Molecular Ecology (2003) **12**, 3349–3357 doi: 10.1046/j.1365-294X.2003.02012.x

Microsatellite markers to assess the influence of population size, isolation and demographic change on the genetic structure of the UK butterfly *Polyommatus bellargus*

G. L. HARPER*, N. MACLEAN† and D. GOULSON†

**Biodiversity and Ecological Processes Research Group, Cardiff School of Biosciences, Main Building, University of Cardiff, Museum Avenue, PO Box 915, Cardiff CF10 3TL, UK; †Biodiversity and Ecology Division, School of Biological Sciences, University of Southampton, Southampton, UK*

Abstract

Five microsatellite DNA markers were isolated and used to quantify population genetic structure among a subset of UK populations of the Adonis blue (*Polyommatus bellargus* Rottemburg). Specifically, whether population size, degree of isolation or history of bottlenecking in 1976–1978 can explain current patterns of genetic variation. The butterfly is at its northern range limit in the UK, where it exists as a highly fragmented metapopulation on isolated pockets of calcareous grassland. Most populations were affected by a severe bottleneck in the late 1970s, when a drought caused the host plant (*Hippocrepis comosa*) to wilt. Mantel tests and spatial autocorrelation analysis indicated a significant effect of isolation by distance among the UK populations, a relationship that broke down at greater geographical scales (> 23.85 km), probably because of large areas of unsuitable habitat presenting barriers to gene flow. Similarly, AMOVA revealed that variation among geographical regions was almost double that observed within regions. Larger populations were found to support significantly higher levels of genetic diversity, suggesting that small populations may lose genetic diversity through drift. If, as in other butterfly species, low genetic diversity increases the probability of population extinction, then these populations are likely to be under threat. Neither isolation nor a history of bottlenecks appeared to influence genetic diversity. The results indicate that adequate population size a crucial factor in the conservation of genetic diversity in *P. bellargus* in the UK.

Keywords: invertebrate, isolation, Lepidoptera, Lycaenidae, microsatellites, metapopulation

Received 19 November 2002; revision received 18 September 2003; accepted 18 September 2003

Introduction

Knowledge of butterflies greatly exceeds that of most other invertebrate groups. For most palaearctic species we have information as to their habitat preferences and larval food plants, and in some regions, notably the UK, remarkably detailed distribution maps and records of changing abundance are available. For this reason, butterflies have been used widely as a model system for examining the population dynamics and genetics of species threatened by habitat loss and fragmentation (Debinski 1994; Hanski & Thomas 1994; Hanski *et al*. 1994; Hill *et al*. 1996; Neve *et al*. 1996; Brookes *et al*. 1997; Lewis *et al*. 1997; Sutcliffe *et al*. 1997; Saccheri *et al*. 1998).

Despite the availability of such detailed population data, most studies of genetic diversity in rare species have focused on vertebrates, and here the general consensus is that inbreeding can have a significant effect (reviewed in Frankham 1995). However, the importance of inbreeding as a factor threatening the long-term persistence of invertebrate metapopulations remains controversial. No direct evidence had been found for the importance of genetic diversity in wild invertebrate populations prior to the work of Saccheri *et al*. (1998). These authors demonstrated that populations of the butterfly *Melitaea cinxia* that had less genetic variation were more likely to become extinct. Further studies of wild invertebrate populations are needed to

Correspondence: G. L. Harper. Fax: +44 292 0 874305, E-mail: harpergl@cf.ac.uk.

establish the effects of isolation and population size on genetic diversity, and the relationship between genetic diversity within populations and extinction rate. Butterflies would be ideal systems for such studies, were it not for the difficulty in finding appropriate genetic markers. Unlike many other molecular approaches, microsatellites are highly variable selectively neutral markers, making them ideal for studying intraspecific variation. However, they have proved to be hard to isolate in Lepidoptera and often have elevated levels of null alleles. (Palo *et al*. 1995; Meglecz & Solignac 1998; Keyghobadi *et al*. 1999). Despite this, we recently isolated five polymorphic microsatellite loci from the butterfly *Polyommatus bellargus* (Harper *et al*. 2000).

P. bellargus is at the northwestern edge of its range in southern England, and is restricted to south facing escarpments of unimproved calcareous grassland (Thomas 1983). More than 90% of unimproved lowland grassland was lost in Britain between 1932 and 1984, and the remaining area of chalk grassland in England is < 40 000 ha (Fuller 1987; Keymer & Leach 1990). Thus, *P. bellargus* underwent a slow decline between the 1950s and 1970s, followed by an acute decline in the late 1970s following the drought in 1976 (Thomas 1983; Emmet & Heath 1990; Asher *et al*. 2001). More than 90% of recorded UK populations had become extinct by 1981, and many more underwent severe bottlenecks in 1976–1978. The species has since recovered slowly, but its status remains a cause for concern.

Surviving populations vary greatly in size and degree of isolation, although populations rarely exceed 1000 individuals (Emmet & Heath 1990; Bourn *et al*., personal observation; Stewart *et al*. 2000). The species is markedly sedentary, so that even where populations are in close proximity, transfer of individuals is rare, with observed dispersal distances generally < 25 m (Thomas 1983). Thus, we would predict that more isolated populations should receive little or no gene flow, and that smaller isolated populations may suffer from a loss of genetic variation through drift.

Here we report patterns of genetic variation across the UK range of *P. bellargus* as measured from the five polymorphic microsatellite loci. In particular, we test whether population size, degree of isolation or history of bottlenecking in 1976–1978 explain current patterns of genetic variation.

Materials and methods

Sample collection was carried out over two years, in 1998 and 1999 (four consecutive generations of the butterfly). Twenty-six populations were sampled, comprising a total of 1173 specimens, and spanning the geographical distribution of the species in the UK (Fig. 1). The detrimental effects of sampling were minimized by removing just a small proportion of the spent male butterflies from each

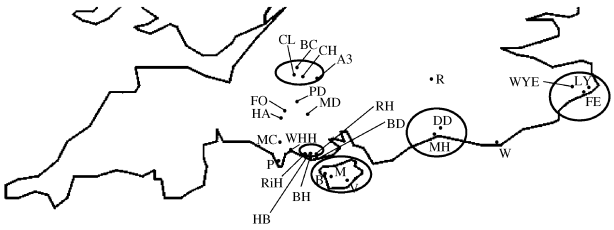

Fig. 1 The study sites in southern England. For AMOVA, populations were grouped into nine geographical regions: East Dorset (BD, BH, HB, RH, RiH), West Dorset (MC, WHH, P), Isle of Wight (V, B, M), Surrey (R), South Wiltshire (FO, HA, PD, MD), North Wiltshire (A3, CL, BC, CH), Kent (FE, WYE, LY), West Sussex (MH, DD) and East Sussex (W). Population groupings from NJ consensus tree (Fig. 3) are also indicated.

population. Population data detailing annual population size (see Table 3), historical bottlenecking (during 1978), and identification of recently founded populations (colonized since 1978) were obtained from numerous unpublished detailed historical records compiled by Butterfly Conservation, The National Trust and English Nature.

Heads from *Polyommatus bellargus* specimens were used as a source of DNA via a salt-extraction technique (after Sunnucks & Hales 1996; Aljanabi & Martinez 1997). The head was homogenized in the presence of 300 µL TEN (250 mM NaCl, 50 mM Tris–HCl, 10 mM EDTA, pH 8.0)/2% SDS and 30 µL of 1 mg/mL proteinase K. This was followed by incubation for 3 h at 55 °C. One hundred microlitres of 5 M NaCl was added to each sample, which was then shaken on a vortex prior to centrifugation at 16 000 *g* for 10 min. The supernatant was then aspirated and pipetted into a fresh tube where the DNA was precipitated using 1 mL of absolute ethanol at –20 °C. After precipitation, the ethanol was decanted off and the DNA pellets were washed twice in 1 mL of 70% ethanol. After both ethanol washes, all the alcohol was removed and the samples air dried and resuspended in 20–40 µL of TE.

Five microsatellites were used. Primer sequences and reaction conditions followed Harper *et al*. (2000). The microsatellites were scored using a 5% denaturing polyacrylamide gel by vertical electrophoresis at 20–60 mA for 2 h on a Perkin–Elmer ABI 377 automated sequencer running GENESCAN and GENOTYPER 2.5 software (PE-Applied Biosystems).

Statistical analysis

Samples collected from more than one generation at the same collecting site (16 sites in total) were assessed for differences between generations using the exact *G*-test advocated by Goudet *et al*. (1996). A total of 5000 permutations of genotypes among samples were performed to assess the significance of temporal differentiation. Computations were performed using FSTAT 2.9.1. (Goudet 2000) updated from Goudet (1995).

Departures from the Hardy–Weinberg equilibrium (HWE) for all 26 populations at each locus were tested using ARLEQUIN (Schneider *et al*. 2000). Where a significant heterozygote deficit was found, estimations of null allele frequency (*n*) were made using the method of Brookfield (1996).

To test for evidence of a recent reduction in effective population size (N_e), the microsatellite data were analysed using BOTTLENECK 1.2.02 (Cornuet & Luikart 1996; Piry *et al*. 1999). This program applies the assumption that within a well-defined population, e.g. with no immigration or population substructure, allele numbers at selectively neutral loci (*k*) will generally decline faster than observed heterozygosity (H_O) (Hedrick *et al*. 1986). Thus, recently bottlenecked populations will display an excess of heterozygosity relative to allele number (Cornuet & Luikart 1996; Luikart & Cornuet 1998; Luikart *et al*. 1998, 1999; Spencer *et al*. 2000). Statistics for all three models of mutation, stepwise (SMM), two-phase (TPM) and the infinite allele (IAM), were calculated. A Wilcoxon signed-rank test was performed to test the hypothesis that the average standardized 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 BOTTLENECK (Cornuet & Luikart 1996; Piry *et al*. 1999).

Estimates of gene diversity (Nei 1987) were made for each of the 5 loci and 26 populations using 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 confounding factors such as sample size, the effects of null alleles, breeding system or the Wahlund effect. Gene diversity was analysed with respect to three potential explanatory factors: (i) population size (assigned as small, medium or large); (ii) whether the population bottlenecked was the result of a recent colonization event or had a stable history; (iii) isolation from other populations, measured as the number of other known populations located within 37.5 km (Table 1). The data were analysed in GLIM with

Table 1 Wilcoxon signed rank tests for excess heterozygosity for 26 *Polyommatus bellargus* populations across the UK. Details of population size (small medium or large), number of sampled local populations (within 37.5 km), Nei's (1987) gene diversity, sample size, numbers of loci exhibiting an excess of heterozygosity, calculated *P*-value, population histories since 1976–1978 (R = recolonized, B = bottlenecked, S = stable, ? = unknown); significance ($P < 0.05$) are all shown. IAM, infinite allele model; TPM, two-phase model

Pop'n	Sample size	% loci with an excess		*P*-value		History	pop'n size	Local pop'n	Gene diversity
		IAM	TPM	IAM	TPM				
A3	35	4	1	0.31	0.98	?	M	6	0.6702
BD	57	5	1	0.02*	0.98	S	L	10	0.7056
BH	51	3	1	0.59	0.98	S	L	8	0.5992
BC	16	3	0	0.69	1.0	B	S	7	0.6448
B	46	3	1	0.41	0.97	S	L	2	0.6878
CL	52	5	5	0.02*	0.02*	R	L	7	0.7114
CH	49	3	2	0.5	0.89	S	M	7	0.637
DD	77	4	1	0.03*	0.95	?	L	2	0.7392
FE	74	5	2	0.02*	0.92	B	L	2	0.7456
FO	11	5	4	0.02*	0.07	B	M	13	0.7228
HA	42	4	2	0.07	0.92	B	L	14	0.7422
HB	41	4	0	0.07	1.0	?	M	9	0.6506
LY	49	4	3	0.03*	0.5	B	M	2	0.6994
MC	60	4	3	0.04*	0.67	R	L	9	0.7292
MD	43	4	2	0.31	0.95	B	S	10	0.6666
MH	36	3	2	0.11	0.92	B	M	2	0.6902
M	49	3	1	0.41	0.97	S	M	2	0.5972
P	25	2	1	0.59	0.97	?	S	7	0.6088
PD	43	4	3	0.04*	0.69	?	M	7	0.684
R	56	4	2	0.07	0.89	S	L	0	0.76
RIH	16	5	2	0.02*	0.92	?	M	9	0.7106
RH	52	2	1	0.68	0.98	?	M	10	0.6524
V	43	3	2	0.31	0.95	S	M	2	0.6646
WHH	57	3	1	0.31	0.95	S	L	9	0.7022
W	47	4	2	0.11	0.59	?	L	2	0.7862
WYE	32	4	3	0.31	0.59	R	M	2	0.664

normal errors (equivalent to a three-factor ANOVA) (see Crawley 1993). The categorizations of population size and demographic change were based on various unpublished surveys from Butterfly Conservation and English Nature.

Estimates of genetic distance between populations were obtained using Cavalli-Sforza & Edwards (1967) chord distance (D_C) calculated in MICROSAT 1.5b (Minch 1997). 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). In particular, this measure outperformed distances designed for the IAM and SMM (Takezaki & Nei 1996). Neighbour-joining (NJ) trees were constructed using D_C. One thousand distance matrices from resampled data sets bootstrapped over loci were created using MICROSAT. These were analysed using the NJ tree subroutine in NEIGHBOUR within PHYLIP 3.57c (Felsenstein 1993). The input order of populations was randomised to ensure that the final tree topology was not dependent on the order the samples were entered. 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.

To establish whether any isolation by distance effect has occurred, matrices of genetic distance data (D_C) and the logarithms of geographical distance data (in km) between all the UK sample sites were constructed. These matrices were analysed for their degree of correlation using a Mantel test, with significance tests performed over 10 000 randomizations (Mantel 1967; Manly 1991, 1997). The analysis was carried out using MANTEL3. This program uses a Pascal translation of the FORTRAN code found in Manly (1991). It is available free from J. Goudet (Université de Lausanne).

The relationship between genetic and geographical distance was further investigated using a spatial autocorrelation. For this analysis, the method of Smouse & Peakall (1999) was used, an approach specifically designed for multi-allelic codominant loci, such as microsatellites. Unlike classical spatial autocorrelation analysis, which is calculated one allele at a time (e.g. Stone & Sunnocks 1993; van Staaden *et al*. 1996), the procedure is intrinsically multivariate. Thus, avoiding the need for allele by allele, locus by locus analysis. By combining alleles and loci, the spatial signal will be strengthened by reducing stochastic noise (Smouse & Peakall 1999; Peakall *et al*. 2003). A pairwise matrix of genetic distance (F_{ST}) and a pairwise geographical distance was used as input files, with distance classes of even sample size. The resulting autocorrelation coefficient generated, *r*, is closely related to Moran's *I*. Tests for statistical significance were performed by random permutation; this generates an estimate of *r* around the null hypothesis of no spatial structure (*rp*). After 1000 permutations, the *rp* values are sorted, and the 25th and 975th *rp* values are taken to define the upper and low bounds of the 95% confidence interval. This analysis was performed using the package GENAIEX V5 (Peakall & Smouse 2001).

The microsatellite data were subjected to a hierarchical analysis of molecular variance (AMOVA), as described by Excoffier *et al*. (1992), using three hierarchical levels; individual, population and geographical region (Fig. 1). The analysis was performed using ARLEQUIN 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.

Results

Of the 16 populations tested for temporal variation in allele frequencies, none showed any significant overall loci differentiation between generations (exact *G*-test, $P > 0.05$). For subsequent analyses, samples from different generations have been pooled.

Of the five microsatellites analysed, there were deviations from HWE in at least a subset of the 26 populations (75% of locus/population combinations). Deviations were all heterozygote deficiencies. When the frequencies of null alleles were calculated and averaged across all sites, the frequencies were 0.2 ± 0.025, 0.18 ± 0.026, 0.03 ± 0.013, 0.09 ± 0.023 and 0.07 ± 0.022 for loci *Lb1/41*, *Lb1/57*, *LbG2*, *Lb4/18* and *Lb4/19*, respectively (means ± SE). Most populations that deviated from HWE displayed null allele frequencies of around 0.15–0.3; these frequencies are similar to those in other published studies where frequencies have been calculated (Neumann & Wetton 1996; Van Treuren 1998).

Of the 26 populations analysed for evidence of a bottleneck, the TPM indicated that one population showed evidence of a heterozygote excess (Table 1). However, under the IAM, estimates of heterozygosity excess were significant for nine populations (Table 1). Under the SMM, no populations indicated heterozygosity excesses (results not shown). None of the allele distributions showed a departure from the standard L-shape in the mode-shift test.

Of the populations known to have undergone a founder event or population bottleneck, 60% showed a significant 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 (Table 1).

The GLIM analysis revealed significantly higher population gene diversity in larger populations (means ± SD; 0.64 ± 0.02, 0.67 ± 0.01 and 0.72 ± 0.01, for small, medium and large populations, respectively) ($F_{2,23} = 6.34$, $P < 0.01$).

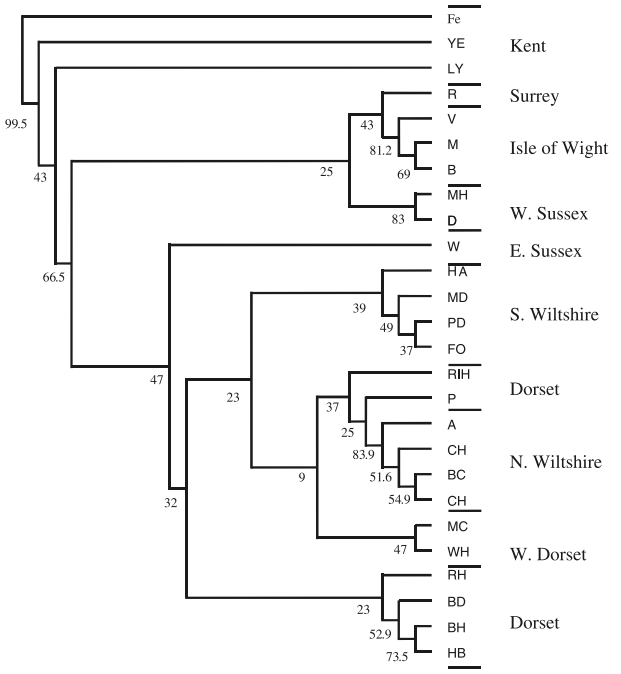

Fig. 2 Consensus tree based on Cavalli-Sforza & Edwards' (1967) chord distances derived from allele frequencies at five microsatellite loci (NJ method of tree construction). Bootstrap values are indicated at each node.

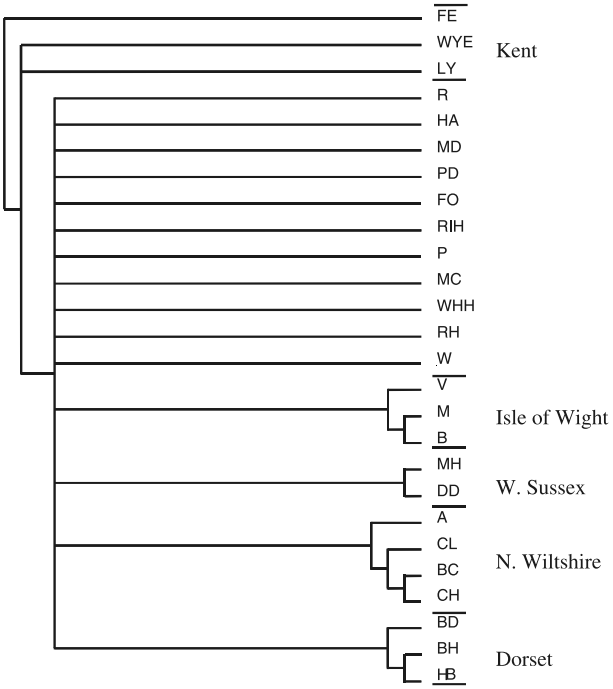

Fig. 3 NJ consensus tree based on Cavalli-Sforza & Edwards' (1967) chord distance. Branches not supported by a bootstrap value of at least 50% have been collapsed.

There were no significant differences in gene diversity according to a history of bottlenecking or degree of isolation.

Phylogenetic reconstructions using Cavalli-Sforza & Edwards' (1967) chord distance resulted in a consensus NJ tree that was only weakly supported at a large number of nodes (< 50%). However, the general trends apparent in this dendogram (Fig. 2) comply well with the geographical distribution of the populations sampled. There were some anomalies, notably the positioning of population 'R' from Surrey in the Isle of Wight clustering. Similarly, two populations from southern Dorset, 'RiH' and 'P' were grouped with populations from northern Wiltshire. After the tree was made more robust, by collapsing nodes supported by < 50% of bootstraps, all unexpected relationships were broken down and five major groupings of populations were revealed, located in the Isle of Wight, Sussex, Wiltshire, Kent and Dorset (Fig. 3). 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.

The Mantel test results gave an *r*-value of 0.3516 (P = 0.0007, for 20 000 randomizations), indicating that there is a strong isolation-by-distance effect present among the populations. However, on further investigation using spatial autocorrelation analysis, this relationship is not of a

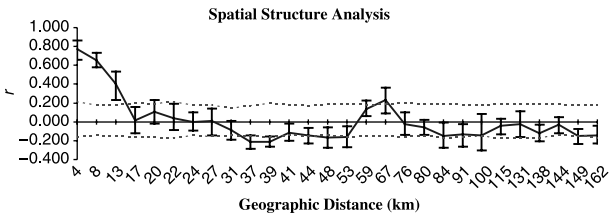

Fig. 4 Correlogram showing the combined genetic correlation (*r*) as a function of distance, 95% CI about the null hypothesis of a random distribution of genotypes, and 95% confidence error bars about r as determined by bootstrapping.

linear nature (Fig. 4). The correlogram shows a sharp, nonlinear decline in *r* across a short geographical distance, with *r* positive and significant at 4 and 8 and 13 km, and an *x*-intercept at 23.85 km. Beyond this point, the relationship remains nonsignificant.

Under the AMOVA analysis, 87% of the variance was explained by within population variation (variance component 1.30). The remainder is partitioned between variation among populations within regions (4.5%, variance component 0.067, $P < 0.0001$), and variation among regions (8.18%, variance component 0.122, $P < 0.0001$). 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.

Discussion

Heterozygote deficiency and null alleles

Genetic data obtained from natural populations of outbreeding organisms usually indicate agreement between observed and expected genotype frequencies according to HWE (Nei 1987). Where heterozygote deficiencies are found consistently across loci, this may be ascribed to a variety of factors including inbreeding and population substructuring (Lade *et al*. 1996; Paxton *et al*. 1996). However, because consistency across loci was not present in our data, it is unlikely that these processes are the sole cause of the observed disequilibrium. Another explanation for heterozygote deficiency is the presence of null alleles. The observation that a small subset of individuals failed to amplify from each locus would support this conjecture. Significant heterozygote deficiency has been found in similar studies of Lepidoptera (Palo *et al*. 1995; Meglecz & Solignac 1998; Meglecz *et al*. 1998; Keyghobadi *et al*. 1999), suggesting a high mutation rate in the regions immediately surrounding Lepidopteran microsatellite repeats.

This widespread presence of null alleles will tend to confound the calculation of many heterozygosity-based genetic distance estimators. In this situation, the application of a Euclidean genetic distance estimator, such as Cavalli-Sforza & Edwards' (1967) chord distance (D_C), may be more appropriate. These measures have also been shown to be more efficient at reconstructing evolutionary relationships using microsatellite data (Takezaki & Nei 1996; Lugon-Moulin *et al*. 1999).

Recent genetic bottlenecks and inbreeding

A consistent excess of heterozygosity, indicating a recent bottleneck, was detected under the IAM in 60% of populations in which documented evidence suggested that a bottleneck had occurred, whereas only 12.5% of populations documented as having remained stable displayed these effects. Although these sample sizes are too small to carry out a statistical test, our data suggest that this approach for detecting past bottlenecks in the absence of direct population data is reasonably effective. However, both TPM and SMM appeared to be rather poor at detecting past bottlenecks in this system.

When gene diversity values were analysed with respect to various aspects of the chronicled population data, no significant effects from population isolation, past bottlenecks or extinction/recolonizations 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 have indicated an increased likelihood of extinction in isolated butterfly populations (Saccheri *et al*. 1998; Nieminen *et al*. 2001). However, population size was strongly and positively related to genetic diversity. This effect has been hinted at in previous studies. For example, in *Plebejus argus*, allozyme and mitochondrial DNA data implied the loss of diversity from small populations with rapid turnover rates (Mallet 1997). Similarly in *Maculinea teleius* a relationship was inferred between small effective population size and low heterozygosity (Figurny-Puchalska *et al*. 2000). A relationship between population size and genetic diversity may seem trivial, but it has profound implications. It suggests that natural populations of butterflies can be sufficiently small for loss of alleles through drift. If so, then they may suffer from deleterious effects of inbreeding, and the probability of local extinction (already high in small populations) is raised. It would be valuable to follow up this work with a study of the relationship between extinction rate and genetic diversity.

Population substructuring, gene flow, and isolation by distance

The dendograms, Mantel and AMOVA analyses all indicate a pattern of connectivity among populations that relates to the geographical distance separating them, in accordance with Wright's (1943) model of isolation by distance. Although the initial phylogram reconstruction resulted in a topology that reflected the geographical 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 (< 50%), many of the relationships broke down, leaving only a few very strong population clusters. These groupings consistently represented 'local' populations, separated by short geographical distances (< 20 km). This suggestion of 'isolation by distance' appeared to be confirmed by both the Mantel and spatial autocorrelation analyses, where genetic and geographical distances were significantly correlated over short distances, with no significant relationship found between populations separated by > 13 km, suggesting that isolation is complete for distances greater than this.

The AMOVA analysis indicates that approximately twice the amount of population variance was identified between regions than was present among populations within regions. Both of these variance components were highly significant, again suggesting significant genetic structuring among populations. The overall estimate of F_{ST} (0.127) is comparable to 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* $F_{ST} = 0.10$ (Johannesen *et al*. 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 obtained similar values for Hungarian populations of *P. mnemosyne* $F_{ST} = 0.070$ (Meglecz *et al*. 1998) and for *Militaea cinxia* $F_{ST} = 0.20$ (Palo *et al*. 1995).

Our analyses unequivocally implicate geographical distance as an important barrier to gene flow. 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 geographical 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 corresponds with the ecology of *Polyommatus bellargus*, where mark–release–recapture studies have inferred that the exchange of individuals among colonies is extremely rare (Thomas 1983). Even in Dorset, where populations are in close proximity, 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 m and on the same side of any putative barriers (such as unsuitable habitat) (Thomas 1983). 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 distances across which dispersal is unlikely. This presumably explains the sharp increase in genetic differentiation across short geographical scales, beyond which no relationship remains between the two variables. The slight contrast between the inferences from the dispersal studies and the 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 second generation can be enough to prevent genetic drift between populations (Slatkin 1995), events which are extremely unlikely to be directly observed. The butterfly has been able to recolonize sites in Dorset that are 10–15 km from the nearest population (Bourn *et al*., personal observation), indicating a higher dispersal ability than that suggested by Thomas (1983).

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 (Peterson & Denno 1998). For example, in *Melitaea didyma*, habitat partitioning prevented gene flow among regions, whereas exchange between local populations remained high (Johannesen *et al*. 1996). In the sedentary lycaenid butterfly, *Euphilotes enoptes*, gene flow appeared to occur 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 parallels with our data for *P. bellargus*, where gene flow has been inferred over distances of up to 37.5 km, yet the butterfly has never been observed to stray much more than 25 m from its home range (Thomas 1983). It is likely that stepping-stone gene flow is playing a part here, but that the structure of the habitat in the UK limits this to ≈ 40 km, because stepping-stone gene flow cannot traverse areas where no populations are present.

Although 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. Conservation effort is often concentrated only on maintenance of corridors between suitable habitats. Our data suggest that population size is another crucial factor in maintaining high levels of genetic diversity, whereas connectivity between populations had a less discernible affect in this study. Thus the preservation of blocks of habitat able to sustain large populations should also be given high priority.

Acknowledgements

This work was funded by a University of Southampton studentship. Thanks to all of those people who helped locate *Polyommatus bellargus* populations, notably Butterfly Conservation, English Nature, The National Trust and English Heritage. Also to Rod Peakall and Peter Smouse for their advice on the spatial autocorrelation analysis.

References

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

Asher J, Warren M, Fox R *et al.* (2001) *The Millennium Atlas of Butterflies in Britain and Ireland*. Oxford University Press, Oxford.

Brookes MI, Graneau YA, King P *et al.* (1997) Genetic analysis of founder bottlenecks in the rare British butterfly *Plejebus argus*. *Conservation Biology*, **11**, 648–661.

Brookfield JYF (1996) A simple method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

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

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Crawley MJ (1993) *Methods in Ecology: GLIM for Ecologists*. Blackwell Science, Oxford.

Debinski DM (1994) Genetic diversity assessment in a metapopulation of the butterfly *Euphydryas gillettii*. *Biological Conservation*, **70**, 25–31.

Emmet MA, Heath J (1990) *The Butterflies of Great Britain and Ireland, 7. Part 1. Heperiidae to Nymphalidae*. Harley Books, Colchester.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

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

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

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

Fuller RM (1987) The changing extent and conservation interest of lowland grasslands in England and Wales: a review of grassland surveys 1930–84. *Biological Conservation*, **40**, 281–300.

Gilpin M (1991) The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society*, **42**, 165–175.

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.

Hanski I, Kuussaari M, Nieminen M (1994) Metapopulation structure and migration in the butterfly *Mellitaea cinxia*. *Ecology*, **75**, 757–762.

Hanski I, Thomas CD (1994) Metapopulation dynamics and conservation: a spatially explicit model applied to butterflies. *Biological Conservation*, **68**, 167–180.

Harper GL, Piyapattanakorn S, Goulson D, Maclean N (2000) The isolation of microsatellite markers from the Adonis blue butterfly (*Lysandra bellargus*). *Molecular Ecology*, **9**, 1948–1949.

Hedrick PW, Brussard BF, Allendorf FW, Beardmore JA, Orzack S (1986) Protein variation, fitness, and captive propagation. *Zoological Biology*, **5**, 91–99.

Hill JK, Thomas CD, Lewis OT (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.

Johannesen J, Veith M, Seitz A (1996) Population genetic structure of the butterfly *Melitaea didyma* (Nymphalidae) along a northern distribution border. *Molecular Ecology*, **5**, 259–267.

Keyghobadi N, Roland J, Strobeck C (1999) Influence of landscape on the population genetic structure of the Alpine butterfly *Parnassius smintheus* (Papillionidae). *Molecular Ecology*, **8**, 1481–1495.

Keymer RJ, Leach SJ (1990) Calcareous grassland — a limited resource in Britain. In: *Calcareous Grasslands: Ecology and Management* (eds Hiller SH, Walton DWH, Wells DA), pp. 11–20. Bluntisham Books, Huntingdon.

Lade JA, Murray ND, Marks CA, Robinson NA (1996) Microsatellite differentiation between Philip Island and mainland Australian populations of the red fox (*Vulpes vulpes*). *Molecular Ecology*, **5**, 81–87.

Lande R (1995) Mutation and conservation. *Conservation Biology*, **9**, 782–791.

Leberg PL (1991) Influence of fragmentation and bottlenecks on genetic divergence of wild turkey populations. *Conservation Biology*, **5**, 522–530.

Lewis OT, Thomas CD, Hill JK, Brookes MI *et al.* (1997) Three ways of assessing population structure in the butterfly *Plebejus argus*. *Ecological Entomology*, **22**, 283–293.

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, Allendorf FW, Cornuet JM, Williams BS (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**, 238–247.

Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology*, **12**, 228–237.

Luikart G, Cornuet J-M, Allendorf FW (1999) Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conservation Biology*, **13**, 523–530.

Lynche M, Conery J, Burger R (1995) Mutation accumulation and the extinction of small populations. *American Naturalist*, **146**, 489–518.

Mallet JLB (1997) Genetic analysis of founder bottlenecks in the rare British butterfly *Plejebus argus*. *Conservation Biology*, **11**, 648–661.

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

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

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

McKechnie SW, Ehrlich PR, White RR (1975) Population genetics of *Euphydryas butterflies*. I. Genetic variation and the neutrality hypothesis. *Genetics*, **81**, 571–594.

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.

Minch E (1997) MICROSTAT, *Version 1.5b*. Stanford University Medical Centre, Stanford, CA.

Napolitano M, Descimon H (1994) Genetic structure of French populations of the mountain butterfly *Parnassius mnemosyne* L. (Lepidoptera: Papilionidae). *Biological Journal of the Linnean Society*, **53**, 325–341.

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

Neumann K, Wetton JH (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.

Nieminen M, Singer MC, Fortelius W, Schops K, Hanski I (2001) Experimental evidence that inbreeding increases extinction risk in butterfly populations. *American Naturalist*, **157**, 237–244.

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.

Paxton RJ, Thoren PA, 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.

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. *Evolution*, **5**, 1182–1195.

Peakall RE, Smouse PE (2001) GENAIEX *V5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research*. Australian National University, Canberra. Available from http://www.anu.edu.au/BoZo/GenAIEx.

Peterson MA (1996) Long-distance gene flow in the sedentary butterfly *Euphilotes enoptes* (Lepidoptera: Lycaenidae). *Evolution*, **50**, 1990–1999.

Peterson MA, Denno RF (1998) The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *American Naturalist*, **152**, 428–446.

Piry S, Luikart G, Cornuet J-M (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Porter AH, Shapiro AM (1991) Genetics and biogeography of the *Oeneis chryxus* complex (Satyrinae) in California. *Journal of Research on the Lepidoptera*, **28**, 263–276.

Saccheri I, Kuussaari M, Kankare M *et al*. (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature*, **392**, 491–494.

Schneider S, Roessli D, Excoffier L (2000) ARLQUIN, Version 2.000: *A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

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.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561–573.

Spencer CC, Neigel JE, Leberg PL (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology*, **9**, 1517–1528.

van Staaden MJ, Michener GR, Chesse RK (1996) Spatial analysis of microgeographic genetic structure in Richardson's ground squirrels. *Canadian Journal of Zoology*, **74**, 1187–1195.

Stone GN, Sunnocks P (1993) Genetic consequences of an invasion through a patchy environment — the cynipid gallwasp *Andricus quercuscalicis* (Hymenoptera: Cynipidae). *Molecular Ecology*, **2**, 251–268.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobium* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510–524.

Sutcliffe OL, Thomas CD, Peggie D (1997) Area dependant migration by ringlet butterflies generates a mixture of patchy population and metapopulation attributes. *Oecologia*, **109**, 229–234.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite data. *Genetics*, **144**, 389–399.

Thomas JA (1983) The ecology and conservation of *Lysandra bellargus* (Lepidoptera: Lycaenidae) in Britain. *Journal of Applied Ecology*, **20**, 59–83.

Van Treuren R (1998) Estimating null allele frequencies at a microsatellite locus in the oystercatcher (*Haemotopus ostralegus*). *Molecular Ecology*, **7**, 1413–1417.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **10**, 97–159.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 139–156.

The work detailed in this study is a part of GLH's PhD thesis, investigating the UK colonization and population structure of *Polyommatus bellargus* using both microsatellites and mitochondrial markers. GLH's primary research interests are applying molecular markers to understanding ecological and evolutionary processes, and applying this to conservation.
