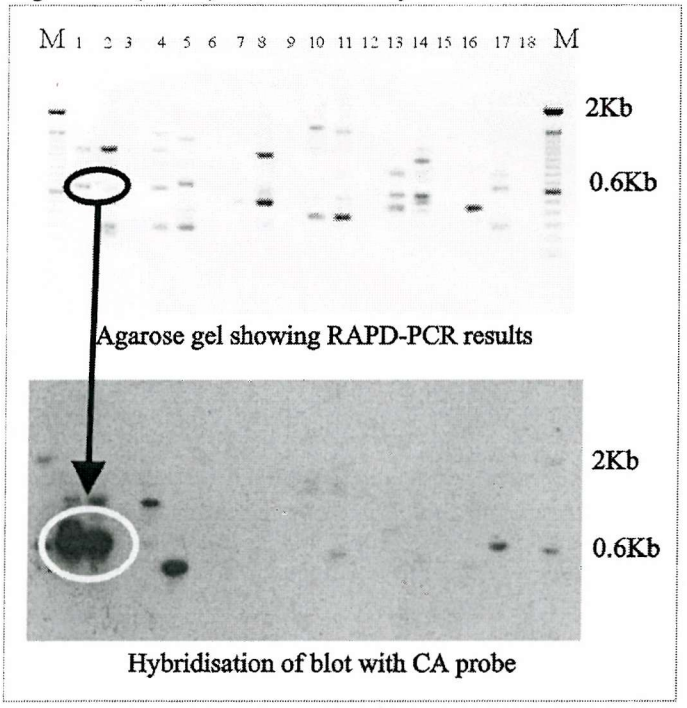
The data obtained is unlikely to result from contamination of the PCR by DNA from modern *P. bellargus*, because whilst polymorphism was present between individuals, no specimen was represented by more than two allelic states. The possibility of non-specific amplification was also ruled out, because the amplifications all showed stutter bands characteristic of microsatellite amplifications. The fact that the primers are species specific should also rule out the possibility of contamination by DNA from other species.

7.3.5. The development of genomic DNA markers using preserved specimens

The results of the RAPD-PCR screening from two 6-month-old specimens of *P. aegeria* show that, as expected, the PCR has produced an assortment of bands of different sizes (figure 7.8). The reliability of the PCR is demonstrated by the fact that each pair of individuals is represented by an almost identical banding pattern when amplified using

the same primer, and that this pattern can consistently be reproduced. Slight discrepancies between the banding patterns for the two individuals would be expected because of the nature of RAPDs, where variation in the DNA sequences causes incongruity in the size of amplicons, or where a mutation in the primer binding site causes their absence altogether.

Figure 7.8 (above): An inverted tone image of a randomly amplified polymorphic DNA (RAPD) PCR amplification with useful arbitrary primers from a commercial selection (n=40) (Operon kit F and kit B) (Lanes 1&2: OPB-04, lanes 4&5: OPB-5, Lanes 7&8: OPB-6, Lanes 10&11: OPB-10, lanes 13&14: OPB-12, lanes 16&17: OPB11; "M" denotes 100bp DNA ladder), on a 1.5% agarose gel stained with ethidium bromide. Negative controls are present between each pair of amplifications (lanes: 3, 6, 9, 12, 15 and 19).
Figure 7.9. (below): Southern blot hybridisation of RAPD-DNA, probed with a radio-labelled $(CA)_n$ repeat.



The radio-active probing of the Southern blotted RAPD-PCR gel, using a ^{32}P labelled $(CA)_n$ probe, shows distinct areas of intense darkness on the photographic film (figure 7.9.). These indicate where the probe has bound to $(GT)_n$ repeats within the RAPD amplified bands. An example has been highlighted between the two figures, demonstrating where an amplified band (figure 7.8) is clearly related to a region on the membrane where the probe has bound (figure 7.9). This is likely to indicate the presence of a $(CA/GT)_n$ microsatellite. The other, weaker, bands probably also contain $(CA/GT)_n$ repeat regions and would also warrant further investigation.

7.4. DISCUSSION

7.4.1. Extraction technique, body part and age of specimen

This preliminary work has given some useful indications of the approach to take concerning the extraction and processing of DNA from preserved specimens. The work has been very encouraging; by demonstrating that DNA extractions using preserved tissue can provide material that is suitable for PCR application. The applicability of this to population studies has also clearly been demonstrated by the amplification of four microsatellites from twenty *P. bellargus* specimens dating from 1896. This opens a whole new approach to the study of population genetics, where not only is a spatial approach possible, but also studies into temporal variation.

The initial investigation assessed four of the simplest DNA extraction procedures mentioned in the literature to date (specifically pertaining to the field of preserved specimens). The techniques varied from very crude methods that do not isolate the DNA from the proteins, to techniques that purify the DNA from other contaminating substances. The results of this study strongly advocate the use of an extraction technique that purifies the DNA away from other contaminants, in this study the salt extraction technique proved most efficient. This method resulted in consistently higher yields of DNA, and more reproducibly successful PCR results. Whilst the guanidium thiocyanate technique separated the DNA from other organic products, it gave far less reliable results than salt extracted DNA.

The crude extractions, which had been applied very successfully in previous experiments (Cano *et al.*, 1993; Lushai *et al.*, 2000), were found to be very unreliable in this study. The Chelex method has been used to isolate the DNA from insects stored in amber for millions of years old (Cano *et al.* 1993), yet in this study it was found to be the least successful method examined. Very low amounts of DNA were isolated, resulting in high levels of PCR failure. This may also be because Chelex is a weak cation-chelating resin, with a high affinity for metal ions; this may cause disturbances in the magnesium

chloride concentrations crucial for the activity of *Taq* polymerase in PCRs. The method additionally produced some kind of spurious fluorescence that interfered with the quantification of the DNA.

None of the crude extraction methods employ a proteinase step to break down the cell walls and enable DNA release from the cells. The methods rely on boiling the samples to cause cell breakdown and subsequent release of DNA. There is also no attempt to separate the proteins from the DNA before PCR amplification, causing lower yields of accessible DNA. What little DNA that is released, is more likely to be subject to degradation, particularly as none of the degradative enzymes naturally found within the cells have been removed. Because of the EDTA content of TE, enzymatic breakdown of DNA should be inhibited. This is because of its affinity for metal ions involved in enzymatic reactions; however, this will also tend to interfere with the magnesium levels in PCR (similar to that described for Chelex).

The investigation into which part of the specimen is most efficient for the extraction of DNA gave very clear results. This strongly advocated the extraction of DNA from butterfly heads, because they consistently yielded the highest quality and quantity of DNA. However, for valuable museum specimens, the removal of the head will be too destructive. The high level of DNA obtained is probably representative of the fact that the head strikes a good compromise between drying speed, and lower amounts of chitin surrounding the tissue. Limbs such as the legs are coated in thick layers of chitin, a substance that is difficult to break down. The thorax is much softer, but its larger size causes the drying out process to be extended, allowing increased degradation to occur. Figure 7.1 exemplifies this pattern, where many of the extractions from the thorax are smeary, indicating degradation, whereas the leg extractions, although exhibiting much smaller amounts of DNA, show a single sharp band. However, most cases where the DNA has smeared, the band intensity indicated a high concentration of DNA. Therefore, this smearing may simply result from a loss of clarity caused by the high concentration of DNA.

The DNA from the antennae of the butterflies was extremely inconsistent, with the quantity of DNA very variable between individuals. The antennae are a particularly delicate part of the specimen, and any rough handling will cause them to break off; thus not many preserved specimens retain them intact. Therefore, collectors are unlikely to be willing to allow removal of antennae from the few specimens that do have them, because they are an aesthetically important part of the butterfly. The wing base and thorax are not really a viable choice, as removal of these will result in the complete destruction of the specimen. Therefore for valuable specimens, the most suitable body part for DNA extraction will be the legs, which being on the underside of the specimen; their removal will have no impact on the aesthetic value of a specimen. For less valuable or fresh-killed butterflies, the head will be the most appropriate body-part for extraction.

When the effect of ageing on DNA quality was investigated, it became clear that the DNA becomes increasingly degraded through time, and also that this 'ageing effect' will be accelerated under damp storage conditions. These effects are exemplified in figure 7.2. The DNA seems to retain its integrity extremely well during the first 5 months of storage, but after this point there is a rapid decrease in quality. It is important to note here that because of the lack of specimens available, there is poor consistency between the samples aged more than 6 months old. The 2.5 year-old samples were stored damp within wax-coated envelopes, undoubtedly encouraged degradation of the biological material. This theory is demonstrated by the 2.5-year-old samples compared with both the 7 and 11-year-old samples. Here the levels of degradation in the older, air-dried specimens are far lower than the younger specimens, which were stored in the wax envelopes. Unfortunately, no specimens aged between 11 years and 103 years old could be located, resulting in a large jump in the age of specimens. However, the analyses on the 105-year-old specimens clearly showed that they contain extractable DNA, seen in figure 7.3 as a smear at the base of the gel. Although this DNA has been severely degraded, and is present as fragments of no more than a few hundred base pairs, attempts to PCR amplify short regions of the mitochondrial genome, even using the DNA from a single leg, were highly successful.

The work detailed in this chapter has clearly demonstrated that even DNA as old as 105 years taken from a single leg is useful for population studies; the only pre-requisite is that small DNA markers are chosen (less than 300 base pairs). Various mitochondrial and nuclear microsatellite amplifications demonstrated this. Of the mitochondrial primer sets used, only the pair targeting the shorter COI region (180bp) amplified consistently in all specimen ages. The second pair targeted a larger region (1.3Kbp), which although amplified the less fragmented DNA, was not suitable for the highly degraded older DNA. This theory was also demonstrated by the amplification of four microsatellite loci, all of which were less than 250bp in length. These loci were all amplified from DNA extracted from twenty 105-year-old *P.* bellargus specimens; representing the oldest group of specimens analysed; nevertheless, the PCR success rate approached 100%. This provides irrefutable evidence for the applicability of microsatellites to the genetic analysis of museum specimens, whilst also demonstrating that even the DNA from a single leg is enough for PCR analysis. This work has exciting implications for the field of population genetics, where microsatellites are already having a huge impact. The demonstration of their applicability to historical investigations opens up a whole new avenue of population research.

7.4.2. Developing microsatellite markers using DNA from preserved specimens

The RAPD-PCR approach to microsatellite isolation seems to have potential as a method for identifying repeat regions, even when using preserved specimens as a source of DNA. The alternative approach to the isolation and characterisation of microsatellites is to use a genomic library, a method requiring large amounts of DNA. The RAPD-PCR technique is ideally suited to situations where only a small amount of DNA is available, such as would be the case where only museum specimens are available. The one proviso for this approach concerns the level of degradation in very old samples. This may have an adverse effect on the reproducibility of the RAPD-PCR profiles, impeding the reliability of the technique. The samples used in this study had been stored dried for 6 months, and no perceivable detrimental effect was found as a result of the specimen age.

The RAPD-PCR gave consistent results, with a small amount of variation between specimens. This low variation would be expected with RAPDs as a result of small sequence discrepancies among individuals. By Southern blotting the gel and probing this with a labelled $(CA)_n$ repeat, the bands containing microsatellite loci could clearly be identified. The next step of this process would be to extract the DNA from the appropriate band on the agarose gel, and to re-amplify it using the original RAPD-PCR primers. Once the identified band has been purified, it would be sequenced; this will identify the microsatellite repeat, and allow subsequent primer design from its flanking regions. The work was cut short at the stage where the repeat regions were identified, and no sequencing was carried out. This was because a set of microsatellite markers has already been isolated and characterised for *P. bellargus* (Harper *et al.*, 2000). However, the work did sufficiently demonstrate that the technique was a feasible approach to the isolation of microsatellites.

8.0 MICROSATELLITE ANALYSIS OF MUSEUM SPECIMENS

8.1. INTRODUCTION

Genetic theory predicts that if a population undergoes a significant demographic contraction, there will be considerable losses of genetic diversity as a consequence of reduced population size (Wright, 1969; Nei *et al.*, 1975; Chakraborty & Nei, 1977; Lacy, 1987). On the basis of this prediction, low levels of genetic diversity have frequently been used to infer past population bottlenecks in a number of studies (Bonnell & Selander, 1974; O'Brien *et al.*, 1983; Ellegren *et al.*, 1996). However, this seemingly obvious conclusion may be erroneous and ignores other alternative explanations such as environmentally selected differences (Bouzat *et al.*, 1998b). In attempts to further elucidate causative factors for observed allele frequency distributions, a number of statistical methods have been devised based on theoretical effects of population declines. For example, it has been proposed that an excess of heterozygosity relative to allele diversity at individual loci can infer a recent population decline, a theory that is the basis of a software analysis programme designed to identify recent population bottlenecks (Cornuet & Luikart, 1996; Luikart *et al.*, 1999). To date, almost all studies concerning changes in population structure have used this type of indirect methodology to infer demographic changes that populations may have undergone (for example: Mallet *et al.*, 1983; Le Page *et al.*, 2000; Spencer *et al.*, 2000; Waits *et al.*, 2000).

However, recent advances in molecular biology have allowed the routine extraction of DNA from historic and even ancient tissue specimens (see chapter 7; Thomas *et al.*, 1989; Hausworth, 1994; Roy *et al.*, 1994; Bouzat *et al.*, 1998a). These new methodologies allow direct measurement of changes in allele frequencies and gene coalescence that previously could only be inferred indirectly from the spatial distribution of alleles in extant populations (Avise, 1989). This ability to observe changes in genetic structure over a temporal scale is of great importance to conservation geneticists who can use this technology to understand the genetic effects of population bottlenecks, founder events or to reconstruct past population structure. To date, there are just two published

studies that have applied these new methodologies to investigate of the effects of population declines, by directly quantifying genetic variability before and after a demographic contraction. This has been achieved by directly comparing allele frequencies in museum specimens with that of individuals from current populations (Bouzat *et al.*, 1998a; Groombridge *et al*, 2000). Both of these studies provide compelling evidence for a link between the occurrence of a bottleneck event and a subsequent reduction in genetic diversity. This effect is exemplified by the presence of what are known as 'ghost' alleles, these are alleles that have been identified in historical populations but can no longer be found in the modern day populations (Bouzat *et al.*, 1998a; Groombridge *et al*, 2000).

The methodologies necessary for this type of analysis have been reviewed and investigated in detail in chapters 7 of this thesis. Briefly, there are a number of limitations to DNA analysis from preserved specimens; the most important of which is the quality of the DNA that can be obtained from these sources. The majority of studies have demonstrated that the DNA will be severely degraded, often to fragments around a few hundred base pairs (Paabo, 1989; Cockburn & Fritz, 1996; see also chapter 7 of this thesis). Clearly this will have important implications concerning the approach taken for the analysis, where a PCR based methodology using short DNA markers would be most applicable and effective (Thomas *et al.*, 1989; Ellegren, 1991; Cano & Poinar, 1993; Mitton, 1994; Hammond *et al.*, 1996). The most obvious choice in this situation would be microsatellite markers; their small size (150-300bp) makes them ideally suited to this task, and they also provide robust population genetic data. The ability to amplify microsatellites in preserved butterfly specimens has clearly been demonstrated by work detailed in this thesis, even with dried butterfly specimens in excess of 100 years old (see chapter 7).

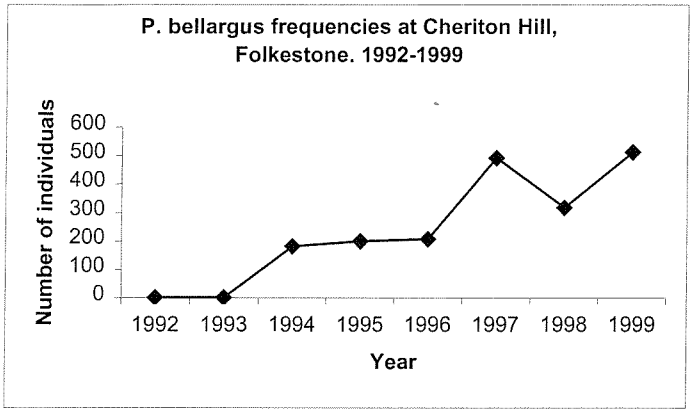
Although destructive sampling for scientific research is somewhat frowned upon nowadays, the eagerness of naturalists over the last few centuries has resulted in vast numbers of preserved specimens, representing countless numbers of species. Not only does this include a large number of extant species, but also many species that have

subsequently become either extinct or endangered. In particular, butterflies have always been a popular taxon for entomological collectors, especially attractive and highly variable species such as *P. bellargus*. Consequently there are extensive butterfly collections around the world. For example, the entomological collections of the London Natural History Museum alone hold over seven thousand *P. bellargus* specimens (David Carter, personal communication). Moreover, detailed data is usually available for each butterfly pertaining to the location and date of its capture. With the advent of molecular genetic techniques, this wealth of carefully documented and preserved biological specimens provides material for a whole new avenue of research. Not only this, but historical population data is available for most butterfly species in Britain, as a result of a nationwide network of amateur lepidopterists. This wealth of data means that specific populations with known histories can be targeted to reveal genetic information. There is an abundance of this kind of historical data available for *P. bellargus*, which if combined with the profusion of museum specimens, provides a rare opportunity for the analysis of temporal genetic variation within a species.

This chapter details a pilot study into the application of microsatellite loci to investigate temporal changes in allele frequencies, using both modern and preserved specimens. This should demonstrate the enormous possibilities of this kind of temporal approach, particularly for this species, where thousands of preserved specimens are available for molecular sampling. The population used was from Folkestone, Kent, and is ideal for the study because it is represented by a substantial number of specimens (74) collected during the 1998 and 1999 field seasons (see chapter 3) and also by twenty specimens at the Oxford Natural History Museum that were all collected from the same location during 1896. According to data compiled by the Kent division of English Nature, this population has rapidly expanded since 1992 to become one of the largest in the UK (see figure 8.1). The data for the site also indicates that subsequent to 1978, when *P. bellargus* underwent a severe decline, and prior to 1992, the population had remained extremely small. By screening the two temporally defined "populations" (the recently collected sample and the 1896 sample) with microsatellite loci, a pattern of variation across the timescale (1896 – present day) should become apparent. The dataset obtained

may indicate a reduction in genetic diversity in the modern specimens relative to those from 1896; a decrease which could be correlated to the decline that the species underwent in 1978. Another trend that may be apparent is a significant change in the pattern of allele frequencies; possibly resulting from genetic drift driven by the demographic contraction and subsequent re-expansion of the population.

Figure 8.1. Temporal variation in population density for *P. bellargus* at Cheriton Hill, Folkestone. This data was compiled by English Nature, Kent.



8.2. MATERIALS AND METHODS

8.2.1. Sample Collection

8.2.1.1. *The 1998 and 1999 specimens*

The sampling was limited to tatty looking male butterflies at the end of their flight season, where the likelihood was that they would already have mated. During the 1998 and 1999 field seasons, a total of 74 *P. bellargus* specimens were collected from Cheriton Hill at Folkestone. This consisted of 32 specimens captured in 1998 and 42 captured in 1999. After analysis for temporal structuring between these two generations (exact G-test, $p > 0.05$), it was deemed that there was no significant overall genetic differentiation (see chapter six for details of this analysis), hence the two groups were combined.

8.2.1.2. *The museum specimens from 1896*

Twenty *P. bellargus* specimens were sampled from the entomological collections at the Oxford Natural History Museum. This involved the removal of a single leg from each butterfly, using a sterile pair of tweezers. All of the butterflies involved in the analysis were from a group that was documented as having been captured at Folkestone during 1896. The fact that these specimens are from the same locality as the modern day specimens should eliminate the possibility of detecting allele frequency differences resulting from geographic differentiation.

8.2.2. DNA extraction

The DNA was extracted from the *P. bellargus* individuals following the salting out procedure described in chapter four, section 4.2.1. The DNA extracted from fresh specimens was stored in 40µl of T.E. buffer (pH8) at –20°C, whereas the DNA from the museum specimens was stored in 20µl of T.E. buffer (pH8) at –20°C. The DNA was assessed for quality and quantity by running it for 1 hour at 100v on an agarose gel stained with 0.5µg of ethidium bromide, then visualisation using a UV light trans-illuminator.

To avoid contamination, the DNA from the museum specimens was extracted in a separate laboratory from the modern day specimens. The fact that the microsatellites are species specific should reduce the possibility of obtaining spurious amplification products. Contamination of the museum DNA was deemed as highly improbable because all individual samples were represented by no more than two alleles for any microsatellite locus and also because novel alleles were amplified from the museum specimens compared with the modern specimens (see results and discussion).

8.2.3. PCR amplification of the microsatellite DNA

Four of the microsatellites developed in chapter 5 (Harper *et* al., 2000) were employed for this analysis (Lb1/57, LbG2, Lb4/18 and Lb 4/19). Whilst attempts were made to optimise the amplification of Lb1/41 from the museum specimens, the PCR consistently failed. This marker was therefore removed from the analysis.

8.2.3.1. *The PCR amplification of the 1998 and 1999 specimens*

For the fresh specimens, each 12.5µl reaction mixture contained 10-50ng of salt extracted *P. bellargus* DNA; 0.15 U *Taq* DNA polymerase (ABgene, UK); 0.2µMoles of each primer (the forward primer of each pair was 5' end labelled with fluorescent phosphoramidites FAM, HEX or TET (MWG-Biotech, Germany)); 20mM $(NH_4)SO_4$; 75mM Tris-HCl, pH 8.8; 0.01%(v/v) Tween®-20; 1.5mM $MgCl_2$; 0.25mM dNTPs (ABgene, UK); in the case of LbG2, 1x Q solution (Qiagen, USA) was added. The following conditions were used for amplification on a HYBAID (UK) "Touchdown" thermal-cycler: 4 min at 94°C; followed by 35 cycles of 30 seconds at 94°C, 30 seconds at annealing temp (50°C for 1/G2, 4/19, 4/18 and 46°C for Lb1/57), 30 seconds at 72°C; followed by 30 minutes at 72°C. The microsatellites were scored using a 5% denaturing polyacrylamide gel by vertical electrophoresis at 20-60mA for 2h on a Perkin Elmer ABI 377 automated sequencer running *GENESCAN* and *GENOTYPER* 2.5 software (PE-Applied Biosystems, USA).

8.2.3.2. *The PCR amplification of the museum specimens*

The PCR reaction and amplification conditions for the museum specimens were identical to those detailed in section 8.2.3.1, with the exception of the amount of DNA template, which was increased to 50-100ng and the concentrations of each primer, which was increased to 0.4µMoles per reaction.

8.2.4. Analysis of the data

8.2.4.1. *Hardy-Weinberg Equilibrium*

In a randomly mating population, observed heterozygosity (H_O) calculated from allele frequencies would be expected to conform to the expected value (H_E) calculated within the Hardy-Weinberg equilibrium. If, as proposed, the twenty specimens captured in 1896 belong to a single population, then theoretically the calculated values for H_O and H_E at each locus should not vary significantly from one another. However, previous work in this thesis (see chapter 6) has already demonstrated that at least four out of the five microsatellites available for *P. bellargus* (Harper *et al.*, 2000) have a high incidence of null alleles (Lb1/41, Lb1/57, Lb4/18 and Lb4/19). Null alleles are where there is nucleotide sequence variation at the primer annealing sites (Lehmann *et al.*, 1996). The result of this is that none of these loci consistently perform as expected under the Hardy-Weinberg equilibrium, almost always being represented by an observed heterozygote (H_O) deficit. Only one locus, LbG2, fits the Hardy-Weinberg equilibrium in most of the modern populations analysed, including that sampled from Folkestone (see chapter 6). Assuming the twenty specimens from 1896 represent a single randomly mating population, the genotypic data obtained for LbG2 should also fit the Hardy-Weinberg equilibrium.

Microsatellite genotypes were tested for departure from Hardy-Weinberg equilibrium at each locus for the twenty museum specimens, using the computer programme *ARLEQUIN* (Schneider *et al.*, 2000). *ARLEQUIN* employs a Markov-chain random walk algorithm (Guo & Thompson, 1992), which is analogous to Fisher's (1935) exact test but extends it to an arbitrary sized contingency table. Where a significant heterozygote deficit was found, estimations were made towards the frequency of null alleles (n) by the method of Brookfield (1996).

8.2.4.2. *Allelic evidence for a population decline*

In order to assess whether the allelic data obtained from the microsatellite analyses showed evidence of a recent demographic contraction, both temporal populations were submitted for analysis by the programme *BOTTLENECK* 1.2.02 (Cornuet & Luikart, 1996; Piry *et al.*, 1999). This program works on the assumption that allele number (k) is generally reduced faster than observed heterozygosity (H_O) (Hedrick *et al.*, 1986), thus recently bottlenecked populations will display an excess of heterozygosity relative to allele number (refer to section 6.2.4.3, chapter 6 for a more detailed explanation of this). When populations with known histories were analysed by this method, under the predictions of the infinite allele model (IAM), the method was found to be a powerful indicator of recent demographic contraction (see section 6.3.1.4, chapter 6). Because very little population data is available prior to 1896, it is important to check whether the microsatellite data obtained infers a population decline previous to this date. If this possibility is not ruled out, then the 1896 data may already represent a population that has experienced a loss of genetic diversity associated with a reduction in population size. This would severely compromise the effectiveness of the work detailed in this chapter to reveal a temporal reduction in genetic diversity associated with a genetic bottleneck.

For the analysis, statistics for all three models of mutation were calculated. This included two-phase model of mutation (TPM), as recommended by Luikart *et al.*, (1998) to analyse microsatellite data. This model is based on a Stepwise Mutation Model (SMM) with 10% multi-step mutations. A Wilcoxon signed rank test was performed to test the hypothesis that the average standardised difference across loci for each population is not significantly different from zero (Spencer *et al.*, 2000). A second indication of a bottleneck is a mode-shift away from an "L-shaped" distribution of allele frequencies, to one with fewer alleles in the low frequency categories. However this qualitative method is not a proper statistical test, because the type I error rate varies with samples size. It cannot be used (with confidence) with samples of fewer than 30 individuals (Cornuet &

Luikart, 1996). These were all assessed using the program *BOTTLENECK* (Cornuet & Luikart, 1996; Piry *et al.*, 1999).

8.2.4.3. *Temporal changes in allelic data*

Basic statistics such as allele number, observed (H_O) and expected heterozygosity (H_E) and frequency data were calculated for both populations. This was carried out using the program *FSTAT* (version 2.9.1.); the frequency data obtained were displayed graphically in combined frequency histograms.

Differences in mean expected heterozygosity (H_E) and mean number of alleles per locus between the 1896 and modern *P. bellargus* populations were tested using a one-way Analysis Of Variance (ANOVA), with sampling date as the main factor. In order to normalise the heterozygosity data; it was subjected to an arcsine transformation. This allows for the fact that heterozygosity data is proportional.

Because the proportion of the population that is sampled will affect parameters such as allelic diversity, because the probability that comprehensive sampling has occurred will increase with increasing sample size (relative to the total population), resulting in a greater likelihood that all alleles are represented. Therefore, because of the disparity between the sample sizes, 20 versus 74, and the resulting dependency effects on allelic diversity, a direct comparison would confound the ANOVA analysis. In order to prevent this, a permutation approach was taken, whereby a random subset of 20 individuals was repeatedly sampled, without replacement, from the 74 individuals from the modern population. Permutations were performed over complete genotypes, as opposed to over loci, in order to avoid assumptions about the interdependence of loci and without replacement to avoid assumptions about the frequency distributions of alleles in the sampled population. The average number of alleles per locus was then calculated from all the artificially generated sample datasets in order to allow an unbiased ANOVA analysis of the 1896 and modern populations.

8.2.4.4. *Testing for population structuring across the temporal scale*

The two populations (modern day and 1896) were directly compared for variation between their allele frequency distributions at each locus; this will elucidate whether temporal genetic structuring has occurred. These changes can be as a result of a number of stochastic population events, such as genetic drift, gene flow, mutation, or extinction and recolonisation (Viard *et al.*, 1997; Lugon-Moulin *et al.*, 1999). If these factors have not had a significant effect, then limited genetic differentiation will have occurred. The mutation rate of microsatellites has been estimated to be in the order of 10^{-5} to 10^{-2} (per generation) (Dallas, 1992; Weber & Wong, 1993; Jarne & Lagoda, 1996). When it is considered that a sum total of 204 generations will have passed since 1896, mutational events are likely to have occurred, causing individuals within the population to have reached allelic states other than those carried by their ancestors.

In order to test for changes over time, the microsatellite data for both temporally defined populations were compared using the exact G-test, as advocated by Goudet *et al.*, (1996). When carried out on genic tables, this was found to be the most powerful test for genetic differentiation, particularly when sample sizes are unbalanced (Goudet *et al.*, 1996), as is the case here. A total of 5000 permutations of genotypes among samples were performed to assess the significance of the temporal differentiation. Computations were performed using *FSTAT* 2.9.1. (Goudet, 2000) updated from Goudet (1995).

In addition to the exact test of population, two other estimates of genetic distance were calculated for the two temporally defined populations. The first of these was θ, Weir and Cockerham's F_{ST} analogue (Weir and Cockerham, 1984), calculated using *FSTAT* version 2.9.1. This estimator is particularly relevant because it has been shown to be independent of the effects of sample size (Weir and Cockerham, 1984). The second distance measure was Slatkin's R_{ST} (Slatkin, 1995), calculated using *RST Calc* version 2.2 (Goodman, 1997). Before calculation of R_{ST}, the data was standardised for variance between sample sizes using a standardisation procedure within the *RST Calc* 2.2 package. Both F_{ST} and R_{ST} are measures of population subdivision, and as such should give an estimate of the

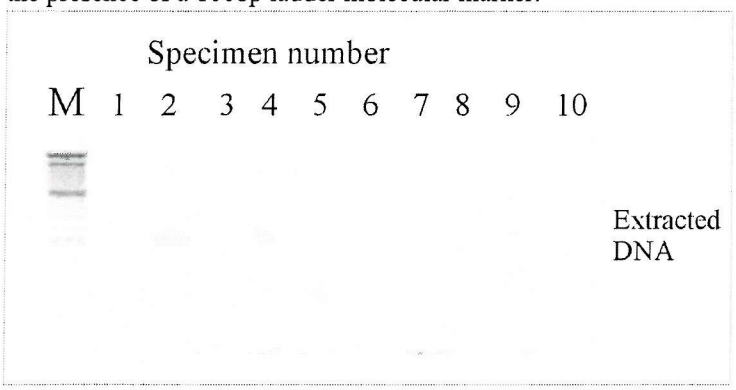
occurrence of population structuring across the temporal scale. Where the calculated values for these indices are deemed as statistically significantly different from zero, population structuring can be assumed to have occurred (see section 6.2.4.5, chapter 6 for more detailed background information about these estimators). It should be noted that as a result of the high number of null alleles detected within the data, it is possible that estimators such as F_{ST} and R_{ST}, which are based on calculated heterozygosities, will not truly reflect the temporal changes.

8.3. RESULTS

8.3.1. DNA extraction and amplification of the microsatellites

DNA was successfully extracted from both the modern and preserved *P. bellargus* specimens. The modern specimens could be visualised as a single high molecular weight band on an ethidium bromide stained agarose gel. In contrast, the DNA from the 1896 specimens was seen as a faint smear at the base of the gel, roughly 100-300 base pairs in size (see figure 8.2). This concurs with the predictions that DNA isolated from dried specimens will have undergone extensive enzymatic degradation and breakdown.

Figure 8.2. An inverted tone image of salt extracted DNA from the legs of ten *P. bellargus* individuals from 1896. The samples have been run on a 1% agarose gel stained with ethidium bromide. Lane "M" denotes the presence of a 100bp ladder molecular marker.



After optimisation of the PCR conditions, four of the five microsatellites amplified consistently and reproducible from both butterfly age groups. All attempts to optimise

Lb1/41 from the museum specimens failed; this marker was therefore not included in the analysis.

8.3.2. Statistical analysis

8.3.2.1. *Hardy-Weinberg Equilibrium*

Of the four microsatellites applied to the study, three failed to conform to the expectations of the Hardy-Weinberg equilibrium in either of the temporally separated populations. As expected, these loci (Lb1/57, Lb4/18 and Lb4/19) all displayed observed heterozygote deficits, possibly indicative of the presence of null alleles. The fourth locus, LbG2, showed no significant deviations between the calculated H_O and H_E, indicating that there are no null alleles at this locus. When the null allele frequencies were calculated for the 1896 sample (see table 8.1), all three loci affected displayed values of between 0.2 and 0.35; this is comparable with other published studies (Neumann & Wetton, 1996; Van Treuren, 1998) and also those calculated for the modern day populations (see table 8.1).

Table 8.1. Calculated values for the Hardy-Weinberg equilibrium and null allele frequencies for all microsatellite loci in each population, showing H_O, H_E, calculated probability rating and null allele frequency (n).

POPULATION	LOCUS	H_O	H_E	*P*-VALUE	NULL ALLELE FREQUENCY
1896	Lb1/57	0.333	0.953	<0.001	0.34
	LbG2	0.588	0.685	0.6	0
	Lb4/18	0.529	0.756	0.02	0.23
	Lb4/19	0.467	0.864	<0.001	0.21
Modern	Lb1/57	0.508	0.836	<0.001	0.25
	LbG2	0.731	0.764	0.07	0
	Lb4/18	0.667	0.828	<0.001	0.27
	Lb4/19	0.6	0.717	0.037	0.07

8.3.2.2. *Allelic evidence for a recent decline*

When both populations (the 1896 and modern (FE)) were analysed for evidence of a genetic bottleneck, only the modern population showed the allelic evidence of a recent decline. Although neither population exhibited a shift in allele frequencies away from an L-shaped distribution, the Wilcoxon signed rank test uncovered an excess of heterozygosity relative to allele number in the modern population, but only under the IAM (assigned *p-value* 0.031). The population from 1896 showed no evidence for a recent contraction in size under any of the mutation models tested (Wilcoxon signed rank test: IAM *p-value* = 0.16, TPM = 0.91, SMM = 0.94).

8.3.2.3. *Testing for population structuring across the temporal scale*

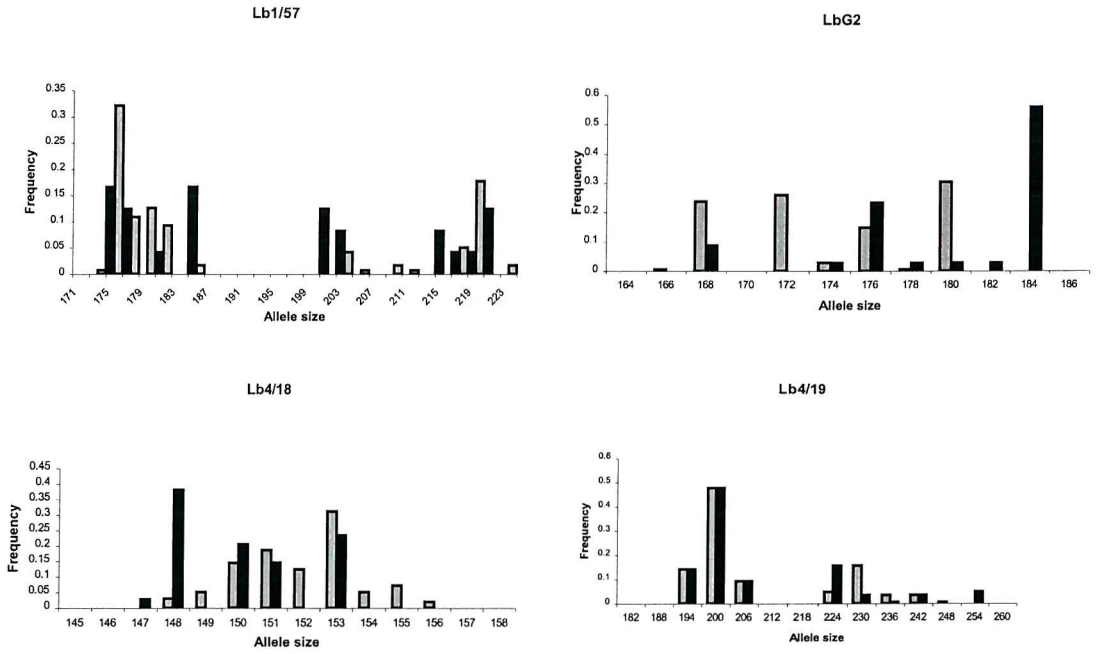
The two temporally separated populations were analysed using an exact test of population differentiation (G-test (Goudet *et al.*, 1996)). The results of this analysis indicated that after 1000 randomisations, three out of the four loci (Lb1/57, LbG2 and Lb4/18) showed probability levels indicative of highly significant differentiation across the temporal scale (exact G-test for each loci: $p = 0.001$). However, one loci, Lb4/19 showed no significant differentiation across the timescale (exact G-test, $p = 0.2$), the similarity between the two temporal allele distributions is clearly shown in figure 8.3. Nevertheless, the overall loci differentiation was deemed as highly significant across the timescale (exact G-test, $p = 0.001$), indicating distinct divergence and genetic drift. . This indication of differentiation was not altered by a sequential Bonferonni correction on the results.

Pairwise F_{ST} and R_{ST} values were calculated for the two temporal populations, to test for genetic subdivision and structuring across the timescale. Both indices indicated a highly significant degree of structuring; the calculated F_{ST} value was 0.0942 with an associated probability of 0.001. The equivalent values for R_{ST} were 0.216 and <0.0001, respectively.

8.3.2.4. *Temporal changes in the allelic data*

It is apparent from figure 8.3 that there have been distinct changes in the allelic distributions between 1896 and 1999, probably because of rapid genetic drift driven by the demographic contraction and re-expansion of the population. The specific alleles found in the modern and 1896 populations at Folkestone, two other local populations (Lydden and Wye), and a sample of the total UK population (data taken from chapter 6) are shown in table 8.2.

Figure 8.3. Combined frequency histograms of allele sizes (in bp) for all four loci in the modern (grey) and 1896 (black) populations of *P. bellargus* at Cherition Hill.



At each locus, the alleles present in the modern (FE) population represent a subset of the alleles present in the total UK population (see table 8.2). In the museum specimens, we found a total of 30 alleles, of which 23 were identical to the 37 alleles currently extant in the population at Folkestone (table 8.2). Of the fourteen alleles missing from the 1896 sample, seven are extant in populations close to FE (Lydden and Wye (see table 8.2)), whereas the remaining seven have not been identified in any of the local populations analysed. In addition to most of the currently extant alleles, seven alleles were identified from the museum specimens that are not found in the current Folkestone population

(table 8.2). Two of these alleles are found in the other Kent populations, and four are present in at least a subset of other UK populations. However, one allele at LbG2 was unique to the 1896 museum specimens, and had not previously been identified in *P. bellargus*, even during an extensive population survey of 1173 butterflies from 26 current UK populations (see chapter 6). This allele was far from rare in 1896, when it was the most common allele identified at this locus, with a frequency of 0.56.

Table 8.2. Individual alleles found at each microsatellite locus for each of the two populations analysed from Cheriton Hill at Folkestone, Kent (1896 and modern) and those present in two other populations close to this site (Lydden and Wye) (data taken from chapter 6) plus also the whole of the modern UK population (data also taken from chapter 6). Mean (and Standard Error) of expected heterozygosity (H_E) over loci is also shown.

POPULATION	LOCI			
	Lb1/57	**LbG2**	**Lb4/18**	**Lb4/19**
1896 (FE)	BC EFGH **LMNO**	B DEFG***H/***	AB DE G	ABCDEFG **I**
Modern (FE)	ABCDEF HIJK MN P	ABCDEFG	BCDEFGHIJ	ABCDEFGH
Lydden (Kent)	BCDE H MN	BC E G	BCDEFG IJ	ABCDEFGH
Wye (Kent)	CDE H MNO	BCD. G	ABCDEFG	ABC FG
UK (Modern)	ABCDEFGHIJKLMNOP*	ABCDEFGH+	ABCDEFGHIJ#	ABCDEFGHI**

Note: Bold letters indicate alleles found in the museum specimens that appear to have been lost from the modern population at FE. Italic letters indicate alleles that are unique to the FE population prior to its decline.

* In addition, alleles QRSTUVWXY are present

\+ In addition, allele J is present

\# In addition, alleles K and L are present

** In addition, alleles JKL and M are present

The calculated mean for the expected heterozygosity (H_E) in the modern population at FE (0.786) was slightly lower than that of the 1896 population (0.816), which had a higher H_E than any other Kent population (see table 8.3). This value was also found to be higher than any other modern population in the UK (data taken from chapter 6). However, when this was submitted to a one-way ANOVA after arcsine transformation, no significant difference was found among any of the means ($F_{3, 12}= 2.01, p = 0.16$). The expected Hardy-Weinberg heterozygosity was used rather than the observed levels because in general H_E is considered a better index of genetic variability (Nei & Roychoudhury, 1974).

After permuting genotypes from the modern population samples, an estimated value for allele number was obtained for each locus, that was unbiased when compared with the smaller sample obtained from the 1896 population (see table 8.3). When the computed values were compared with the 1896 and other Kent populations, the mean number of alleles per locus was lower for all the modern populations (FE: 6.75, LY: 6.5, WYE: 5.75) compared with the 1896 population (7.5) (see table 8.3). However, when this was submitted to an ANOVA, no significantly difference was found among them ($F_{3,12} = 0.63, p = 0.61$).

Table 8.3. Expected heterozygosity (H_E), number of alleles (n), and effective number of alleles (N_E), for each microsatellite locus within each population. Standard errors for mean heterozygosities (H_E) and mean numbers of alleles are indicated in parentheses.

	POPULATIONS								
	FE(1896)		FE(modern)			LY		WYE	
LOCUS	H_E	n	H_E	n	N_E*	H_E	n	H_E	n
Lb1/57	0.953	10	0.834	13	8.6*	0.580	6	0.509	7
LbG2	0.685	7	0.764	7	4.4*	0.690	4	0.617	4
Lb4/18	0.756	5	0.828	9	6.9*	0.741	8	0.823	7
Lb4/19	0.864	8	0.717	8	7.1*	0.853	8	0.576	5
Mean H_E	0.816		0.786			0.716		0.631	
SE of H_E	(0.06)		(0.03)			(0.01)		(0.09)	
Mean Allele/Locus	7.5		9.25		6.75*	6.5		5.75	
SE of allele/locus	(0.9)		(1.1)		(0.87)*	(0.8)		(0.6)	
Effective sample size	20		74		20*	49		32	

* Estimated values calculated by the permutation method.

8.4. DISCUSSION

Although this chapter only represents a pilot study into the potential application of museum specimens to population studies, the resulting work is unsurpassed in its focus on a single temporally and geographically defined historical population. The detailed information accompanying the *P. bellargus* museum specimens has enabled their geographic origin and sampling date to accurately be ascertained, allowing direct measurements of allele frequency changes within a single population over a temporal scale. Although other studies have also concentrated on temporal variations (e.g. Bouzat *et al.*, 1998a; Nielsen *et* al., 1999; Groombridge *et al.*, 2000), their studies have been hampered by a lack of specimens available and the imprecise data accompanying them,

resulting in an inability to achieve temporal consensus within the historical data. In some cases, these studies have also combined individuals from geographically separated populations (e.g. Bouzat *et al.*, 1998a). This kind of partitioning within the sampled 'population' may lead to erroneous calculations of population heterozygosity, because of the Wahlund effect, where mixing of separate populations can cause a deficiency of heterozygotes.

The Hardy-Weinberg analysis of both temporally defined populations shows significant deviations from expectations for all but one of the loci (LbG2). The indication of a heterozygosity deficit in Lb1/57, Lb4/18 and Lb4/19 corresponds with previous observations that these three loci have elevated levels of null alleles (see chapters 5 and 6). Whilst these heterozygote deficits could be ascribed to inbreeding or population structuring (Lade *et al.*, 1996; Paxton *et al.*, 1996), these effects would be unlikely to account for the extent of the deviation indicated here (Palo *et al.*, 1995). Not only this, but there is also inconsistency among loci, whereby one locus remains unaffected, population effects, such as structuring or inbreeding, would tend to affect all loci simultaneously. This strongly suggests the cause of the disequilibrium to be the presence of null alleles at Lb1/57, Lb4/18 and Lb4/19, rather than population subdivision. For both temporal populations, the calculated null allele frequencies are within the same range, a level that has also been observed in other studies (Neumann & Wetton, 1996; Van Treuren, 1998). The parallels between the Hardy-Weinberg equilibrium and null allele frequencies for both temporal datasets tend to support the conjecture that both samples were removed from single populations.

Whilst the population data for the modern population at Folkestone indicates a demographic decline in the early 1990's, the allelic evidence is somewhat less conclusive. When the data was analysed for an excess of heterozygosity relative to allele number, a proposed characteristic of a population bottleneck (Hedrick, 1986; Cornuet & Luikart, 1996), significant excess was only found under the IAM (assigned $p = 0.031$) rather than the TPM and SMM. This may indicate constraints on the analytical power of the test, such as: (i) The presence of null alleles in the population may have a

confounding effect on the analysis (Cornuet & Luikart, 1996; Le Page *et al.*, 2000), (ii) the disjunctive patterns of allele size distributions present for most loci (see figure 8.3) may indicate a departure from the SMM, possibly making this test uninformative. (iii) Alternatively, four loci may be too few to allow the test to achieve sufficient analytical power (Cornuet & Luikart, 1996). (iv) It must also be considered that the historical data on population changes is anecdotal, as a result of having not been collected under any methodological consensus by amateur entomologists.

Whilst the comparisons of allelic diversity are indicative of significant temporal structuring at Folkestone, the results are inconclusive about any losses of genetic diversity since 1896. With the exception of Lb4/19, there are significant differences between the allele distributions for all loci, as determined by the G-test, and temporal change was also indicated by significant pairwise F_{ST} and R_{ST} values. Therefore, it seems that rapid genetic drift has occurred, which may have been intensified by a decline and re-expansion of the population. This data is in accordance with other similar studies of temporal changes in allele frequencies associated with population declines (Taylor *et* al., 1994; Bouzat *et* al., 1998a; Groombridge *et al.*, 2000).

The modern population sample contains a total of 37 alleles, of which 23 were also found within the 1896 population. The fact that all of these modern alleles are shared with other UK populations suggests they may also have been present before the demographic contraction, but are not represented in the museum sample because of its limited size. Alternatively, of the fourteen alleles missing from the 1896 sample, seven are extant in populations close to Folkestone (Lydden and Wye (see table 8.2)); these alleles could have entered the population since 1896, by immigration (gene flow) from these local sources. However, seven of the modern alleles missing from the 1896 population are also absent from the local populations, their current presence at FE almost certainly reflects a combination of the rarity of these alleles (all have frequencies <0.02 in the current FE population, range 0.007 – 0.017) with the small sample size of the 1896 population (rare alleles are less likely to be identified from a smaller sample size). Even with the disparity between the sample sizes, there are seven alleles present in the 1896 population that have

not been observed from the modern population, of which five are also not found in any of the local populations (WYE & LY). Although, with the exception of a single allele at LbG2, six are present in at least a subset of the other UK populations. This allele (184bp) at LbG2 has never before been identified in *P. bellargus*, yet has a frequency of 0.56, and represents the most common allele identified at this locus in 1896. This drastic shift in allelic distribution is unlikely to have occurred in a population at equilibrium, and can only be attributed to one or more major demographic changes, such as bottlenecks or cycles of population extinction and recolonisation, because these events will cause population differentiation via rapid and intense genetic drift (e.g. Wright, 1940; Slatkin, 1977; Wade & McCauley, 1998; Whitlock & McCauley, 1990). Not only this, but anything other than a severe bottleneck or extinction event would be unlikely to remove such a common allele, because theoretical predictions suggest that declines will generally only remove rare alleles (Nei *et al*, 1975; Leberg, 1992).

Whilst the values for H_E and mean number of alleles per locus were higher in the 1896 population than for any other modern Kent population, and the 1896 value for H_E was also higher than any other modern UK population, none of these comparisons presented a significant difference under an ANOVA. This observed trend for a lower H_E and mean number of alleles per locus in modern populations may indicate a long-term effect from the 1978 bottleneck that the species encountered. The results of this study are suggestive that the Folkestone population may have originally had higher levels of genetic diversity, that were consequently lost either through the demographic contraction of the late 1970's or that of the early 1990's. Bouzat *et al.* (1998a) demonstrated a similar trend for the greater prairie chicken, a species that underwent a bottleneck of fewer than 50 individuals in 1993. This study also used museum specimens to provide clear evidence for the prediction that bottlenecks will have a large impact on population allelic diversity (Nei *et al.*, 1975; Leberg, 1992), particularly for alleles at low frequencies.

The lack of statistical significance for changes in diversity may be attributable to a number of reasons, of which the most likely is that the declines were less severe than suggested by direct observations and transect surveys (e.g. Thomas, 1983a). The ecology

of *P. bellargus* is such that even slight changes in weather may affect the observed number of butterflies, and hence transect counts. For example, during the fieldwork, it was observed that minor events, such as increased cloud cover, is enough to discourage *P. bellargus* from flying, and once they close their wings, they are well camouflaged against their background. Not only this, but the structure of the habitat at Folkestone is such that many small populations are in very close proximity to one another, separated by just tens to hundreds of metres. Whilst a decline was identified at Cheriton Hill during 1992, a population at Folkestone Escarpment, less than a kilometre away, is suggested to have had a population peak during the same time period (Nick Johanssen, personal communication). Previous work in this thesis has already demonstrated the small, but significant, propensity of *P. bellargus* individuals to move between neighbouring populations, thus the bottleneck effect at Cheriton Hill may have been greatly weakened by gene flow from Folkestone Escarpment.

Even if the 1978 bottleneck did have severe consequences for the population at Folkestone, the rapid mutation rate of the microsatellite DNA combined with the short generation time of *P. bellargus* may have considerably diluted these effects. The study of the greater prairie chickens (Bouzat *et al.*, 1998a) was estimating the effects of a bottleneck in 1993, a much more recent event, which combined with the slower generation turnover of the species would mean that the effects are more likely to remain visible in the population.

The only alternative approach to this study would have been to use "control" populations (i.e. conspecific populations that have undergone different demographic histories) to give a general insight into the relationship between genetic diversity and population size. However, this method would have been unable to disentangle the effects of the local environment from the effects of a decline, so that factors such as natural selection may favour a given set of variants at a specific site. For example, in the study of a bottleneck in lions (*Panthera leo*) in the Ngorongoro crater (Tanzania), a larger population from Serengeti was used as a reference (Packer *et al.*, 1991). However, each of these populations may have been affected differently by the environmental conditions present

at each site. Whilst this problem may be overcome by increasing the number of reference populations (e.g. Bouzat *et al.*, 1998a), only sampling genetic variation from past generations in the same population (i.e. before the decline) would provide direct evidence concerning the role of population size in the preservation of genetic diversity.

This pilot study has clearly demonstrated the enormous potential of this experimental approach, which combined with the vast number of documented *P. bellargus* specimens available in entomological collections, would enable the elucidation of detailed information concerning the origin and route of temporal changes in allele frequencies. Not only could this address questions of ecological/conservation importance, particularly relevant to species existing in closed populations, but also towards enhancing the current understanding of mutational patterns and evolution in microsatellite DNA. Thus the value of museum specimens must not be underestimated, particularly for butterflies and other insect species, where unlike most other animals, the number of preserved specimens available for analysis is unlikely to present a limiting factor to the study.

9.0 GENERAL DISCUSSION

9.1. SUMMARY

The main aims of this thesis (see chapter 1, page 8) were to investigate a variety of aspects of the UK population structure of *P. bellargus*. Specifically, the work has considered its colonisation of the UK alongside a range of factors affecting interpopulation relationships and genetic diversity at both the local and national scale, including the effects of population size, isolation, genetic bottlenecks, and barriers to gene flow. These aspects have all been considered in terms of their relevance to the conservation of *P. bellargus* in the UK, and also for other arthropod species existing in similar metapopulation structures. In order to achieve this, I carried out a survey using molecular markers to assess genetic structure within and between carefully chosen populations across the UK range of *P. bellargus*. This chapter summarises the results from these investigations, outlines the limitations and suggests avenues for further work.

Although *P. bellargus* underwent a severe (~90%) decline between 1978 and 1981 (Thomas, 1983a; Emmet & Heath 1990; Asher *et al.*, 2001), the species is currently recovering, and has recolonised many sites where it had formerly become extinct (Emmet & Heath 1990; Bourn & Warren, 1998; Bourn *et al.*, 1999; Stewart *et al.*, 2000; Asher *et al.*, 2001). Only in extremely isolated areas, such as the Cotswolds and Bedfordshire, has the species found recolonisation difficult. This reflects the sedentary nature of the species, where mark-release-recapture studies have shown dispersal is often less than 100 metres (Thomas, 1983a; Asher *et al.*, 2001). During this project, a vast amount of data was collected concerning the number of UK populations and their locations; the results confirmed the status of the butterfly as a species that is slowly recovering some of its historical range. The fieldwork also confirmed some of the predictions about metapopulations (Gilpin & Hanski, 1991; Gilpin, 1991; Hanski and Gilpin, 1996), whereby large fluctuations occur in subpopulation size over time. This was exemplified by many of the populations, where numbers often fluctuated greatly between successive sampling generations (see chapter three). Such rapid changes in colony size can cause

problems in the estimation of population size, because conservation volunteers (from organisations such as Butterfly Conservation, English Nature and the National Trust) generally calculate these from census counts of butterflies. Therefore, whilst the data may be relevant for the generation it was taken from the population size may be quite different in successive generations, particularly if environmental conditions change. For this reason, a cautious approach was taken when applying the chronicled population data in this study.

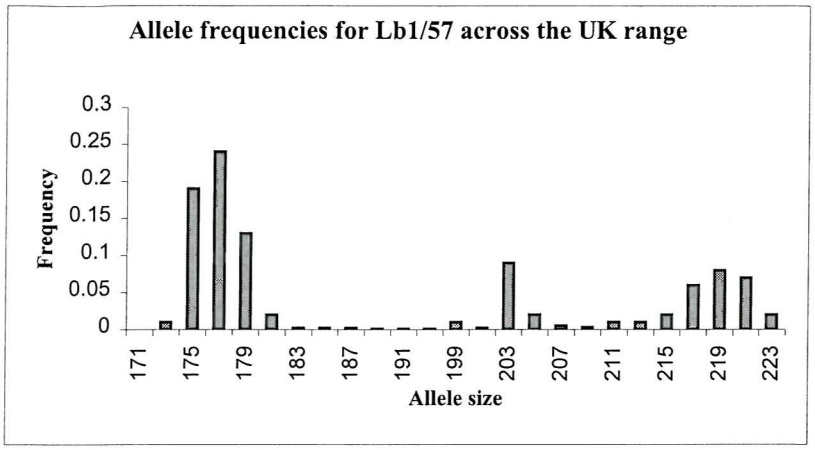
The volunteers who supplied the population data to conservation organisations are often themselves keen amateur entomologists, with a knowledge of the habitat suitable for *P. bellargus*. It was noted from the raw data that there were records of occasions when *P. bellargus* had been unofficially released at various sites, with no reference to the source population that the butterflies were removed from; it is likely that unrecorded releases have also taken place. Obviously this has important implications to a population genetics study, where the phylogenetic relationships among populations are being reconstructed, and may explain why some of the population relationships remained poorly resolved under the microsatellite analysis.

A potentially exceptional example of this kind of anthropogenic colonisation may have been revealed by the mitochondrial DNA analysis (see chapter four). Although the molecular data cannot suggest a method by which the UK colonisation was mediated, it does suggest that it was probably a recent event that was centred on very few female butterflies. It has been noted that *P. bellargus* was not recorded in the UK until 1775 (Harris, 1775; Emmet & Heath, 1990), and then appears to have undergone a rapid invasion of the UK. It may be coincidence that this colonisation happened to occur during a period of escalating interest in lepidopteran natural history, but it is also possible that an enthusiastic collector released a few mated females from France on some chalk downland in southern England, initiating its UK colonisation. Of course, this explanation is very speculative, and it may simply be a matter of a chance natural event bringing the species to the UK, such as a mated female blowing across the English Channel in a storm.

It is hard to find a more plausible explanation than a recent colonisation event to account for the patterns of mtDNA diversity in France and England, because of the lack of sequence variation in the UK relative to that found in other lycaenid species (Taylor *et* al., 1993; Brookes *et al.*, 1997). These studies have clearly demonstrated a propensity for mutation within the control region. Although the occurrence of a severe UK bottleneck is an alternative scenario, it is far less likely to be the explanation. In order to obtain the current level of variation observed for *P. bellargus*, the decline would have literally had to wipe out all but a few butterflies at a single site. Anything less severe would have resulted in the fixation of different alleles in geographically separated populations.

In contrast to the mtDNA evidence, the microsatellite data (chapter six) demonstrates a high level of interpopulation variability in the UK. There are a variety of possible explanations for this paradox. Firstly the maternal inheritance pattern and lack of recombination of mtDNA results in a marker that, unlike microsatellites, ignores male mediated gene flow and variation. A number of *P. bellargus* field studies show that males are generally more mobile than females (Emmet & Heath, 1990; Thomas, 1983a; Rusterholz & Erhardt, 2000), so males may have a higher likelihood of occasionally crossing the English Channel, leading to gene flow that would increase microsatellite diversity but have no influence on mitochondrial variability. An observation from the microsatellite data, which may support this conjecture, is that most of the allele distributions do not show the normal distribution commonly associated with microsatellite DNA, but instead they display a more disjunctive pattern of alleles (e.g. see figure 9.1 and 9.2). This could infer occasional male mediated gene flow from France, whereby a small number of diverse microsatellite alleles have arrived in the UK, which have each mutated over time to give rise to distinct bell shaped distributions, hence the roughly tri-modal distribution observed in Lb1/57 and Lb1/41. Obviously, if the microsatellite analysis had included one or more sizeable French populations and the allele distributions were equally disjunctive, then this possibility could be ruled out. An alternative explanation for the allele patterns could be mutations in the flanking sequences around the microsatellite, causing large jumps in the allelic distribution.

Figure 9.1. Allele frequency distribution for Lb1/57 across the UK range.



Another possible reason for the disparity between the levels of diversity between the mitochondrial and microsatellite DNA is the high mutation rate of microsatellites. At least ~450 discrete generations of *P. bellargus* have passed since its putative colonisation of the UK (between 1710 and 1775), and the mutation rate of microsatellites has been estimated to be in the order of 10^{-5} to 10^{-2} (per generation) (Dallas, 1992; Weber & Wong, 1993; Jarne & Lagoda, 1996), there has been ample opportunity for variation to accumulate at each locus. Whatever the basis for these observations, other studies of colonisation events have also observed a similar disparity between the patterns elucidated by various types of DNA marker (e.g. Provan *et al*, 1999).

The isolation of five microsatellites from *P. bellargus* represents the most successful piece of work of this kind (chapter five). No other study has managed to obtain more than four microsatellites for any lepidopteran species (Keyghobadi *et al.*, 1999). Nevertheless, the five markers obtained have aspects in common with all other lepidopteran microsatellites isolated, including the high frequency of compound or interrupted repeat motifs and the pervasive presence of null alleles at most loci (Palo *et al.*, 1995; Meglecz *et al.*, 1998; Meglecz & Solignac, 1998; Keyghobadi *et al.*, 1999). These observations infer that lepidopteran microsatellites are often associated with areas of very simple sequence, which are prone to mutational events (Neve & Meglecz, 2000). When mutations occur within primer binding sites, they will cause elevated levels of null alleles. It is hard to theorise about why particular taxa should be affected in this manner,

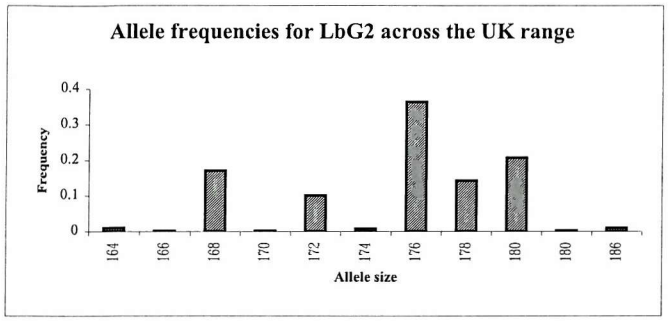
but suffice to say that it implies a level of complexity within lepidopteran genomes that is not found in most other species.

The presence of null alleles means that caution must be taken when analysing population data. For instance, coefficients such as F_{ST} and R_{ST}, which are calculated using observed heterozygosity, are likely to be confounded by the observed homozygosity excesses, leading to erroneous conclusions from the data. The lack of consensus over repeat motif is also likely to cause problems with many of the estimators applied to microsatellite data, because these indices generally assume a unanimous mutation model amongst loci. However, of the five microsatellites isolated for this thesis, two are represented by perfect repeats, though of differing motif length: $(CA)_n$ and $(GACGGT)_n$, whilst the other three are either compounded, interrupted or both. These distinctions make it impossible to unanimously allocate a particular model of mutation to all five loci, because a perfect dinucleotide repeat, such as Lb1/57, will probably follow a stepwise mutation model, whilst a compounded locus, such as LbG2, may mutate in a manner similar to the infinite allele model. This is because LbG2 consists of five compounded repeat motifs ($(CA)_n$, $(GC)_n$, $(GA)_n$, $(GAGC)_n$, and $(CGCA)_n$), thus slippage could occur during replication at more than one of these sites, preventing a consensus in size change among descendant alleles. The allele frequency distribution for this locus clearly demonstrates a propensity for mutation at one, or more, of the tetra-nucleotide repeats, because the alleles are commonly separated by four nucleotides, but occasional dinucleotide changes must occur to give rise to the low frequency of intermediate alleles observed between them (see figure 9.2). Similarly, Lb4/18 consists of a dinucleotide repeat ($(CA)_n$) compounded against a single nucleotide repeat ($(T)_n$). The characteristics of these loci make it extremely difficult to make assumptions about allelic relationships. .

This disparity amongst loci precludes the application of measures based on the stepwise mutation model, such as R_{ST}, and it is for this reason that numerous approaches were taken to the statistical analysis of the microsatellite data, where no single class of distance estimator was relied upon and an effort was made to include Euclidean measures. Nevertheless, the results from the population analysis chapter (chapter six) were

cohesive, with the same inferences being drawn from all of the analytical approaches, thus providing an extremely robust picture of the current population structure of *P. bellargus* in the UK.

Figure 9.2. The allele frequency distribution for LbG2 across the UK range.



The overall indication from the population analysis (chapter six) is that *P. bellargus* occupies a rather subdivided distribution in the UK. The level of gene flow among the populations varies depending on the geographic scale, i.e. local populations (defined as occurring within 37.5 kilometres of one another) have a much higher potential for gene flow than populations separated by large expanses of unsuitable habitat (>37.5km), where no clear interpopulation relationships could be discerned. However, these widely separated populations may actually be more distantly related than these data indicate, because where migration is low, highly polymorphic markers (such as microsatellites) can lead to an underestimation of population divergence (Balloux *et al.*, 2000). This underlines the need for a second genetic marker to be used in parallel with microsatellites, for example a sequenced mitochondrial region, because this would have been invaluable for verifying the trends shown by the five microsatellites. However, because this marker was found to be devoid of variation within the UK, no comparisons could be made.

The data from the five microsatellites provided evidence for a link between small population size and reduced diversity. This tendency has been found in numerous captive populations (e.g. Ralls & Ballou, 1986; Laikre & Ryman, 1991; Saccheri *et al.*, 1996), but although low diversity has been observed in small wild populations, to date only a few recent studies have found a relationship between heterozygosity and population size

(e.g. Luijten *et al.*, 2000; Podolsky, 2001). Inevitably, a lower level of heterozygosity infers the presence of inbreeding (Frankham, 1995a), leading to concerns over the effects of inbreeding and conservation of these populations.

Whilst an effect of population size was clearly present, no effect could be identified from either population isolation or a history of recent declines on genetic diversity. The lack of an association between these variables may result from inadequacies within the data rather than the absence of a link between them. For example, the history of declines was taken from numerous anecdotal data sources that, as already discussed, are likely to be affected by confounding factors. The population isolation data was also subject to errors, because the level of isolation for each population depended only on its proximity to the other 25 populations used in this study. In order to obtain a more accurate measure, a detailed survey of all UK populations of *P. bellargus* would have been necessary; but this was not possible within the time constraints of this project.

This association between population size and genetic diversity may have affected the results of the temporal investigation of the affects of a bottleneck at Folkestone (chapter eight). The work in this chapter shows that whilst the chronicled data suggest a recent severe population bottleneck, the slight trend for reduced genetic diversity relative to pre-bottleneck museum specimens from the same site was not significant. This is one of the largest populations of *P. bellargus* in the UK, occupying an expansive area of downland. The large population size may have encouraged increased levels of diversity, masking the effects of the decline. The two factors are almost certainly intermingled to some extent, whereby a small population occupying limited habitat resources will be more vulnerable to declines than a large population with extensive habitat space.
This application of museum specimens to an investigation of historical genetic diversity (chapter eight), and the associated preliminary work leading up to it (chapter seven), is the most extensive investigation of its kind in any arthropod species. Whilst a few studies have touched upon the potential for this kind of analysis in insects (e.g. Bulat *et al.*, 1992; Hammond *et al.*, 1996; Meglecz *et al.*, 1999; Zakharov *et al.*, 2000), none have applied the techniques to specimens older than one or two decades, or to the elucidation

of quantitative genetic data. Nonetheless, the work in this thesis has clearly demonstrated the ability of molecular techniques to reveal temporal genetic changes within a population, using dried specimens preserved for as long as 105 years. Whilst this field in genetics is currently in its infancy, and to date only a handful of studies have successfully been carried out (e.g. Bouzat *et al.*, 1998a; Groombridge *et al*, 2000), this pilot study illustrates the enormous potential for the vast entomological collections around the world to provide a unique opportunity for detailed examination of temporal genetic change.

9.2. THE IMPLICATIONS FOR CONSERVATION

The emerging patterns from this study are that the sedentary nature of *P. bellargus* does result in isolation by distance, although typically at a higher geographic scale than might be expected from direct observational studies of individual butterflies. It appears that gene flow can be very persistent, though almost certainly via a stepping-stone model rather than long distance exchange of individual butterflies, thus enabling genetic exchange across distances much greater than butterflies could be expected to traverse: almost 40 kilometres rather than a few hundred metres. This kind of population interconnectivity is extremely vulnerable to the effects of habitat fragmentation, a disturbing fact when it is considered that increases in 'agricultural improvement' has already resulted in an 80% loss of calcareous grassland since the 1940s (Nature Conservancy Council, 1984) and the current estimated average size of calcareous grassland patches is only 7.7ha (Blackstock *et al.*, 1999). Whilst attempts have been made to reverse this trend, via various restoration schemes, these almost never result in habitats as diverse as ancient ones and a complete community may take hundreds of years to develop, or not at all if the constituent plants and animals are no longer available in the vicinity (Gibson, 1995; Gilbert & Anderson, 1998). For example, *Hippocrepis comosa*, the host plant of *P. bellargus*, is extremely slow to spread, often taking decades to recover from periods of abandonment, where it has been outcompeted by taller grasses. In many sites it has disappeared altogether and its ability to regenerate from buried seeds is uncertain (Asher *et al.*, 2001). Whilst these restoration schemes comprise over 1,000 hectares of chalk downland, so far these areas have mainly been recolonised by common butterflies, such as *Maniola jurtina*, but very little by downland specialists (Asher *et al.*, 2001). Therefore it seems unlikely that these schemes will have much impact on *P. bellargus* in the near future, and whilst the species has recovered around 20% of its range since the severe decline of 1978-1981, this still only amounts to around 50% of its pre-1940 range (Asher *et al.*, 2001), with habitat fragmentation almost certainly limiting its recovery. Another especially worrying aspect of this is the possibility that there may be a lag time between habitat loss and fragmentation and the eventual extinction of a species from a region (Hanski, 1988).

Improvements in management policies for many calcareous grassland areas are now making a significant contribution to protecting existing habitats for butterflies (Warren & Bourn, 1997). However, these schemes are not without their problems, because the grazing schemes adopted cannot suit all of the species requiring protection. For example, many butterflies benefit from autumn or winter grazing, whereas *P. bellargus* relies on spring grazing to provide the short turf necessary for its survival (Butterflies Under Threat Team, 1986). Many species, such as the Marsh Fritillary (*Euphydryas aurinia*), the Marbled White (*Melanargia galathea*) and the Lulworth skipper (*Thymalicus action*), benefit from reduced grazing pressure (Pearman *et al.*, 1998), and therefore whilst management schemes encourage some areas to be suitable for *P. bellargus*, the requirements of other species results in managed areas that are not conducive to colonisation.

Many studies have already demonstrated that habitat fragmentation and isolation has a severe impact on butterflies, for example Thomas, *et al.* (1992 & 1998) have shown that many suitable sites remain unoccupied by habitat specialists because they are too small or too isolated for population survival. The metapopulation structure adopted by many lepidopteran species is particularly vulnerable to these effects, whereby if the effects of fragmentation reduce movement between patches, persistence of the entire metapopulation may be at risk. It has been suggested that a minimum of 20 connected patches is necessary to ensure metapopulation survival (Thomas, 1995).

Whilst many butterflies, including *P. bellargus*, also occur widely across Europe, it is likely that many of these species have adapted to the local conditions in the UK. The fact that in the UK, *P. bellargus* has a preference for closely grazed areas, as opposed to taller pasture in parts of Europe (Thomas, 1983a), may be evidence for this. Therefore, the conservation of the species in the UK may represent the preservation of habitat specific genes, vital to future evolution and survival. Similarly, slight regional habitat variations in the UK may also result in local adaptation, possibly explaining the failure of some attempts to reintroduce *P. bellargus* to regions where it had formerly become extinct, such as Gloucestershire (Barker, 1990 & 1991; Meredith, 1992; Joy, 1997). Not only

this, but the maintenance of genetic variation within populations is also important to ensure robust populations that are able to adapt and survive in the long term (Frankham, 1995a), a theory exemplified by Saccheri *et al*'s (1998) demonstration that inbreeding increases the likelihood of extinction in butterfly metapopulations. This accentuates the importance of the inferred relationship found between population size and heterozygosity in *P. bellargus*, whereby small populations appear to be prone to the effects of inbreeding (as deduced from their reduced heterozygosity (Frankham, 1995a)). This relationship may result from the observed susceptibility of small populations to declines caused by climatic and demographic effects (Kuussaari *et al.*, 1998) and will undoubtedly compromise the ability of smaller populations to adapt to environmental change. Therefore, it seems likely that these populations of *P. bellargus* will have an increased tendency to become extinct, emphasising not only the importance of preserving suitable habitat, but that it is crucial to ensure that these areas are big enough to sustain large, healthy populations.

Therefore, it seems that there are two important requirements for conserving populations of *P. bellargus* in the UK; these are to maintain population connectivity within regions and to encourage the preservation of large habitat areas. Gene flow is a vital factor in maintaining population genetic diversity, because the susceptibility of the species to demographic declines would otherwise gradually impoverish the populations of variation, leading to the eventual fixation of mildly deleterious alleles (Lande, 1994). Therefore small, isolated populations have the highest risk of extinction because of the combined effects of their innate lack of genetic variability and deficiency of new alleles entering the population via gene flow. This corresponds to the observation by Frankham (1995a) that endangered populations and species have a tendency to display lower genetic diversity.

9.3. FUTURE WORK

The field of population genetics has been the subject of hundreds of publications, particularly since the advent of sophisticated molecular techniques. The availability of selectively neutral markers, which undergo Mendelian segregation, has allowed geneticists to routinely test many of the theoretical predictions about genetic change. To date, literally hundreds of species have been investigated in this manner, particularly those where wild populations are suggested to be threatened or endangered. In the vast majority of these species, low genetic diversity has been found (Frankham, 1995a). However, without adequate reference populations, it is impossible to define the nature of this relationship – whether inbreeding has reduced reproductive fitness resulting in endangerment or the low population size associated with endangerment has resulted in low genetic variation.

The application of molecular analysis to museum specimens (chapter seven) should be able to elaborate on this relationship, and whilst the work in this thesis was only a pilot study, the results demonstrated the general applicability of the technique and identified a trend for reduced genetic diversity subsequent to a decline. The vast availability of preserved specimens from populations across the UK, collected prior to the 1978-1981 bottleneck could provide enormous insights into the effect that a severe decline has on the genetic variability and fitness of populations. If the data acquired from museum specimens is analysed with respect to the levels of diversity in the modern populations, then the effects of population declines may become apparent. This kind of approach may also have implications for increasing our understanding of microsatellite mutation patterns, whereby allelic variation over a temporal scale can be analysed to infer what kind of changes have occurred.

The mitochondrial analysis chapter has provided some fascinating insights into the colonisation of the UK by *P. bellargus*. However, there are a number of ways in which this work could be further extended, such as by analysing more French specimens, or backing up the data using a second molecular marker. All eight of the French specimens

analysed were captured in southern France. Analysis of more French specimens could make the study more informative, particularly if specimens from the more northern areas were used, such as the calcareous downland around Calais, where the UK population probably originates. It is possible that the species becomes increasingly impoverished of diversity towards the northern limits of its range, perhaps because of climatic pressure resulting in population instability encouraging the loss of diversity. This could be examined by analysing individuals from a north to south transect across France. The application of a second molecular marker to the study, either another mitochondrial region or microsatellites, would strengthen the inferences drawn about the colonisation process. The Cytochrome-B and COI regions of the mitochondrial genome in Lepidoptera have been observed to display variation among and within populations, particularly at the third codon positions (Zhang & Hewitt, 1997; Joyce & Pullin, 2001). If the same results were obtained by sequencing one of these genes in parallel with the control region, a more robust case for a recent colonisation could be made. As already discussed, the analysis of microsatellite DNA in the French populations may reveal less disjunctive allele distributions relative to those in the UK, which would also support the idea of a recent UK founder event. However, it is unlikely microsatellite DNA could infer any relationship between the French and UK populations, because it has already been demonstrated that populations within the UK that are separated by large geographic distances show no clear relationship to one another (chapter six).

9.4. CONCLUSIONS

The work detailed in this thesis is not only an evaluation of the current population structure of *P. bellargus* in the UK, but also the effects of a number of other influencing factors. These include the colonisation of the species in the UK, the effects of population declines, and the potential for temporal genetic change. This study is one of the most detailed analyses of factors affecting population structure in a lepidopteran species to date. Importantly, the work has revealed the applicability of preserved entomological specimens to molecular analysis, and a successful method for isolating lepidopteran microsatellites. Although this work represents another step in understanding aspects of population genetics, there is still much to learn about the consequences of demographic and stochastic processes, and how these effect population survival and genetic fitness.

10.0 APPENDICES

Appendix I.

Silver Staining Protocol

Acrylamide Gel and Silver staining protocols

For a 6% gel.

Total volume for 4 gels	150 mls170 mls	1 gel	
lOX TBE buffer	15 mls	17 mls	4.25 mls
Acrylamide	30 mls	34 mls	8.5 mls
Distilled Water	105 mls	119 mls	29.75 mls
APS 10%	1.0 mls	1.13 mls	282.5µl
TEMED	130µl	130 µl	32.5 µl

Mix the TBE, Water and Acrylamide (care neurotoxic) together.
Swirl the flask to degas the contents
Add the APS and TEMED and pour the gel.
Place the combs and cover with cling film. Bubbles can be removed by shaking the comps.
Leave for at least 1 hour to polymerise the gels. Remove the combs after at least half an hour and rinse the wells with buffer to remove any unpolymerised acrylamide.
Run at 3OmA per gel en 1X TBE

Silver Staining

Make up 3 fresh solutions For 4 gels

Buffer A Make up 2 lots as each gel is rinsed twice.

Distilled Water	360 mls
Absolute Ethanol	40 mls
Acetic acid	2 mls

Buffer B

Distilled water	400 mls
Silver Nitrate	400 mg

Buffer C

Distilled Water	600 mls
Sodium Hydroxide (NaOH)	9g
Sodium Borohydride ($NaBH_4$)*	60 mg
Formaldehyde	2.4 mls

*Always add the NaBH~ to the dissolved sodium hydroxide. This is because the $NaBH_4$ will liberate hydrogen when added directly to water which can be explosive.

Wash in buffer A for 4 minutes. Discard and repeat.
Wash in buffer B for 10 minutes, then was twice in distilled water.
Wash in Buffer C until the bands appear. Rinse with water and leave in water for 10 minutes.
Seal the gel in Polythene.

Appendix II

Calculated values for the Hardy Weinberg Equilibrium, F_{IS} and associated probabilities.

POPULATION	POPULATION SIZE	H_O	H_E	F_{IS}	*P*-VALUE
A303	35	0.49	0.68	0.275	0.0001
BD	57	0.46	0.71	0.35	0.0001
BH	51	0.37	0.6	0.377	0.0001
BC	16	0.4	0.65	0.376	0.0001
B	46	0.49	0.69	0.282	0.0001
CL	52	0.53	0.72	0.249	0.0001
CH	49	0.54	0.64	0.148	0.0003
DD	77	0.58	0.74	0.206	0.0001
FE	74	0.52	0.75	0.306	0.0001
FO	11	0.42	0.76	0.408	0.0001
HA	42	0.53	0.74	0.285	0.0001
HB	41	0.43	0.66	0.252	0.0001
LY	49	0.51	0.72	0.273	0.0001
MC	60	0.5	0.74	0.315	0.0001
MD	43	0.5	0.68	0.25	0.0001
MH	36	0.58	0.7	0.156	0.0005
M	49	0.42	0.6	0.3	0.0001
P	25	0.4	0.62	0.344	0.0001
PD	43	0.52	0.68	0.245	0.0001
R	56	0.57	0.77	0.256	0.0001
RiH	16	0.52	0.73	0.274	0.0001
RH	52	0.34	0.66	0.44	0.0001
V	43	0.43	0.68	0.357	0.0001
WHH	57	0.44	0.7	0.38	0.0001
W	47	0.43	0.79	0.449	0.0001
WYE	32	0.46	0.63	0.309	0.0001

Appendix III

Calculated null allele frequencies for all population and loci combinations

	AB1/41	AB1/57	ABG2	AB4/18	AB4/19
A303	0.25	0.20	0.00	0.28	0.00
BD	0.26	0.31	0.00	0.00	0.27
BH	0.29	0.32	0.00	0.17	0.15
BC	0.00	0.00	0.22	0.00	0.00
B	0.29	0.26	0.00	0.25	0.00
CL	0.29	0.28	0.00	0.00	0.00
CH	0.15	0.00	0.00	0.00	0.19
DD	0.30	0.24	0.00	0.23	0.00
FE	0.27	0.25	0.00	0.27	0.00
FO	0.00	0.00	0.00	0.00	0.00
HA	0.00	0.00	0.00	0.15	0.00
HB	0.00	0.00	0.00	0.00	0.26
LY	0.28	0.24	0.00	0.31	0.00
MC	0.25	0.32	0.00	0.00	0.28
MD	0.00	0.27	0.00	0.00	0.00
MH	0.00	0.00	0.00	0.00	0.00
M	0.35	0.23	0.24	0.00	0.00
P	0.21	0.29	0.00	0.00	0.00
PD	0.26	0.31	0.08	0.00	0.00
R	0.36	0.15	0.00	0.00	0.00
RiH	0.00	0.00	0.00	0.00	0.00
RH	0.35	0.38	0.00	0.23	0.26
V	0.33	0.00	0.17	0.00	0.00
WHH	0.21	0.31	0.00	0.23	0.29
W	0.28	0.29	0.00	0.26	0.23
WYE	0.27	0.26	0.00	0.00	0.00

Appendix IV

Allele Frequency data for LbG2

	ALLELE SIZE										
Population	**164**	**166**	**168**	**170**	**172**	**174**	**176**	**178**	**180**	**180**	**186**
A3	0	0	0.111	0.019	0.037	0	0.481	0.111	0.241	0	0
BD	0	0	0.074	0	0.032	0	0.638	0.191	0.064	0	0
BH	0	0	0	0	0.033	0	0.889	0.056	0.022	0	0
BC	0	0	0	0	0.031	0.063	0.156	0.563	0.188	0	0
B	0	0	0	0	0.109	0	0.315	0.13	0.424	0.011	0.011
CL	0	0	0.398	0	0	0	0.173	0.316	0.112	0	0
CH	0	0	0.266	0	0	0	0.266	0.277	0.191	0	0
DD	0	0	0.132	0	0.056	0	0.16	0.146	0.507	0	0
FE	0	0.008	0.231	0	0.338	0.031	0.023	0.008	0.362	0	0
FO	0	0	0.188	0	0.063	0	0.188	0.5	0.063	0	0
HA	0	0	0.118	0	0.015	0	0.382	0.324	0.162	0	0
HB	0	0	0.154	0	0.026	0	0.692	0.103	0.026	0	0
LY	0	0	0.429	0	0.122	0	0.133	0	0.316	0	0
MC	0.013	0	0.132	0	0.105	0.039	0.632	0.066	0.013	0	0
MD	0	0	0.263	0.026	0.197	0.026	0.211	0.237	0.039	0	0
MH	0	0	0.2	0	0.067	0	0.267	0.167	0.3	0	0
M	0	0.013	0.038	0	0.115	0	0.359	0.103	0.372	0	0
P	0	0	0	0	0.02	0	0.92	0.02	0.04	0	0
PD	0	0.019	0.202	0	0.077	0	0.202	0.24	0.231	0.029	0
R	0	0	0.078	0.01	0.324	0	0.098	0.01	0.48	0	0
RIH	0	0	0.125	0	0.031	0	0.531	0.219	0.094	0	0
RH	0	0	0.1	0	0.08	0.01	0.71	0.09	0.01	0	0
V	0	0	0.032	0	0.129	0.016	0.403	0.129	0.29	0	0
WHH	0	0	0.235	0	0.029	0	0.657	0.069	0.01	0	0
W	0	0	0.17	0	0.148	0	0.273	0.25	0.125	0.034	0
WYE	0	0	0.586	0	0.207	0.052	0	0	0.155	0	0
All	0	0.002	0.171	0.002	0.101	0.008	0.363	0.143	0.207	0.003	0

Appendix V

Allele Frequency data for Lb4/18

	ALLELE SIZE													
Population	**145**	**146**	**147**	**148**	**149**	**150**	**151**	**152**	**153**	**154**	**155**	**156**	**157**	**158**
A3	0	0	0	0	0.019	0.074	0.037	0.074	0.667	0.019	0	0.111	0	0
BD	0	0	0	0	0	0.256	0.186	0.047	0.5	0.012	0	0	0	0
BH	0.011	0	0	0.034	0.034	0.182	0.205	0.17	0.364	0	0	0	0	0
BC	0	0	0	0	0	0.083	0.042	0.292	0.583	0	0	0	0	0
B	0	0	0.028	0.028	0.069	0.125	0.083	0.25	0.292	0.014	0.028	0.083	0	0
CL	0	0	0	0	0	0.044	0.356	0	0.589	0	0.011	0	0	0
CH	0	0.011	0	0.011	0.034	0.023	0.091	0.091	0.739	0	0	0	0	0
DD	0	0.024	0.008	0.016	0.024	0.105	0.331	0.056	0.403	0.016	0	0.016	0	0
FE	0	0	0	0.031	0.052	0.146	0.188	0.125	0.313	0.052	0.073	0.021	0	0
FO	0	0	0	0	0	0.438	0.313	0.063	0.188	0	0	0	0	0
HA	0	0.041	0.054	0.014	0.027	0.554	0.122	0.041	0.135	0.014	0	0	0	0
HB	0	0	0	0.026	0	0.154	0.167	0.154	0.487	0	0.013	0	0	0
LY	0	0	0	0.025	0.025	0.25	0.138	0.088	0.425	0	0.025	0.025	0	0
MC	0	0.038	0.013	0	0	0.282	0.205	0.013	0.41	0.026	0	0.013	0	0
MD	0	0	0	0	0.013	0.538	0.225	0.013	0.2	0.013	0	0	0	0
MH	0	0	0	0	0	0.061	0.318	0.03	0.576	0.015	0	0	0	0
M	0	0	0	0.012	0.012	0.134	0.195	0.159	0.427	0.012	0	0	0.049	0
P	0	0	0	0.022	0.065	0.522	0.13	0.022	0.152	0	0	0.087	0	0
PD	0	0	0.021	0	0.053	0.319	0.33	0.043	0.234	0	0	0	0	0
R	0	0.029	0.029	0	0.049	0.186	0.324	0.039	0.294	0.01	0.01	0.02	0	0.01
RIH	0	0	0	0	0	0.125	0.219	0	0.563	0	0.031	0.063	0	0
RH	0	0	0.027	0.027	0.014	0.203	0.257	0.081	0.324	0.014	0.041	0	0.014	0
V	0	0	0	0	0	0	0.081	0.145	0.742	0.032	0	0	0	0
WHH	0.021	0	0.021	0	0	0.394	0.149	0.021	0.351	0.021	0.011	0	0.011	0
W	0	0	0.037	0.012	0.037	0.22	0.232	0.061	0.244	0.049	0.049	0.049	0.012	0
WYE	0	0	0.021	0.042	0.146	0.25	0.25	0.104	0.188	0	0	0	0	0
All	0.002	0.007	0.011	0.012	0.026	0.212	0.208	0.079	0.397	0.014	0.012	0.016	0.004	0.001

Appendix VI

Allele Frequency data for Lb4/19

	ALLELE SIZE												
Population	**182**	**188**	**194**	**200**	**206**	**212**	**224**	**230**	**236**	**242**	**248**	**254**	**260**
A3	0	0	0.02	0.1	0	0	0.3	0.02	0	0	0	0.56	0
BD	0.027	0.108	0.027	0.324	0	0.041	0.027	0.257	0.014	0.014	0.081	0.054	0.027
BH	0	0	0	0.558	0	0.058	0	0.221	0.023	0.012	0.128	0	0
BC	0	0	0.1	0	0	0.033	0.7	0	0	0	0	0.167	0
B	0.034	0.011	0	0.75	0.08	0.011	0	0	0.102	0	0.011	0	0
CL	0	0	0.088	0.157	0	0	0.353	0.098	0	0	0	0.304	0
CH	0	0	0.083	0.226	0.024	0.012	0.405	0.024	0	0	0	0.226	0
DD	0	0.071	0.103	0.421	0.048	0	0.087	0.04	0.159	0.063	0.008	0	0
FE	0	0	0.143	0.479	0.093	0	0.05	0.157	0.036	0.036	0.007	0	0
FO	0	0	0.3	0.35	0	0	0.35	0	0	0	0	0	0
HA	0	0.015	0.5	0.121	0	0	0.182	0.03	0.015	0	0.136	0	0
HB	0	0	0.016	0.468	0.016	0.032	0	0.113	0.065	0	0.097	0.194	0
LY	0	0	0.181	0.213	0.17	0	0.138	0.16	0.085	0.043	0.011	0	0
MC	0	0	0.207	0.228	0.011	0.033	0.196	0.25	0.065	0	0.011	0	0
MD	0	0	0.218	0.423	0.09	0	0.192	0.064	0	0	0	0.013	0
MH	0	0.029	0.176	0.544	0	0	0.015	0.015	0.132	0.088	0	0	0
M	0	0.022	0.022	0.822	0.022	0	0.033	0	0.067	0	0	0.011	0
P	0	0	0.104	0.104	0	0.021	0.333	0.125	0	0	0.083	0.229	0
PD	0	0	0.08	0.52	0	0	0.35	0	0.05	0	0	0	0
R	0	0	0.1	0.24	0	0	0	0.06	0.38	0.2	0	0.02	0
RIH	0	0.045	0	0.545	0	0	0.227	0.182	0	0	0	0	0
RH	0	0.025	0.125	0.663	0.038	0	0.025	0.1	0	0	0.025	0	0
V	0	0	0.21	0.371	0.113	0	0.048	0.145	0.081	0.032	0	0	0
WHH	0	0	0	0.278	0	0.178	0.2	0.344	0	0	0	0	0
W	0	0.122	0.11	0.39	0	0	0	0.244	0.037	0.073	0.024	0	0
WYE	0	0	0.1	0.6	0.033	0	0	0.25	0.017	0	0	0	0
All	0.003	0.018	0.112	0.396	0.034	0.017	0.137	0.115	0.062	0.027	0.023	0.057	0.001

Appendix VII

Allele Frequency data for Lb1/41

	ALLELE SIZE																				
Population	172	174	176	178	180	188	190	192	198	200	202	204	206	216	222	224	226	228	230	232	234
A3	0	0	0	0.017	0	0	0	0	0	0	0.155	0	0	0	0	0	0	0.034	0.431	0.362	0
BD	0.01	0	0.143	0.048	0.024	0	0	0	0	0	0.107	0	0	0	0	0	0	0.012	0.179	0.476	0
BH	0	0	0.025	0.025	0	0	0	0	0	0	0.213	0	0	0	0	0	0	0.025	0.188	0.5	0.025
BC	0	0	0	0	0	0	0	0	0	0	0.179	0.036	0	0	0	0.036	0	0.179	0.071	0.5	0
B	0	0	0.023	0.023	0.058	0	0	0	0	0	0	0	0	0	0	0	0.012	0.058	0.465	0.349	0.012
CL	0	0	0	0.021	0	0	0	0	0	0	0.245	0	0	0	0	0	0.011	0.074	0.16	0.447	0.043
CH	0	0	0.023	0	0	0	0	0	0	0	0.239	0	0	0	0	0	0	0.091	0.227	0.409	0.011
DD	0.016	0	0	0	0.056	0	0.016	0	0	0	0.21	0	0	0	0	0.008	0	0.121	0.169	0.395	0.008
FE	0	0	0	0	0	0	0	0	0	0	0.122	0	0	0	0	0	0.02	0.061	0.276	0.52	0
FO	0	0	0.05	0	0	0	0	0	0	0	0.2	0	0	0	0	0	0	0	0.2	0.55	0
HA	0	0.016	0.21	0.194	0.113	0	0	0	0	0	0.016	0	0	0	0	0	0	0	0.081	0.371	0
HB	0	0	0.073	0	0	0	0	0	0	0	0.232	0.012	0	0	0	0	0	0.024	0.122	0.524	0.012
LY	0	0	0	0.061	0.024	0	0	0	0	0	0.195	0	0	0	0	0	0	0	0.183	0.537	0
MC	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.174	0.267	0.453	0.035
MD	0	0	0.026	0	0.013	0	0	0	0	0	0.079	0	0	0	0	0	0	0.013	0.039	0.803	0.026
MH	0	0	0	0	0	0	0	0	0	0	0.192	0	0	0	0	0	0	0.231	0.096	0.462	0.019
M	0	0	0.012	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.72	0.256	0.012
P	0	0	0.16	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0.02	0.62	0.1	0.04
PD	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0	0.11	0.1	0.73	0.01
R	0.025	0	0.025	0.013	0.013	0	0	0	0	0	0.05	0	0.013	0.013	0.013	0	0	0.1	0.413	0.3	0.025
RIH	0	0	0	0	0	0	0	0	0	0	0.133	0	0.067	0	0	0	0.033	0.1	0.333	0.333	0
RH	0	0	0.077	0.026	0.038	0	0	0	0	0	0.128	0	0.013	0	0	0	0	0	0.103	0.603	0.013
V	0	0	0.097	0	0.016	0	0	0	0	0	0	0	0	0	0	0	0	0.065	0.581	0.21	0.032
WHH	0	0	0.011	0	0	0	0	0	0	0.022	0.065	0	0	0	0.033	0	0	0.076	0.283	0.489	0.022
W	0.013	0.038	0	0.05	0.038	0	0	0.013	0	0	0.038	0	0.013	0	0	0	0	0.063	0.163	0.575	0
WYE	0	0	0	0	0	0.019	0	0	0.222	0.037	0.111	0	0	0	0	0	0	0.074	0.278	0.185	0.074
All	0.003	0.002	0.037	0.018	0.017	0.001	0.001	0.001	0.006	0.002	0.113	0.001	0.003	0.001	0.002	0.001	0.004	0.065	0.255	0.452	0.016

Appendix VIII

Allele Frequency data for Lb1/57

	Allele Size																								
Population	**171**	**173**	**175**	**177**	**179**	**181**	**183**	**185**	**187**	**189**	**191**	**197**	**199**	**201**	**203**	**205**	**207**	**209**	**211**	**213**	**215**	**217**	**219**	**221**	**223**
A3	0	0	0.07	0.29	0.05	0	0	0	0	0	0	0	0	0	0.2	0.02	0	0	0.02	0.04	0	0	0.21	0.09	0.02
BD	0	0.07	0.29	0.35	0.1	0	0.01	0	0	0	0	0	0	0	0.04	0.01	0	0	0.01	0.02	0.04	0	0	0.04	0
BH	0	0.05	0.42	0.35	0.06	0	0.01	0	0	0	0	0	0	0.012	0.04	0.04	0	0	0	0	0	0	0.01	0.02	0
BC	0	0	0.07	0.03	0	0.07	0	0	0	0	0.033	0	0	0	0.17	0	0	0	0	0	0.07	0	0.47	0.03	0.07
B	0.03	0.04	0.11	0.06	0.07	0.03	0	0	0	0	0	0	0	0.014	0	0	0	0	0	0.03	0.03	0.22	0	0.35	0.03
CL	0	0	0.02	0.16	0.03	0	0	0	0	0	0	0	0	0	0.23	0.05	0	0	0	0	0.03	0.05	0.17	0.22	0.05
CH	0	0	0.02	0.07	0	0	0	0	0	0	0	0	0	0	0.12	0.06	0	0.04	0	0	0	0.01	0.67	0	0.01
DD	0	0.01	0.37	0.22	0.22	0.04	0.01	0	0	0	0	0	0.09	0	0	0	0	0	0	0	0.01	0.02	0.01	0.01	0
FE	0	0.02	0.33	0.11	0.12	0.1	0	0.009	0	0	0	0	0	0	0.04	0.01	0	0.02	0.01	0	0	0.05	0.17	0	0.02
FO	0	0	0.41	0.27	0	0	0	0	0	0	0	0	0	0	0.18	0.05	0.091	0	0	0	0	0	0	0	0
HA	0	0.03	0.15	0.24	0.24	0.02	0	0	0	0	0	0	0	0.016	0.13	0.08	0.032	0	0.02	0.03	0.02	0	0	0	0
HB	0	0.03	0.21	0.53	0.1	0	0	0	0	0	0	0	0	0	0.08	0	0	0	0	0	0	0	0.06	0	0
LY	0	0	0.16	0.63	0.09	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0.01	0	0
MC	0	0.03	0.14	0.22	0.22	0.01	0	0	0	0	0	0	0	0	0.21	0	0.014	0	0.01	0	0.08	0.06	0	0	0
MD	0	0.01	0.06	0.15	0.21	0	0	0	0	0.013	0	0	0	0	0.35	0	0	0	0.03	0.06	0.05	0.04	0	0.03	0
MH	0	0	0.17	0.47	0.01	0.04	0	0	0	0	0	0.014	0.06	0	0	0	0	0	0.11	0	0.06	0.01	0	0.04	0
M	0	0	0.15	0.08	0.01	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0.05	0.3	0.04	0.34	0
P	0	0	0.39	0.05	0.14	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0.09	0	0.02	0.23
PD	0	0.03	0.17	0.33	0.17	0.01	0	0	0	0	0	0	0	0.011	0.16	0.02	0	0	0	0.04	0	0.01	0.01	0.01	0.02
R	0	0	0.11	0.12	0.08	0	0	0	0.012	0	0	0	0	0	0.05	0	0	0	0.05	0	0.07	0.2	0.12	0.21	0
RIH	0	0	0.19	0.28	0.31	0	0	0	0	0	0	0	0	0	0.06	0.03	0	0	0	0.03	0.06	0	0	0.03	0
RH	0	0	0.19	0.24	0.25	0.01	0	0	0	0	0	0	0	0	0.08	0.03	0	0	0	0	0.04	0.07	0.1	0	0
V	0	0	0.16	0.05	0.03	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0.05	0	0	0.19	0.02	0.45	0.02
WHH	0	0	0.24	0.13	0.22	0	0	0.023	0.023	0.012	0	0	0	0	0.07	0.02	0.047	0	0	0	0	0.11	0.05	0.04	0.02
W	0	0.01	0.15	0.2	0.19	0.04	0	0	0	0	0	0	0.01	0	0.19	0	0	0	0.04	0.08	0.01	0.04	0.01	0	0.05
WYE	0	0	0	0.71	0.15	0.02	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0.02	0.02	0.04	0
All	0.00	0.01	0.19	0.24	0.13	0.02	0.002	0.002	0.002	0.001	0.001	0.001	0.01	0.002	0.09	0.02	0.005	0.003	0.01	0.01	0.02	0.06	0.08	0.07	0.02

11.0 REFERENCES

Albin, E. (1720) *The natural history of English insects*. London.

Aljanabi, S.M. & Martinez, I. (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*. **25**: 4692-4693.

Altukhov, Y.P. & Abramova, A.B. (2000) Species-specific random amplified monomorphic DNA. *Russian Journal of Genetics*. **36**: 1411-1417.

Archie, J.C., Simon, C. & Wartenberg, D. (1985) Geographical patterns and population structure in periodical cicadas based on spatial analysis of Allozyme frequencies. *Evolution*. **39**: 1261-1274.

Arnold, M.l., Buckner, C.M. & Robinson, J.J. (1991) Pollen mediated introgression and hybrid speciation in Louisiana irises. *Proceedings of the national Academy of Sciences of the USA*. **88**: 1398-1402.

Asher, J. (1994) The butterflies of Berkshire, Buckinghamshire and Oxfordshire. Pisces Press, Newbury, Berkshire.

Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G & Jeffcoate, S. (2001) *The millennium atlas of butterflies in Britain and Ireland.* Oxford University Press, Oxford.

Ashley, M.V. and Dow, B.D. (1994). The use of microsatellite analysis in population biology: Background, methods and potential applications. In: *Molecular Ecology and Evolution, Approaches and Applications*. B. Schierwater, Streit, B., Wagner, G. P., DeSalle, R. (ed.). Birkhauser Verlag Base, Switzerland.

Audisio, P., De Biase, A., Romanelli, P., Angelici, M.C., Ketmaier, V. & De Matthaeis, E. (2000) Molecular re-examination of the taxonomy of the *Meligethes viridescens* species complex (Coleoptera : Nitidulidae) *Biochemical Systematics and Ecology*. **28**: 1-13

Austin, J.J., Ross, A.J., Smith, A.B., Fortey, R.A. & Thomas, R.H. (1997) Problems of reproducibility: does geologically ancient DNA survive in amber preserved insects? *Proceedings of the Royal Society of London (series B) Biological Sciences*. **264**: 467-474

Avise, J.C. (1986) Mitochondrial DNA and the evolutionary genetics of higher animals. *Philosophical Transactions of the Royal Societyof London. B-Series*. **312**: 325-242.

Avise, J.C. (1989) A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology and Evolution*. **9**: 279-281.

Avise, J.C. (1991) Ten unorthodox perspectives on evolution prompted by comparative population genetics findings on mitochondrial DNA. *Annual Review of Genetics*, **25**: 45-69

Avise, J.C. (1994) *Molecular markers, Natural History and Evolution.* Chapman & Hall, New York U.S.A.

Ayres, K.L. & Balding, D.J. (1998) Measuring departures from Hardy-Weinberg: a Markov-chain Monte Carlo method for estimating inbreeding coefficient. *Heredity*. **80**: 769-777

Baker, C.S., Parry, A. & Bannister, J.L. (1993) Abundant mitochondrial DNA variation and world wide population structure in humpback whales. *Proceedings of the National Academy of Sciences of the USA,* **90**, 8239-8243.

Balloux, F., Brunner, H., Lugon-Moulin, N., Hausser, J., Goudet, J. (2000) Microsatellites can be misleading: An empirical and simulation study. *Evolution*. **54**: 1414-1422

Barascud, B., Martin, J.F., Baguette, M. & Descimon, H. (1999) Genetic consequences of an introduction-colonization process in an endangered butterfly species. *Journal of Evolutionary Biology*. **12**: 697-709.

Barker, J.S.F., Moore, S.S., Hetzel, D.J.S., Evans, D., Tan, S.G. & Byrne, K. (1997) Genetic diversity of Asian water buffalo (*Bubalus bubalis*): microsatellite variation and a comparison with protein coding loci. *Animal Genetics*. **28**: 103-115.

Barker, S. R. J. (1990). Cotswold Adonis Blue re-establishment project - report on feasibility study 1989. Unpublished report to Butterfly Conservation, West Midlands Branch.

Barker, S. R. J. (1991). Cotswold Adonis Blue re-establishment project - report on monitoring of release 1990. Unpublished report to Butterfly Conservation, West Midlands Branch.

Barton, N.H. (1992). The genetic consequences of dispersal. In: *Animal Dispersal: Small Mammals as a Model*. N. C. Stenseth, Lidicker, W. Z. (ed.). Chapman & Hall, London. pp.37-63.

Berg, E.E., & Hamrick, J.L. (1997) Quantification of genetic diversity at allozyme loci. *Canadian journal*

of forest research- revue Canadienne de recherche forestiere. **27**: 415-424
Berry, R.J., Triggs, G.S., King, P., Nash, H.R. & Noble, L.R. (1991) Hybridisation and gene flow in house mice introduced in to an existing population on an island. *Journal of Zoology* (London). **225**: 615-632.
Birkle, L.M. & Douglas, A.E. (1999) Low genetic diversity among pea aphid (*Acyrthosiphon pisum*) biotypes of different plant affiliation. *Heredity*. **82**: 605-612.
Black, W.C. (1993) PCR with arbitrary primers: approach with care. *Insect Molecular Biology* **2**: 1-6
Blackstock, T.H., Rimes, C.A., Stevens, D.P., Jefferson, R.G., Robertson, H.J., Mackintosh, J. & Hopkins, J.J. (1999) The extent of semi-natural grassland communities in lowland England and Wales: a review of conservation surveys 1978-96. *Grass and Forage Science*. **54**: 1-18
Blanchetot, A. (1989) Detection of highly polymorphic regions in insect genomes. *Nucleic Acids Research.* **17**: 3313
Blanchetot, A. (1991a) *Musca domestica* satellite sequence detects individual polymorphic regions in insect genomes. *Nucleic Acids Research*. **19**: 929-932
Blanchetot, A. (1991b) Genetic relatedness in honey bees as established by DNA fingerprinting. *J. of Heredity*. **82**: 391-396
Blanchetot, A. (1992) DNA fingerprinting in the solitary bee *Megachile rotunda*: variability and nest mate genetic relationships. *Genome.* **35**: 681-688
Blanchetot, A. & Packer, L. (1992) Genetic variability in the social bee *Lasioglossum marginatum* and a cryptic undescribed sibling species, as detected by DNA fingerprinting and allozyme electrophoresis. *Insect Molecular Biology*. **1**: 89-97.
Blouin, M.S., Parsons, M., Lacaille, V. & Lotz, S. (1996) use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology*. **5**: 393-401
Bonnell, M.L. & Selander, R.K. (1974) Elephant Seals: genetic variation and near to extinction. *Science.* **184**: 908-909.
Bossart, J.L. & Prowell, D.P. (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution*. **13**: 202-206.
Boughton, D.A. (2000) A dispersal system of a butterfly: a test of source-sink theory suggests the intermediate-scale hypothesis. *American Naturalist*. **256**: 131-144.
Bourn, N. A. D. & Warren, M. S. (1998) Species Action Plan: Adonis Blue, *Lysandra bellargus* (*Polyommatus bellargus*). Butterfly Conservation
Bourn, A.D., Whitfield, K.E.J., Pearman, G.S. & Roberts, E. (1999) *Site dossier and status of the Adonis Blue Polyommatus (lysandra) bellargus in Dorset between 1997 and 1999.* Butterfly Conservation.
Bouzat , J.L., Lewin, H.A. & Paige, K.N. (1998a) The ghost of genetic diversity past: historical DNA analysis of the greater prairie chicken. *The American Naturalist*. **153**: 1-6
Bouzat, J.L., Cheng, H.H., Lewin, H.A., Westmeier, R.L., Brawn, J.D. & Paige, K.N. (1998b) Genetic evaluation of a demographic bottleneck in the greater prairie chicken. *Conservation Biology*. **12**: 836-843
Britten, H.B., Brussard, P.F., Murphy, D.D. & Ehrlich, P.R. (1995) A test for isolation-by-distance in central Rocky mountain and Great Basin populations of Edith's checkerspot butterfly (*Euphydryas editha*). *Journal of Heredity*. **86**: 204-210.
Brookes, M.I., Graneau, Y.A., King, P., Rose, O.C., Thomas, C.D. & Mallet, J.L.B. (1997) genetic analysis of founder bottlenecks in the rare British butterfly *Plejebus argus*. *Conservation Biology*. **11**: 648-661.
Brookfield, J.Y.F. (1996) A simple method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*. **5**: 453-455.
Brower, A.V.Z. & Boyce, T.M. (1991) Mitochondrial DNA variation in Monarch butterflies. *Evolution*, **45**: 1281-1286
Brunton, C.F.A. & Hurst, G.D.D. (1998) Mitochondrial DNA phylogeny of Brimstone butterflies (genus *Genopteryx*) from the Canary Islands and Madeira. *Biological Journal of the Linnean Society*, **63**: 1-11
Bryant, E.H., Meffert, L.M. & McCommas, S.A. (1990) Fitness rebound in serially bottlenecks of the housefly. *American Naturalist*, **136**: 542-549.

Bulat, S.A. & Zakharov, I.A. (1992) Detection of DNA by means of the polymerase chain reaction in the material of entomological collections. *Zh. Obshch. Biology.* **53**: 861-863.

Burke, T. (1994) Spots before the eyes: molecular ecology. *Trends in Ecology and Evolution,* **9**: 355-357

Butterflies Under Threat Team (BUTT). (1986) *The management of chalk Downland for Butterflies*, Focus on Nature Conservation Series, No. 17. Nature Conservancy Council, Peterborough.

Buxton, R.D. & Connolly, M.L. (1973) 1973 Survey of the Adonis Blue, *Lysandra bellargus* Rot. Unpublished.

Cabin, R.J. & Mitchell, R.J. (2000) To Bonferroni or not to Bonferroni: when and how are the questions. *Bulletin of the Ecological Society of America.* **81**: 246-248

Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H.A., Richards, R.I., Mulley, J.C. & Sutherland, G.R. (1993) Incidence and origin of "null" alleles in the $(CA)_n$ microsatellite markers. *American Journal of Human Genetics.* **52**: 922-927.

Calos, J.B. & Sakaluk, S.K. (1998) Paternity of offspring in multiply-mated female crickets: the effect of nuptial food gifts and the advantage of mating first. *Proceedings of the Royal Society of London (Series B) Biological Sciences.* **265:** 2191-2195

Cameron, S.A. (1993) Multiple origins of advanced eusociality in bees inferred from mitochonrial DNA sequences. Proceedings of the National Academy of Sciences USA, 90: 8687-8691.

Cano, R..J. & Poinar, H.N. (1993) Rapid isolation of DNA from fossil and museum specimens suitable for PCR. *Biotechniques.* **15**: 432-436.

Cano, .R.J., Hendrik, .N.P., Pieniazek, .N.J. & Poinar, G.O. (1993) Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature.* **363**: 536-538.

Carbone, I., Anderson, J.B. & Kohn, L.M. (1999) Patterns of descent in clonal lineages and their multilocus fingerprints are resolved with combined gene genealogies. *Evolution.* **53**: 11-21

Cardoso, S.R.S., Cardoso, N.B., Cardoso, J., Cardoso, M.A. & Ferreira, P.C.G. (2000) Genetic Differentiantion of *Euterp edulis* Mart. Populations estimated by AFLP analysis. *Molecular Ecology.* **9**: 1753-1760

Caro, T.M. & Laurenson, M.K. (1994) Ecological and genetic factors in conservation: a cautionary tale. *Science*, **263**: 485-486

Caughley, G. (1994) Directions in conservation biology. *Journal of Animal Ecology.* **63**: 215-244

Carvalho, G.R., McLean, N., Wratten, S.D., Carter, R.E. & Thurston, J.P. (1991) Differentiation of Aphid clones using DNA fingerprints from individual aphids. *Proceedings of the Royal Society of London B.* **243**: 109-114.

Cavalli-Sforza LL and Edwards AWF (1967) Phylogenetic analyses : models and estimation procedures. *Evolution* 32 : 550-570

Cervera MT, Cabezas JA, Simon B, Martinez-Zapater JM, Beitia F, Cenis JL (2000) Genetic relationships among biotypes of *Bemisia tabaci* (Hemiptera : Aleyrodidae) based on AFLP analysis. *Bulletin of Entomological Research.* **90**: 391-396

Chakraborty, R. & Jin, L. (1993) Determining the relatedness between individuals using DNA fingerprinting. *Human Biology.* **65**: 875-895.

Chakraborty, R. & Nei, M. (1977) Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution.* **31**: 347-356.

Chapco, W., Ashton, N.W., Martel, R.K.B. & Antonishyn, N. (1992) A feasibility study of the use of random amplified polymorphic DNA in the population genetics and systematics of grasshoppers. *Genome.* **35***:* 569-574.

Chapman, H.F., Hughes, J.M., Jennings, C., Kay, B.H. & Ritchie, S.A. (1999) Population structure and dispersal of the saltmarsh mosquito *Aedes vigilax* in Queensland, Australia. *Medical and Veterinary Entomology.* **13**: 423-430.

Clarke, G.M. & O'Dwyer, C. (2000) Genetic variability and population structure of the endangered golden sun moth, *Synemon plana. Biological Conservation.* **92**: 371-381.

Clary, D.O. & Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organisation and genetic code. *Journal of Molecular Evolution.* **22**: 252-271.

Clutton-Brock, T.H. (1989) Female transfer and inbreeding avoidance in social mammals. *Nature.* **337**: 70-72.

Cockburn.A.F. & Fritz.G.A. (1996) Isolation and purification of insect DNA. In: *Methods in Molecular Biology, vol 50: species diagnostic protocols: PCR and other nucleic acid methods* (Clapp.J.P. ed) Humana Press Inc., Totowa, NJ.

Cockerham, C.C. & Weir, B.S. (1993) Estimation of gene flow from F-statistics. *Evolution*. **47**: 855-863.

Colin, R. (2000) Phylogeny of the Crepidula plana (Gastropoda: Calyptraeidae) cryptic species complex in North America. *Canadian Journal of Zoology-Revue Canadienne de Zoologie*. **78**: 1500-1514

Collevati, R.G., Grattapaglia, D., Hay, J.D. (2001) Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. *Molecular Ecology*. **10**: 349-356

Cooper, A. (1994) Dried samples: soft tissues. DNA from museum specimens. Ancient DNA (Hermann, B. and Hummel, S. eds), pp. 149-165. Springer, New York.

Cornuet, J.M. & Luikart, G. (1996) Decription and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*. **144**: 2001-2014.

Costa, J.T. & Ross, K.G. (1993) Seasonal decline in intracolony genetic relatedness in eastern tent caterpillars: Implications for social evolution. *Behavioural Ecology and Sociobiology*. **32**: 47-54.

Crawley, M.J. (1993) *Methods in ecology: GLIM for Ecologists*. Blackwell Science. Oxford.

Crease, T.J., Lee, S.K., Yu, S.L., Spitze, K., Leheman, N. & Lynch, M. (1997) Allozyme and mtDNA variation in populations of the *Daphnia pulex* complex from both sides of the Rocky mountains. *Heredity* **79**: 242-251.

Crespi, B.J. (1991) Heterozygosity in the haplodiploid Thysanaoptera. *Evolution*. **45**: 458-464.

Crozier, R.H., Crozier, Y.C. & Mackinlay, A.G. (1989) The CO-I and CO-II region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolution rates. *Molecular Biology and Evolution*. **6**: 399-411.

Dallas, J.F. (1992) Estimation of microsatellite mutation rates in recombinant inbred strains of mouse, *Mamm. Genome*. **3**: 452-456.

DeBarro, P. Sherrat, T., Wratten, S. & Maclean, N. (1994) DNA fingerprinting of cereal aphids using $(GATA)_4$. *European Journal of Entomology*. **91**: 109-114

Debinski, D.M. (1994) Genetic Diversity assessment in a metapopulation of the butterfly *Euphydryas gillettii. Biological Conservation.* **70:** 25-31.

Dempster, J.P. (1991) Fragmentation, isolation and mobility of insect populations. In: Collins, N.M. & Morris, M.G. (Eds.) *The conservation of insects and their habitats.* London: Academic.

Dempster, J.P., King, M.L. & Lakhani, K.H. (1976) The status of the swallowtail butterfly in Britain. *Ecological Entomology.* **1**: 71-84.

Deowanish, S., Nakamura, J., Matsuka, M. & Kimura, K. (1996) MtDNA variation among subspecies of *Apis cerana* using restriction fragment length polymorphism. *Apidologie*. **27**: 407-413.

Department of the Environment (1995) *Biodiversity: The U.K. Steering group report.* H.M.S.O. London.

Desalle, R. (1994) Implications of ancient DNA for phylogenetic studies. *Experientia*. **50**: 543-550

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M and Friemer NB (1994) Mutational processes of simple sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences USA* **91** : 3166-3170

Dillon, N., Austin, A.D. & Bartowsky, E. (1996) Comparison of preservation techniques for DNA extraction from Hymenopterous insects. *Insect Molecular Biology*. **5(1)**: 21-24

Dobson, F.S. (1982) Competition for mates and predominant juvenile male dispersal in mammals. *Animal Behaviour*. **30**: 1183-1192.

Dobzhanskky, T. Wright, S. (1943) Genetics of natural populations. X dispersion rates in *Drosophila pseudoobscura. Genetics*.**32**: 303-324.

Eaton, L.C. & Medel, R.G. (1994) Allozyme variation and genetic relatedness in a population of *Campotanus-chilensis* in Chile. *Revista Chilena De Historia Natural*. **67**: 157-161.

Edwards, A., Hammond, H.A. , Jin, L., Caskey, C.T. & Chakraborty, R. (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, **12**: 241-253.

Eliot, J.N. (1973) The higher classification of the Lycaenidae (Lepidoptera): a tentative arrangement. *Bulletin of British Museum of Natural History (Entomology)*. **28**: 373-506.

Ellegren, H. (1991) DNA typing of museum birds. *Nature*. **354**: 113.

Ellegren, H., Mikko, S., Wallin, K & Andersson, L. (1996) Limited polymorphism at major histocompatibility complex (MHC) loci in the Swedish moose *A. alces. Molecular Ecology*. **5**: 3-9.

Emmet, M. A. & Heath, J. (1990) *The butterflies of great Britain and Ireland.* Volume 7, part 1. *Heperiidae* to *Nymphalidae.* Colchester: Harley books.

Ender, A., Schwenk, K., Stadler, T., Striet, B. & Schierwater, B. (1996) RAPD identification of microsatellites in *Daphnia. Molecular Ecology*. **5**: 437-442

Endler, J.A. (1977) *Geographic variation, speciation, and clines*. Princeton University Press, Princeton, New Jersey.

England, P.E., Briscoe, D.A. & Frankham, R. (1996) Microsatellite polymorphisms in a wild population of *Drosophila melanogaster. Genetical Research,* **67**: 285-290.

Engel, S.R., Linn, R.A., Taylor, J.F. & Davis, S.K. (1996) Conservation of microsatellite loci across species of Artiodactyls: implications for population studies. *Journal of Mammalogy*. **77**: 504-518

Erlich, H.A. & Arnheim, N. (1992) Genetic analysis using the polymerase chain reaction. *Annual Review of Genetics,* **26**: 479-506

Excoffier, L. (1993) MINSPNET. http://anthroplogie .unigie.ch/lgb/software/win/min-span-net/

Excoffier, L., Smouse, P.E. and Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics.* **131**(479-491).

Fanciulli, P.P., Melegari, D., Carapelli, A., Frati, F. & Dallai, R. (2000) Population structure, gene flow and evolutionary relationships in four species of the genera Tomocerus and Pogonognathellus (Collembola, Tomoceridae). *Biological Journal of the Linnean Society*. **70**: 221-238

Fauron, C.M-R. & Wolstenholme, D.R. (1980) Intraspecific diversity of nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules of *Drosophila Mauritania, Drosophila melanogaster* and *Drosophila simulans. Nucleic Acids Reaserch*. **8**: 5391-5410

Feldman MW, Bergman A, Pollock DD, and Golstein DB (1997) Microsatellite genetic distances with range constraints : analytic description and problems of estimation. *Genetics* 145 : 207-216

Felsenstein, J. (1993) *PHYLIP* (Phylogeny Inference Package) v3.57c. Department of Genetics, University of Washington, Seattle.

Felsenstein, J. (1995). *PHYLIP* (Phylogeny Inference Package). Seattle, University of Washington.

Filchak, K.E., Feder, J.L., Roethele, J.B., & Stolz, U. (1999) A field test for host-plant dependent selection on larvae of the apple maggot fly, *Rhagoletis pomonella. Evolution*. **53**: 187-200

Figurny-Puchalska, E., Gadeberg, R.M.E. & Boomsma, J.J. (2000) Comparison of genetic population structure of the large blue butterflies *Maculinea nausithous* and *M-teleius. Biodiversity and Conservation*. **9**: 419-432

Frankel, O.H. & Soule, M.E. (1981) *Conservation and Evolution*. Cambridge University Press, Cambridge, UK.

Frankham, R. (1995a) Conservation genetics. *Annual Review of Genetics*, **29**: 305- 327

Frankham, R. (1995b) Inbreeding and Extinction: a threshold effect. *Conservation Biology*. **9**: 792 – 799

Fuller, M. (1995) The butterflies of Wiltshire: their history, status and distribution. Pisces publications, Newbury, Berkshire.

Gaggiotti, O.E., Lange, O., Rassmann, K. & Gliddon, C. (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*. **8**: 1513-1520

Gaines, M.S. & McClenaghan, L.R. (1980) Dispersal in small mammals. *Annual Review of Ecology and Systematics*. **11**: 163-196.

Gay, J, & Gay, P. (1996) Atlas of Sussex butterflies. Butterfly Conservation Sussex Branch, Henfield, Sussex.

Gerber, S., Mariette, S., Streiff, R., Bodenes, C. & Kremer, A. (2000) Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Molecular Ecology*. **9**: 1037-1048.

Geurgas, S.R., Infante-Malachias, M.E. & Azeredo-Espin, A.M.L. (2000) Extreme mitochondrial DNA variability and lack of genetic structure in populations of Dermatobia hominis (Diptera : Cuterebridae) from Brazil. *Annals of the Entomological Society of America.* **93**: 1085-1094.

Gibson, C.W.D. (1995) *Creating chalk grassland on former arable land: a review*. Bioscan, Oxford

Gilbert, O.L. & Anderson, P. (1998) *Habitat Creation and Repair.* Oxford University Press. Oxford.
Gillespie, J.H. (1991) *The causes of Molecular Evolution.* Oxford University Press, New York.
Gilpin, M. (1991) The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society.* **42**: 165-175.
Gilpin, M. & Hanski, I. (1991), editors. (1991) Metapopulation dynamics: Empirical and theoretical investigations. Academic Press, London.
Gilpin, M.E. & Soule, M.E. (1986) in: *Conservation Biology: The science of scarcity and diversity.* (ed. Soule, M.E.) Sinauer, Sunderland, MA. pp19-34.
Goldstein, DB. & Pollock, DD (1997) Launching microsatellites: A review of mutation processes and methods of phlogenetic inference. *Journal of Heredity.* **88**: 335-342
Goodman, S.L. (1997) R-ST Calc: A collection of computer programmes for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology.* **6**: 881-885.
Goudet, J. (1995) *FSTAT* (vers. 1.2): A computer programme to calculate F-statistics. *Journal of Heredity.* **86**: 485-486.
Goudet, J. 2000. *FSTAT*, a program to estimate and test gene diversities and fixation indices (version 2.9.1). Available from http://www.unil.ch/izea/softwares/fstat.html.
Goudet, J., Raymond, M., Demeeus, T. & Rousset, F. (1996) Testing genetic differentiation in diploid populations. *Genetics.* **144**: 1933-1940.
Goulson, D. (1993) Allozyme variation in the butterfly *Maniola jurtina* (Lepidoptera: Satyrinae) (L): evidence for selection. *Heredity.* **71**: 386-393
Graur, D. (1985) Gene diversity in Hymenoptera. *Evolution.* **39**: 190-199.
Greenwood, P.J. (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour.* **28**:
1140-1162.
Grevelding, C.G., Kampkotter, A., Hollman, M., Schafer, U. & Kunz (1996) Direct PCR on fruitflies and blood flukes without prior DNA isolation. *Nucleic Acids Reasearch.* **24**: 4100-4101
Groman, J.D. & Pellmyr, O. (2000) Rapid evolution and specialization following host colonization in a yucca moth. *Journal of Evolutionary Biology.* **13**: 223-236.
Groombridge, J.J., Jones, C.G., Bruford, M.W. & Nichols, R.A. (2000) 'Ghost' alleles of the Mauritius kestrel. *Nature.* **403**: 616.
Guo, S. & Thompson, E. (1992) Performing the exact test of Hardy-Weinberg proportions to multiple alleles. *Biometrics.* **48**: 361-372.
Haag, K.L., Dearaujo, A.M. & Zaha, A. (1993) Genetic structure of natural populations of *Dryas-iulia* (Lepidoptera Nymphalidae) revealed by enzyme polymorphism and mitochondrial DNA (mtDNA) restriction fragment length polymorphism (RFLP). *Biochemical Genetics*, 31: 449-460.
Haddad, N.M. (1999) Corridor and distance effects on interpatch movements: a landscape experiment with butterflies. *Ecological Applications.* **9**: 612-622.
Hadrys, H., Balick, M., & Schierwater, B. (1992) Applications of randomly amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology.* **1**: 55-63.
Hale, L.R. & Singh, R.S. (1991) A comprehensive study of genic variation in natural populations of *Drosophila melanogaster.* IV. Mitochondrial DNA variation and the role of history *vs.* selection in the genetic structure of geographic populations. *Genetics*, **129**: 103-117.
Hall, H.G. (1990) Parental analysis of introgressive hybridization between African and European honeybees using nuclear DNA RFLPs. *Genetics* **125**: 611-621
Hammond, B.W., Spanswick, G. & Mawn, J.A. (1996) extraction of DNA from preserved animal specimens for use in randomly amplified polymorphic DNA analysis. *Analytical Biochemistry.* **240**: 298-300
Hanski, I. (1994) Patch occupancy dynamics in a fragmented habitat. *Trends in Evolution and Ecology.* **9**: 131-135.
Hanski, I. (1998) Metapopulation Dynamics. *Nature.* **396**: 41-49
Hanski, I., Kuussaari, M. & Nieminem, M. (1994) Metapopulation structure and migration in the butterfly *Mellitaea cinxia. Ecology.* **75**:757-762.
Hanski, I. & Thomas, C.D. (1994) Metapopulation dynamics and conservation: a spatially explicit model

applied to butterflies. *Biological Conservation*. **68**: 167-180.
Hanski, I., & Gilpin, M.E., Editors. (1996) Metapopulation dynamics: Ecology, genetics and evolution. Academic Press, London.
Harper, G.L., Piyapattanakorn, S., Goulson, D. & maclean, N. (2000) The isolation of microsatellite markers from the Adonis blue butterfly (*Lysandra bellargus*). *Molecular Ecology*. **9**: 1948-1949
Harris, M. (1775) *The English Lepidoptera: or, the Aurelian's pocket companion.* London.
Harris, M. (1766) *The Aurelian or natural history of English insects: namely moths and butterflies. Together with the plants on which they feed.* London.
Harris, H (1966) Enzyme polymorphisms in man. *Proceedings of the Royal Society of London B*, **164**: 298-310.
Harrison, R.G. (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution*. **4**: 6-11
Harry, M., Robin, S. & Lachaise, D. (1998) Use of polymorphic genetic markers (RAPDs) in evolutionary and applied entomology. *Annales de la Societe Entomologique de France*. **34**: 9-32
Hartl, D.L. (1988) *A primer of population genetics 2nd ed.* Sinauer, Sunderland, Massachusetts.
Hartl, G.B. & Hell, P. (1994) Maintenance of high levels of allelic variation in spite of a severe bottleneck in population size: the brown bear (*Ursus arctos*) in the Western Capathians. *Biodiversity and Conservation*. **3**: 546-554.
Haucke, H. & Gellissen, G. (1988) Different mitochondrial gene orders among insects: exchanged tRNA gene positions in the COII/COIII region between an orthopteranand a dipteran species. *Current Opinions in Genetics*. **14**: 471-476.
Hausworth, W.W. (1994) Ancient DNA: an introduction. *Experientia*. **50**: 521-523
Hayashi, K. (1991) PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods and Applications*. **1**: 34-38
Heath, DD., Iwama, GK. & Devlin, RH. (1993) PCR Primed with VNTR core sequences yields species-specific patterns and hypervariable probes. *Nucleic Acids Research*. **21**: 5782-5785
Heath, J., Pollard, E. and Thomas, J.A. (1984) Atlas of butterflies in Britain and Ireland. Viking, London.
Hedrick, P.W., Brussard, B.F., Allendorf, F.W., Beardmore, J.A. & Orzack, S. (1986) Protein variation, f fitness, and captive propagation. *Zoological Biology*. **5**: 91-99.
Heschel, M.S. & Paige, K.N. (1995) Inbreeding and depression, environmental stress and population size in scarlet gilia (*Ipomopsis aggregata*). *Conservation Biology*. **9**: 126– 33.
Hewitt, G.M. (1999) Post-glacial recolonisation of Eurpean biota. *Biological Journal of the Linnean Society*. **68**: 87-112.
Hill, J.K., Thomas, C.D. & Lewis, O.T. (1996) Effects of habitat patch size and isolation on the dispersal by *Hesperia comma* butterflies: implications for metapopulations structure. *Journal of Animal Ecology*. **65**: 725-735.
Hoelzel, A.R. & Dover, G.A. (1991) *Molecular Genetic Ecology*. IRL/ Oxford University press.
Hoffmann, A.A. & Parsons, P.A. (1991) *Evolutionary Genetics and Environmental Stress.* Oxford: Oxford University Press. 284pp.
Hoffman, R.J., Boore, J.L. & Brown, W.M. (1992) A novel genome organisation for the blue mussel *Mytilus edulis*. *Genetics*. **131**: 397-412.
Houlden, B.A., England, P.R., Taylor, A.C., Greville, W.D. & Sherwin, W.B. (1996) Low genetic diversity of the Koala *Phascolarctos cinereus* in south-eastern Australia following a severe population bottleneck. *Molecular Ecology*. **5**: 269-281.
Howard, W.E. (1960) Innate and environmental dispersal of individual vertebrates. *American Naturalist*. **63**: 152-161.
Hughes, C.R. & Queller, D.C. (1993) Detection of highly polymorphic microsatellite loci in a species with little allozyme polymorphism. *Molecular Ecology*. **2**: 131-137
Jackson, J.K & Resh, V.H. (1998) Morphologically cryptic species confound ecological studies of the caddisfly genus *Gumaga* (Trichoptera : Sericostomatidae) in Northern California. *Aquatic Insects*. **20**: 69-84.
Jarne, P. & Lagoda, J.L. (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*. **11**: 424-429
Jeffreys, A.J., Wilson, V. & Thein, S.W. (1985a) Hypervariable 'minisatellite' regions in human DNA.

Nature **314**: 67-73
Jeffreys, A.J., Wilson, V. & Thein, S.W. (1985b) Individual specific 'fingerprints' of human DNA. *Nature* **316**: 76-79
Jiminez, J.A., Hughes, K.A., Alaks, G., Graham, L. & Lacy, R.C. (1994) An experimental study of inbreeding depression on a natural habitat. *Science*. **266**: 271-273.
Johannesen, J., Veith, M. & Seitz, A. (1996) Population genetic structure of the butterfly *Melitaea didyma* (Nymphalidae) along a norther distribution border. *Molecular Ecology*. **5**: 259-267.
Joy, J. (1997). The West Midlands Regional Action Plan. Unpublished report to Butterfly Conservation.
Joyce, D.A. & Pullin, A.S. (2001) Phylogeography of the Marsh Fritillary *Euphydryas aurinia* (Lepidoptera: Nymphalidae) in the UK. *Biological Journal of the Linnean Society*. **72**: 129-141.
Kandpal, R., Kandpal, G. & Weissman, S.M. (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridisation selection for region-specific markers. *Proceedings of the National Academy of Sciences*. **91**. 88-92
Kaufman, B. & Wool, D. (1992) Gene flow by immigrants into isolated recipient populations: a laboratory model using flour beetles. *Genetica*. **85**: 163-171.
Keyghobadi, N., Roland, J. & Strobeck C (1999) Influence of landscape on the populations genetic structure of the alpine butterfly *Parnassius smintheus* (Papillionidae). *Molecular Ecology*. **8**: 1481-1495
Kimmel, M., Chakraborty, R., King, J.P., Bamshap, M., Watkins, W.S. &Jorde, L.B. (1998) Signatures of population expansion in microsatellite loci. *Genetics*. **148**: 1921-1930
Kimura, M. & Crow, J.F. (1964) The number of alleles that can be maintained in a population. *Genetics* **49**: 725-738
King, P.S. (1987) Macro- and microgeographic structure of a spatially subdivided beetle species in nature. *Evolution*. **41**: 401-416.
Koenig, W.D., Van Vuren, D. & Hooge, P.N. (1996) Detectibility, philopatry and the distribution of dispersal distances in mammals. *Trends in Ecology and Evolution.*. **11**: 514-517.
Krafsur, E.S., Vos, J., Nariboldi, P. & Marquez,G. (1995) Gene diversity in the cottonwood leaf beetle (*Chrysomela scripta*). *Biochemical Genetics*. **3-4**: 83-89
Kreher, S. A., Fore, S. A. & Collins, B. S. (2000). Genetic variation within and among patches of the clonal species, *Vaccinium stamineum* L.. *Molecular Ecology*. **9**, 1247-1252.
Kukuk, P.F., Crozier, R.H., Smith, R.R., Eickwort, G.C. & May, B. (1987) Macrogeographic genetic variation in the sweat bee (*Lassioglossum* (*Dialictus*) *zephyrum*. P. 364 *in* J. Eder & H. Rembold (eds.) *Chemistry and Biology of social insects*. Verlag, Munich, Germany.
Kuussaari, M., Nieminem, M. & Hanski, I. (1996) An experimental study of migration in the Glanville Fritillary. *Journal of Animal Ecology*. **65**: 791-801.
Lacy, R.C. (1987) Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology*. **2**: 143-158.
Lade, J.A., Murray, N.D., Marks, C.A. & Robinson, N.A. (1996) Microsatellite differentiation between Philip island and mainland Australian populations of the red fox (*vulpes vulpes*). *Molecular Ecology*. **5**: 81-87
Lagercrantz, U., Ellegren, H. and Andersson, L. (1993). The abundance of various polymorphic microsatellite motifs differ between plants and vertebrates. *Nucleic Acids Research*. **21**. 1111-1115.
Laikre, L. & Ryman, N. (1991) Inbreeding depression in a captive wolf (*Canis lupus*) Population. *Conservation Biology*. **5**: 33-40.
Lande, R. (1994) Risk of population extinction from fixation of new deleterious mutations. *Evolution*. **48**: 1460-1469.
Lande, R. (1995) Mutation and Conservation. *Conservation Biology*. **9**: 782-791.
Le Page, S.L., Livermore, R.A., Cooper, D.W. & Taylor, A.C. (2000) Genetic analysis of a documented population bottleneck: introduced Bennet's wallabies (*Macropus rufogriseus rufogriseus*) in New Zealand. *Molecular Ecology*. **9**: 753-763.
Leberg, P.L. (1991) Influence of fragmentation and bottlenecks on genetic divergence of wild turkey populations. *Conservation Biology*. **5**: 522-530
Leberg, P. (1992) Effects of population bottlenecks on genetic diversity as measured by allozyme

electrophesis. *Evolution*. **46**: 477-494.
Lehman, T., Hawley, A.W. & Collins, H.C. (1996) An evaluation of evolutionary constraints on microsatellite alleles using null alleles. *Genetics*. **144**: 1155-1163
Lessa, E.P. & Applebaum,G. (1993) Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology*, **2**: 119-129.
Lewis, O.T., Thomas, C.D., Hill, J.K., Brookes, M.I., Crane, T.P.R., Graneau, Y.A., Mallett, J.L.B. & Rose, O.C. (1997) Three ways of assessing population structure in the butterfly *Plebejus argus*. *Ecological Entomology*. **22**: 283-293.
Lewontin, R.C. (1991) Electrophoresis in the development of evolutionary genetics: milestone or millstone. *Genetics*. **128**: 657-662.
Lewontin, R.C., & Hubby, J.L. (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, **54**: 595-609
Li W. H. & Garaur, D. (1991,) *Fundamentals of Molecular Evolution*. Sinaur, Sunderland,MA.
Liebherr, J.K. (1986) Comparison of genetic variation in two carabid beetles (Coleoptera) of differing vagility. *Annals of the Entomological Society of America*. **79**: 424-433.
Lister, A.M. (1994) Ancient DNA: not quite *Jurassic Park*. *T.R.E.E.* **9(3)**: 82-84
Litt, M. & Lutty, J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a di-nucleotide repeat within the cardiac muscle gene. *American Journal of Human Genetics*. **44**: 397-401.
Loxdale, H.D., Tarr, I.J., Weber, C.P., Brookes, C.P., Digby, P.G.N. & Castanera, P. (1985) Electrophoretic study of enzymes from cereal aphid populations. III. Spatial and temporal genetic variation of populations of *Sitobion avenae*. (F.) (Hemiptera: Aphididae). *Bulletin of Entomological Reasearch*. **75**: 121-141.
Loxdale, H. & Lushai, G. (1998) Molecular markers in entomology. *Bulletin of Entomological Research*. **88**, 577-600
Lugon-Moulin, N., Brunner, H., Balloux, F., Hausser, J. & Goudet, J. (1999) Do riverine barriers, history or introgression shape the genetic structuring of a common shrew (*Sorex araneus*) population? *Heredity*. **83**: 155-161.
Luikart, G. & Cornuet, J.M. (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology*. **12**: 228-237
Luikart, G., Allendorf, F.W., Cornuet, J.M. & Williams, B.S. (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*. **89**: 238-247.
Luikart, G., Cornuet, J-M. & Allendorf, F.W. (1999) Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conservation Biology*. **13**: 523-530.
Luikart, G. & England, P.R. (1999) Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution*. **14**: 253-255
Luijten, S.H., Dierick, A., Oostermeijer, J.G.B., Raijmann, L.E.L., Den Nijs, H.C.M. (2000) Population size, genetic variation, and reproductive success in a rapidly declining, self-incompatible perennial (Arnica montana) in The Netherlands. *Conservation Biology*. **14**: 1776-1787
Lunt, D.H., Whipple, L.E. & Hyman, B.C (1998) Mitochondrial DNA variable number tandem repeats (VNTR's): utility and problems in molecular ecology. *Molecular Ecology*. **7**: 1441-1445
Lunt, D.H., Hutchinson, F., Carvalho, G.R. (1999) An efficient method for PCR based isolation of Microsatellite arrays (PIMA). *Molecular Ecology*. **8**: 891-894
Lushai, G., Fjellsted, W., Marcovitch, O., Aagaard, K., & Sherratt, K., Allen, J. & McLean, N. (2000) Application of molecular techniques to non-lethal tissue samples of endangered butterfly populations (*Parnassius appolo* L.) in Norway for conservation management. *Biological Conservation*: **94**, 43-50
Lynch, M., & Milligan, B.G.(1994) Analysis of population structure with RAPD markers. *Molecular Ecology*. **3**: 91-99
Lynche, M., Conery, J. & Burger, R. (1995) Mutation accumulation and the extinction of small populations. *American Naturalist*. **146**:489-518.
Mader, H.J. (1984) Animal habitat isolation by roads and agricultural fields. *Biological Conservation*. **29**: 81-96

Mallett, J. (1996) The genetics of biological diversity: from varieties to species. Pages 13-47 in K. J. Garston, editor. *Biodiversity: Biology of numbers and difference*. Blackwell, Oxford, United Kingdom.

Mallett, J. (1997) Genetic analysis of founder bottlenecks in the rare British butterfly *Plejebus argus*. *Conservation Biology*. **11**: 648-661.

Mallett, J., Korman, A., Heckel, D.G. & King, P. (1993) Biochemical genetics of *Heliothis* and *Helicoverpa* (Lepidoptera: Noctuidae) and evidence for a founder event in *Helicoverpa zea*. *Annals of the Entomological Society of America*. **86**: 189-197.

Manly, B.F.J. (1991) *Randomization and Monte Carlo Methods in Biology*, 1st edn. Chapman and Hall, London.

Manly, B.F.J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd edn. Chapman and Hall, London.

Mantel, N. (1967) The detection of disease clustering as a generalised regression approach. *Cancer Research*. **27**: 209-220.

Martin, J.F., Gilles, A. & Descimon, H. (2000) Molecular phylogeny and evolutionary patterns of the European *satyrids* (Lepidoptera : *Satyridae*) as revealed by mitochondrial gene sequences. *Molecular Phylogenetics and Evolution*. **15**: 70-82

McCarthy, C., CHROMAS, version 1.45. Griffith University, Queensland, Australia.

McCauley, D.E. & Eanes, W.F. (1987) Hierarchical population structure analysis of the milkweed beetle, *Tetraopes tetraophthalamus* (Forster). *Heredity*. **58**: 193-201.

McKechnie, S.W., Ehrlich, P.R. & White, R.R. (1975) Population genetics of *Euphydryas butterflies*. I. Genetic variaytion and the neutrality hypothesis. *Genetics*. **81**: 571-594.

Meglecz, E., Pecsenye, K., Peregovits, L. & Varga, Z. (1997). Allozyme variation in *parnassius mnemosyne* (l.) (lepidoptera) populations in north-east hungary: variation within a subspecies group. *Genetica*, 101, 59-66.

Meglecz, E., Pecsenye, K., Varga, Z. & Solignac, M. (1998). Comparison of differentiation pattern at allozyme and microsatellite loci in *Parnassius mnemosyne* (Lepidoptera) populations. *HEREDITAS* , 128, 95-103.

Meglecz, E. & Solignac, M., (1998) Microsatellite loci for Parnassius mnemosyne (Lepidoptera) *Hereditas*. **128**, 179-180

Meglecz, E., Neve, G., Pecsenye, K. & Varga, Z. 1999. Genetic variations in space and time in *Parnassius mnemosyne* (L.) (Lepidoptera) populations in north-east Hungary: implications for conservation. *Biological Conservation*. **89**: 251-259.

Meredith, G.H.J. (1992) Checklist of Gloucestershire butterflies 1975-1988 *The Gloucestershire Naturalist*, Gloucestershire Naturalists Society.

Merola, M. (1994) A reassessment of homozygosity and the case for inbreeding depression in the cheetah (*Acininyx jubatus*): implications for conservation. *Conservation Biology*, **8**: 961-971.

Miller, P.S. (1994) Is inbreeding depression more severe in a stressful environment? *Zoo Biology*, **13**: 195-208

Minch, E. (1997). MICROSAT, Version 1.5b. Stanford, Stanford University Medical Centre.

Mitton, J. (1994) Molecular approaches to population biology. *Annual Review of Ecology and Systematics*.. **25**: 45-69

Mitten, J.B. (1994) Molecular approaches to population biology. *Annu. Rev. Ecol. Syst.* **25**: 45-69.

Moritz, C. (1991) The origin and evolution of parthenogenesis in *Heteronotia binoei* (Gekkonidae): evidence for recent and localized origins of widespread clones. *Genetics*, **129**, 211-219.

Moritz, C. (1994) Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, **3**: 401-411

Moritz, C., Dowling, T.E. &Brown, W.M. (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics*. **18**: 269-292.

Morton, A.C. (1982) The effects of marking and capture on recapture frequencies in butterflies. *Oecologia*. **53**: 105-110.

Mouson, L., Neve, g. & Baguette, M. (1999) Metapopulation structure and conservation of the cranberry fritillary *Boloria aquilonaris* (Lepidoptera: nymphalidae) in Belgium. *Biological Conservation*. **87**: 285-293.

Mundy, N.I. & Woodruff, D.S. (1996) Polymorphic microsatellite markers in the loggerhead shrike *Lanius ludoviacanus* isolated from a library enriched for CA repeats. *Molecular Ecology*. **5**: 811-813.

Napolitano, M. & Descimon, H. (1994) Genetic stucture of French populations of the mountain butterfly *Parnassius mnemosyns* L. (Lepidoptera: papilionidae). *Biological Journal of the Linnean Society*. **53**: 325-341.

Nature Conservancy Council (1984) *Nature Conservation in Great Britain*. Nature Conservancy Council, Shrewsbury, UK.

Negi, M.S., Singh, A. & Lakshmikumuran, M. (2000) Genetic Variation and Relationship Among and Within Withania species as Revealed by AFLP markers. *Genome*. **43**: 975-980

Nei, M. (1973) Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences*. **70**: 3321-3323

Nei, M. (1978) Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics*. **89**: 583-590

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

Nei, M., Maruyama, T. & Chakraborty, R. (1975) The bottleneck effect on genetic variability in populations. *Evolution*. **29**: 1-10.

Nei, M. & Roychoudhury, A.K. (1974) Sampling variances of heterozygosity and genetic distance. *Genetics*. **76**: 379-390.

Nei M, Tajima F and Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* 19 : 153-170

Neumann, K. & Wetton, J.H. (1996) Highly polymorphic microsatellites in the house sparrow, *Passer domesticus*. *Molecular Ecology*. **5**: 307-309.

Neve, G., Mousson, L. & Baguette, M. (1996) Adult dispersal and genetic structure of butterfly populations in a fragmented habitat. *Acta Oecologia International Journal of Ecology*. **17**: 621-626.

Neve, G. & Meglecz, E. 2000. Microsatellite frequencies in different taxa . *TRENDS IN ECOLOGY & EVOLUTION* , **15**, 376-377.

New, T.R., Pyle, R.M., Thomas, J.A., Thomas, C.D. & Hammond, P.C. (1995) Butterfly Conservation Managemant. *Annual Review of Entomology*. **40**: 57-83.

Newman, D. & Pilsen, D. (1997) Increased probability of extinction due to decreased genetic effective population size: Experimental populations of *Clarkia pulchella*. *Evolution*. **51**: 354-362

Nielsen, E.E., Hansen, M.M. & Loeschke, V. (1999) Genetic variation in time and space: microsatellite analysis of extinct and extant populations of Atlantic salmon. *Evolution*. **53**: 261-268

Nieminem, M., Singer, M.C., Fortelius, W., Schops, K & Hanski, I. (2001) Experimental evidence that inbreeding increases extinction risk in butterfly populations. *American Naturalist*. **157**: 237-244

Nunney, L. & Campbell, K.A. (1993) Assessing minimum viable population size: Demography meets population genetics. *Trends in Ecology and Evolution*. **8**: 234-239.

O'Brien, S.J. (1994) The cheetah's conservation controversy. *Conservation Biology*. **8**: 1153-1155.

O'Brien, S.J., Wildt, D.E., Goldman, D., Merril, C.R. & Bush, M. (1983) The Cheetah is depauperate in genetic variation. *Science*. **221**: 459-462.

Oakenfull, E.A. (1994) Vodka, meths and DNA. Trends in Ecology and Evolution. 9: 26

Ohta and Kimura (1973) The model of mutation appropriate to estimate the number of electrophetically detectable alleles in a genetic population. *Genetic Research* **22** : 201-204

Ojala, D., Montoya, J. & Attardi, G. (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature*. **290**: 470-474.

Orita, M., Iwahana, H.m Kanazawa, H., Hayashi & K. Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis and single strand conformation polymorphisms, *Proceedings of the National Academy of Sciences of the USA*. **86**: 2766-2770

Osakabe, M., Hinomoto, N., Toda, S., Komazaki, S. & Goka, K. (2000) Molecular cloning and characterization of a microsatellite locus found in an RAPD marker of a spider mite, *Panonychus citri* (Acari : *Tetranychidae*). *Experimental and Applied Acarology*. **24**: 385-395.

Oyler-McCance, S.J., Kahn, N.W., Burnham, K.P., Brauns, C.E. & Quinn, T.W. (1999) A population genetic comparison of large- and small- bodied sage grouse in Colerado using microsatellite and mitochondrial markers. *Molecular Ecology*. **8**: 1457-1465

Paabo.S. (1989) Ancient DNA: extraction, characterisation, molecular cloning, and enzymatic

amplification. *Proc. Natl. Acad. Sci. USA.* **86**: 1939-1943.
Packer, C., Pusey, A.E., Rowley, H., Gilbert, D.A., Martenson, J. & O'Brien, S.J.O. (1991) Case study for a bottleneck: lions of the Ngorongoro crater. *Conservation Biology*. **5**: 219-230.
Paektau D, Waits LP, Clarkson PL, Craighead L and Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* 147 : 1943-1957
Page, R.D.M. (1996) TREEVIEW: an application to display phylogenetic treeson personal computers. *Computer Applications in Biosciences.* **12**:357-358.
Page, R.E. & Robinson, G.E. (1994) Reproductive competition in queenless honey-bee colonies (*Apis mellifera,* L). *Behavioural Ecology and Sociobiology*. **35**: 99-107.
Palo, J., Varvio, S-L., Hanski, I. & Vainola, R. (1995) Developing microsatellite markers for insect population structure: complex variation in a checkerspot butterfly. *Hereditas*. **123**: 295-300.
Palumbi, S. R. 1995. Nucleic Acids. III. The Polymerase Reaction. In: *Molecular Systematics* (Ed. by D. Hillis & C. Moritz). pp. -. Sunderland, MA: Sinauer
Palumbi, S.R. & Baker, C.S. (1994) Contrasting Population Structure from Nuclear Intron Sequences and MtDNA of Humpback Whales. *Molecular Biology and Evolution*. **11**: 426-435
Parker, P.G., Snow, A.A, Schug, M.D., Booton, G.C. & Fuerst, P.A. (1998) What Molecules can tell us about Populations: Choosing a Molecular Marker. *Ecology*. **79**: 361-382
Pashley, D.P. & Ke, L.D. (1992) Sequence evolution in the mitochondrial ribosomal and ND-1 genes in Lepidoptera: implications for phylogenetic analyses. *Molecular Biology and Evolution*. **9**: 1061-1075
Paxton, R.J., Thoren, P.A., Tengo, J., Estoup, A. & Pamilo, P. (1996) Mating structure and nestmate relatedness in a communal bee, *Andrena jacobi* (*Hymenoptera, Andrenidae*). *Molecular Ecology*. **5**: 511-519.
Pearman, G.S., Goodger, B., Bourn, N.A.D. & Warren, M.S. (1998) *The changing status of the Lulworth* Skipper (Thymelicus acteon) and Adonis Blue (Lysandra bellargus) in south-east Dorset over two *decades*. Butterfly Conservation.
Pemberton, J.M., Slate, J., Bancroft, D.R. & Barrett, J.A. Non amplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology*. **4**: 249-252
Peterson, M.A. (1996) Long-distance gene flow in the sedentary butterfly *Euphilotes enoptes* (Lepidoptera: Lycaenidae). *Evolution*. 50: 1990-1999.
Peterson, M.A. & Denno, R.F. (1998) The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *American Naturalist*. **152**: 428-446.
Pettiver, J. (1717) *Papilionum Britanniae icones*. London.
Philip, E. G. (1993). *The butterflies of Kent : an atlas of their distribution*. The Kent Field Club, Sittingbourne
Phillips, A.J. & Simon, C. (1995) Simple, efficient, and nondestructive DNA extraction protocol for arthropods. *Entomological Society of America*. **88**: 281-283.
Piertney, S.B., MacColl, A.D.C., Lambin, X., Moss, R. & Dallas, J.F. (1999) Spatial distribution of genetic relatedness in moorland populations of red grouse (*Lapogus lapogus scoticus*). *Biological Journal of the Linnean Society*. **68**: 317-331.
Piry, S., Luikart, G. & Cornuet, J-M. (1999) BOTTLENECK: a computer programme for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*. **90**: 502-503.
Podolsky, R.H. (2001) Genetic variation for morphological and allozyme variation in relation to population size in *Clarkia dudleyana*, an endemic annual. *Conservation Biology*. **15**: 412-423
Porter, A.H. & Shapiro, A.M. (1991) Genetics and biogeography of the *Oeneis chryxus* complex (Satyrinae) IN California. *Journal of Research on the Lepidoptera*. **28**: 263-276.
Post, R.J., Flook, P.K. & Millest, A.L. (1993) Methods for the preservation of insects for DNA studies. Biochemical Systematics and Ecology. 21: 85-92.
Pouyaud, L., Desmarais, E., Chenuil, A., Agnese, J.F. & Bonhomme, F. (1999) Kin cohesiveness and possible inbreeding in the mouthbrooding tilapia *Sartherodon melanotheron* (Pisces: Cichlidae) *Molecular Ecology*. **8**: 803-812.
Powers, A.M., Piyapattanakorn, S., O'Riordan, R., Iyengar, I., Myers, A.A., Hawkins, S.J., Delaney, J.,

McGrath, D., Maclean, N. (1999) Verification of cyprid size as a tool in the identification of two European species of *Cthalamus* barnacles using mtDNA-RFLP analysis. *Marine Ecology Progress Series*. **191**: 251-256

Pray, L.A., Schwartz, J.M., Goodnight, C.J. & Stevens, L. (1994) Environmental dependency of inbreeding depression: Implications for conservation biology. *Conservation Biology*. **8**: 562-568.

Provan, J., Powell, W., Dewar, H., Bryan, G., Machray, G.C. & Waugh, R. (1999) An extreme cytoplasmic bottleneck in the modern European cultivated potato (Solanum tuberosum) is not reflected in decreased levels of nuclear diversity. *Proceedings of the Royal Society of London Series-B Biological Sciences*. **266**: 633-639

Pusey, A.E. (1980) Inbreeding avoidance in chimpanzees. *Animal Behaviour*. **28**: 543-552.

Queller, D.C., Strassman, J.E. & Hughes, C.R. (1993) Microsatellites and kinship. *Trends in Ecology and Evolution*. **8**: 285-288

Raijmann, L.E.L. & Menken, S.B.J. (2000) Temporal variation in the genetic structure of host-associated populations of the small ermine moth *Yponomeuta padellus* (Lepidoptera, Yponomeutidae). *Biological Journal of the Linnean Society*. **70**: 555-570

Ralls, K. & Ballou, J.D. (1986) Captive breeding programmes for populations with a small number of founders. *Trends in Ecology and Evolution*. **1**: 19-22.

Ramser, J., Weising, K., Chikaleke, V. & Kahl, G. (1997) Increased informativeness of RAPD analysis by detection of microsatellite motifs. *Biotechniques*. **23**: 285-290

Ray, J. (1710) *Historia Insectorum*. London.

Raymond, M. & Rousset, F. (1995) *GENEPOP* (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*. **86**. 248-249

Reeve, H.K., Westneat, D.F. & Queller, D.C. (1992) Estimating average within-group relatedness from DNA fingerprints. *Molecular Ecology*. **1**: 223-232.

Regner, F., Wiedeck, E. & Stadlbauer, A. (2000). Differentiation and identification of White Riesling clones by genetic markers. *VITIS* , 39, 103-107.

Roderick, G.K. (1996) Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annu. Rev. Entomol.* **41**: 325-352

Rodriguez, C., Piccinali, R., Levy, E. & Hasson, E. (2000) Contrasting population genetic structures using allozymes and the inversion polymorphism in *Drosophila buzzatii*. *Journal of Evolutionary Biology*. **13**: 976-984.

Roehrdanz, R.L. (1995) Amplification of complete insect mitochondrial genome in two easy pieces, *Insect Molecular Biology*, **4**: 169-172.

Rogers, A.R. and Harpending, H. (1992) Population Growth makes Waves in the Distribution of Pairwise genetic Differences. *Molecular Biology and Evolution*. **9**: 552-569

Roland, J., Keyghobadi, N. & Fownes, S. (2000) Alpine *Parnassius* butterfly dispersal: effects of lanscape and population size. *Ecology*. **81**: 1642-1653.

Rose, O.C., Brooks, M.I. & Mallet, J.B. (1994) A quick and non-lethal method for extracting DNA from butterfly wings. *Molecular Ecology*. **3**: 275.

Rosenberg, R. (1989) Genetic differentiation among populations of Weidemeyer's admiral butterfly. *Canadian Journal of Zoology*. **67**: 2294-2300.

Roy, M.S., Girman, A.C., Taylor A.C. & Wayne, R.K. (1994) The use of museum species to reconstruct the genetic variability and relationships of extinct populations. *Experientia* **50**: 551-557.

Rozen, R. & Skaletsky, H.J. (1996,1997) Primer3. Code available at http://wwwgenome.wi.mit.edu/genome_software/other/primer3.html .

Rusterholz, .H.P. & Erhardt, A. (2000) Can nectar properties explain sex-specific flower preferences in the Adonis Blue butterfly *Lysandra bellargus*? *Ecological Entomology*. 25: 81-90

Ruzzante, DE. (1998) A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. *Canadian Journal of Fisheries and Aquatic Science*. **55**: 1-14

Saccheri, I.J., Brakefield, P.M. & Nichols, R.A. (1996) severe inbreeding depression and rapid fitness rebound in the butterfly *Bicyclus anynana* (Satyridae). *Evolution*, **50**: 2000-2013

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, I.J., Wilson, I.J., Nichols, R.A., Bruford, M.W. & Brakefield, P.M. (1999) Inbreeding of bottlenecked butterfly populations: estomation using the likelihood of changes in marker allele frequencies. *Genetics*, **151**: 1053-1063

Saiki, R.K., Gelfland, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science,* **239**: 487-491.

Sambrooke, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual. 2nd edition*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Scarpassa, V.M., Tadei, W.P. & Suarez, M.F. (1996) Allozyme differentiation among allopatric populations of *Anopheles nuneztovari* (Diptera, Culicidae). *Brazilian Journal of Genetics.* **19**: 65-269.

Schlotterer, C. (1998) Microsatellites (in Molecular Genetic Analysis of Populations – a Practical Approach, ed AR Hoelzel IRL Press)

Schneider, S., Kueffer, J.M., Roessli, D.& Excoffier, L. (1997). *ARLEQUIN* : A software for population genetic data analysis. University of Geneva, Switzerland, Genetics and Biometry Laboratory.

Schneider, S., Roessli, D. & Excoffier, L (2000) *ARLEQUIN* , version 2.000: a software for population genetics data analysis. Genetics & Biometry Laboratory, University of Geneva, Switzerland.

Segura, J.A.L. (2000) Highly polymorphic DNA markers in an Africanized honey bee population in Costa Rica. *Genetics and Molecular Biology.* **23**: 317-322.

Shchyolkina AK, Borisova OF, Livshits MA, Pozmogova GE, Chernov BK, Klement R, Jovin TM (2000) Parallel-stranded DNA with mixed AT/GC composition: Role of trans G center dot C base pairs in sequence dependent helical stability. *Biochemistry*. **39**: 10039-10044

Shriver MD, Jin L, Chakraborty R and Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* **139** : 983-993

Simon, J.C., Herbert, P.D.N., Carillo, C. & Demelo, R. (1995) Lack of clonal variation among Canadian populations of the corn leafaphid, *Rhopalosiphum-maidis fitch* (Homoptera: aphididae). *Canadian Entomologist*. **127**: 623-629

Simon, S., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting, phylogenetic utility of mitochondrial sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*. **87**: 652-701

Slatkin, M. (1977) Gene flow and genetic drift in a species subject to local extinctions. *Theoretical Population Biology.* **12**: 253-262.

Slatkin, M. (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics.* **16**: 393-430.

Slatkin, M. (1987) Gene flow and the geographical structure of natural populations. *Science*. **236**: 787-792.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **136**: 457-462.

Smith, D.R. & Brown, W.M. (1990) Restriction endonuclease cleavage site and length polymorphisms in mitochondrila DNA of *Apis mellifera* and *A.m. carnica* (Hymenoptera: Apidae). *Annals of the Entomological Society of America*. **83**:81-88.

Sokal, R.R., Oden, N.L. & Barker, J.S.F. (1987) Spatial structure in *Drosophila buzzatii* populations: simple and spatial autocorrelation. *American Naturalist*. **129**: 122-142.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry, the principles and practice of statistics in Biological research.* W.H. Freeman and Company, New York.

Soule, M.E., editor. (1987) *Viable populations for conservation*. Cambridge University Press, Cambridge, UK.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*. **98**: 503-517

Spencer, C.C., Neigel, J.E. & Leberg, P.L. (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology.***9**: 1517-1528.

Sperling, F.A.H. & Harrison, R.G. (1994) Mitochondrial DNA variation within and between species of the *Papillio machaon* group of swallowtail butterflies. *Evolution*, **48**: 408-422

Stefanini, F.M. & Feldman, M.W. (2000) Bayesian estimation of range for microsatellite loci. *Genetical*

Research, **75**: 167-177
Stephens JC, Gilbert DA, Yuhki N and O'Brien SJ (1992) Estimation of heterozygosity for single-probe multilocus DNA fingerprints. Molecular Biology and *Evolution* 9 : 729-743
Stewart, K.E.J., Bourn, N.A.D., Pearman, G.S. & Roberts, E. (2000) *status of the Adonis blue, Polyommatus (Lysandra)bellargus, in Dorset between 1997 and 2000. Volume 2 – update 2000.* Butterfly Conservation.
Stockley, P., Searle, J.B., MacDonald, D.W. & Jones, C.S. (1993) Female multiple mating behaviour in the common shrew as a strategy to reduce inbreeding. *Proceedings of the Royal Society of London, Series B.* **254**: 173-179.
Sula, J. & Spitzer, K. (2000) Allozyme polymorphism in isolated populations of the moth *Coenophila subrosea* (Lepidoptera : Noctuidae) from three Central European peat bogs. *European Journal of Entomology*. **97**: 7-12.
Sunnucks, P., & Hales, D.F. (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobium* (Hemiptera: Aphididae). *Molecular Biology & Evolution*. **13**: 510-524
Sutcliffe, O.L., Thomas, C.D. & Peggie, D. (1997) Area dependant migration by ringlet butterflies generates a mixture of patchy population and metapopulation attributes. *Oecologia*. **109**: 229-234.
Sutherlands, G.R. & Richards, R.I. (1995) Simple tandem DNA repeats and human genetic disease. *Proc. Natl. Acad. Sci. USA*. **92**: 3636- 3641
Tajima, F. (1983) Evolutionary Relationship of DNA sequences in finite populations. *Genetics*. **105**: 437-460.
Takezaki, N. & Nei, M. (1996) Genetic Distances and Reconstruction of Phylogenetic Trees from Microsatellite Data. *Genetics*. **144**: 389-399
Tamura, K. (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. *Molecular Biology and Evolution*. **9**: 678-687.
Tan, An-Ming. & Orrego, C. (1992) DNA stabilisation and amplification from museum collections of extracts originally intended for allozyme analysis. Molecular Ecology, 1: 195-197.
Tautz, D. & Renz, M. (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*. **12**: 4127- 4138
Taylor, A.C., Marshall Graves, J., Murray, N.D., O'Brien, S.J., Yukhi, N. & Sherwin, B. (1997) Conservation genetics of the Koala (*Phascolarctos cinereus*): low mitochondrial DNA variation amongst southern Australian populations. *Genetical Research*, **69**: 25-33.
Taylor, A.C., Sherwin, W.B. & Wayne, R.K. (1994) Genetic variation of simple sequence in bottlenecked species: the decline of the northern hairy nosed wombat (*Lasiorhinus krefftii*). *Molecular Ecology*, **3**: 277-290
Taylor, B.L., Chivers, S.J., Sexton, S. & Dizon, A.E. (2000) Evaluating dispersal estimates using mtDNA data: Comparing analytical and simulation approaches. *Conservation Biology*. **14**: 1287-1297.
Taylor M.F.J, McKechnie, S.W., Pierce, N. & Kreitman, M. (1993) The Lepidopteran Mitochondrial Control Region: Structure and Evolution. *Molecular Biology and Evolution*. **10**: 1259-1272
Thomas, C.D. (1995) Ecology and conservation of butterfly metapopulations in the fragmented British Landscape. In *The Ecology and Conservation of Butterflies.* (ed. A.S. Pullin) pp. 46-63. Chapman & Hall, London.
Thomas, C.D. & Harrison, S. (1992) Spatial dynamics of a patchily distributed butterfly species. *Journal of Animal Ecology*. **61**: 437-446.
Thomas, C.D., Thomas, J.A. & Warren, M.S. (1992) Distribution of vacant and occupied habitats in fragmented landscapes. *Oecologia*. **92**: 563-567
Thomas, C.D., Hill, J.L. & Lewis, O.T. (1998) Evolutionary consequences of habitat fragmentation in a localized butterfly. *Journal of Animal Ecology*. **67**: 485-497.
Thomas, J.A.. (1983a) The ecology and conservation of *Lysandra bellargus* (Lepidoptera: Lycaenidae) in Britain. *Journal of Applied Ecology*, **20**: 59-83
Thomas, J.A. (1983b) A quick method for estimating butterfly numbers during surveys. *Biological Conservation*, **27**, 195-211.
Thomas, J.A. (1984) The conservation of butterflies in temperate countries: past efforts and lessons for the future. In: Vane-Wright, R.I. & Ackery, P.R. (eds.) *The Biology of Butterflies*. London: Academic.

pp 333-353
Thomas, R.H., Schaffner, W., Wilson, A.C., & Paabo, S. (1989) DNA phylogeny of the extinct marsupial wolf. *Nature*. **340**: 465-467
Thomas, W.K., Paabo, S., Villablanca, F.X. & Wilson, A. (1990) Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. *Journal of molecular Evolution*. **31**: 101-112.
Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997)The ClustalX windows interface: flexible strategies for multiple sequencealignment aided by quality analysis tools. Nucleic Acids Research, 24:4876-4882.
Thornhill, N.W. Editor. (1993) *The natural history of inbreeding and outbreeding: Theoretical and empirical perspectives.* Chicago: University of Chicago Press. 575pp.
Traut, W. & Epplen, J.E. (1992) Inheritance and mutation of hypervariable $(GATA)_n$ microsatellite loci in a moth *Ephestia kuehniella. Genome,* **35**: 659-666
Upholt, W.B. & David, I.B. (1977) Mapping of mitochondrial DNA of individual sheep and goats: rapid evolution of the D-loop regions of mitochondrial DNA. *Cell*. **11**: 571-583.
Valdes, A.M., Slatkin, M. & Freimar, N.B. (1993) Allele frequencies at microsatellite loci revisited. *Genetics* **133** : 737-749
Van der Peer, Y.J-M., Neefs, P., De Rijk, P. & De Wachter, R. (1993) Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. *Journal of Molecular Evolution*. **37**: 221-232.
Van Treuren, R (1998) Estimating null allele frequencies at a microsatellite locus in the oystercatcher (*Haemotopus ostralegus*). *Molecular Ecology*. **7**: 1413-1417
Vandewoestijine, S., Neve, G. & Baguette, M. (1999) Spatial and temporal population genetic structure of the butterfly *Aglais urticae*
Vawter, L. & Brown, W. (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science*, **234**: 194-196
Viard, F., Justy, F. & Jarne, P. (1997) Population dynamics inferred from temporal variation at microsatellite loci in the selfing snail *Bulinus truncates. Genetics*. **146**: 389-399.
Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandalee, T., Hornes, M., Frijters, A., Pot, A., Peleman, J., Kuiper, J. & Zabeau, M. (1995) AFLP- a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**: 4407-4414.
Vrijjenhoek, R.C. (1994) Genetic diversity and fitness in small populations. Pp. 37-53. In *Conservation Genetics.* Basel: Birkhauser Verlag.
Wade, M.J. & McCauley, D.E. (1988) Extinction and recolonisation: their effects on the differentiation of local populations. *Evolution*. **42**: 995-1005.
Waits L., Taberlet, P., Swenson, J.E., Sandegren, F. & Franzen, R. (2000) Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus ursus*). *Molecular Ecology*. **9**: 421-431.
Wang, J.S. & Grace, J.K. (2000) Allozyme patterns of protozoan symbionts of *Coptotermes formosanus* (Isoptera : Rhinotermitidae). *Sociobiology*. **35**: 467-474
Warren, M.S. (1992) The Conservation of British Butterflies. In: Dennis, R.L.H. (ed.) *The ecology of butterflies in Britain*. Oxford University Press, Oxford. pp 247-274.
Warren, M.S. (1987) The Nature Conservancy Council's Butterfly site register for the Isle of Wight.
Warren, M.S. & Bourn, N.A.D. (1997) *The impact of grassland management in environmentally sensitive areas*, (ed. R.D. Sheldrick), pp. 138-143. British Grassland Society.
Wauters, L., Hutchinson, Y., Parkin, D.T. & Dhondt, A.A. (1994) The effects of habitat fragmentation on demography and the loss of genetic variation in the red squirrel. *Proceedings of the Royal Society of London, B- series*. **255**: 107-111.
Wayne, R.K., & Jenks, S.M. (1991) Mitochondrial DNA analysis implying extensive hybridisation of the endangered red wolf *Canis rufus. Nature*. **351**: 565-568.
Watt, W.B. (1994) Allozymes in evolutionary biology- self imposed burden or extraordinary tool. *Genetics* **136**: 11-16
Weber, J.L. & Wong, C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics*. **2**: 1123-1128.

Wehrhehn, C. (1975) The Evolution of Selectively Similar Electrophoretically Detectable Alleles in Finite Natural Populations. *Genetics*. **80**: 375-394

Weir, B.S. & Cockerham, C.C (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. **38**:1358-1370

Welsh, J. & McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**: 7213-7219

Wenburg, J.K., Olsen, J.B. & Bentzen, P. (1996) Multiplexed Systems for Microsatellites for Genetic Analysis in Coastal Cutthroat Trout (*Oncorhynchus clarki clarki*). *Molecular Ecology*. **7**: 733-749

Whitfield, J.B. (1996) Destructive sampling and information management in molecular systematic research: an entomological perspective: in Managing the modern Herbarium (Metsger, D., Byers, S. and Wood, E., eds). Royal Ontario Museum Press.

Whitlock, M.C. (1999) Indirect measures of gene flow and migration: F-ST not equal to 1/[4Nm+1]. *Heredity*. **82**: 117-125.

Whitlock, M.C. & McCauley, D.E. (1990) Some population consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*. **44**: 1717-1724.

Whitlock, M.C. & McCauley, D.E. (1990) Some population consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*. **44**: 1717-1724.

Wilkerson, R.C., Parsons, T.J., Albright, D.G., Klein, T.A. & Braun, M.J. (1993) Random amplified polymorphic DNA (RAPD) markers readily distinguish cryptic mosquito species (Diptera: Culicidae: *Anopheles*). *Molecular Ecology* **1**: 205-211.

Wilkes, B. (1741-1742) *The British Aurelian. Twelve new designs of English butterflies*. London.

Wilkes, B. (1747-1749) *The English moths and butterflies*. London.

Wilkinson, G.S., Mayer, F., Kerth, G. & Petri, B. (1997) Evolution of repeated sequence arrays in the d-loop region of bat mitochondrial DNA. *Genetics*. **146**: 1035-1048.

Williams, J.K.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A. & Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers as useful genetic markers. *Nucleic Acids Research,* **18**: 6531-6535

Wilson, I.J. & Balding, D.J. (1998) Geneological inference from microsatellite data. *Genetics*. **150**: 499-510

Wolstenholme, D.R. (1992) Animal mitochondrial DNA: structure and evolution, pp. 173-216. *In* Jeon, K.W. & Wolstenholme, D.R. (eds), *Mitochondrial Genomics*. International Revue of Cytology. **141**: 173-216.

Wolstenholme, D.R. & Clary, D.O. (1985) Sequence evolution of *Drosophila* mitochondrial DNA. *Genetics*. **109**: 725-744.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics*, **10**: 97-159

Wright, S. (1940) Breeding structure of populations in relation to speciation. *American Naturalist.* **74**: 232-248.

Wright, S. (1943) Isolation by distance. *Genetics.* **28**: 139-156.

Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics* **15**: 323-354

Wright, S. (1969) Evolution and genetics of populations. Vol. 2. *The Theory of gene frequencies.* University of Chicago Press, Chicago.

Wright, S. (1977) *Evolution and the genetics of populations,* Vol. 3. *Experimental results and evolutionary deductions.* Chicago: University of Chicago Press.

Wright, S. (1978) *Evolution and the genetics of populations,* Vol. 4. *Variability between and among natural populations.* Chicago: University of Chicago Press.

Yu, K. & Peter Pauls, K. (1994) The use of RAPD analysis to tag genes and determine relatedness in heterogeneous plant populations using tetraploid alfalfa as an example. Chapt. 23 In: *PCR technology: current innovations*. CRC Press Inc.

Zhou, X.F., Faktor, O., Applebaum, S.W. & Coll, M. (2000) Population structure of the pestiferous moth *Helicoverpa armigera* in the Eastern Mediterranean using RAPD analysis. *Heredity*. **85**: 251-256.

Zakharov, E.V., Chelomina, G.N. & Zhuravlev, YN. (2000) Isolation and analysis of DNA from museum specimens of butterflies (Lepidoptera: Papillionidae) with the aid of polymerase chain reaction using arbitrary and universal gene-specific primers. *Russian Journal of Genetics*. **36**: 1016-1024

Zhang, D.X. & Hewitt, G.M. (1997) Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology.* **25**: 99-120

Zimmermann, M., Wahlberg, N. & Descimon, H. (2000) Phylogeny of Euphydryas checkerspot butterflies (Lepidoptera : *Nymphalidae*) based on mitochondrial DNA sequence data. *Annals of the* Entomological Society of America. **93**: 347-355.