Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity

NUSHA KEYGHOBADI,*† JENS ROLAND,* and CURTIS STROBECK*

*Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada, T6G 2E9, †Department of Biology, Okanagan University College, 3333 University Way, Kelowna, British Columbia, Canada, V1V 1V7

Abstract

Levels of gene flow among populations vary both inter- and intraspecifically, and understanding the ecological bases of variation in levels of gene flow represents an important link between the ecological and evolutionary dynamics of populations. The effects of habitat spatial structure on gene flow have received considerable attention; however, most studies have been conducted at a single spatial scale and without background data on how individual movement is affected by landscape features. We examined the influence of habitat connectivity on inferred levels of gene flow in a high-altitude, meadow-dwelling butterfly, *Parnassius smintheus*. For this species, we had background data on the effects of landscape structure on both individual movement and on small-scale population genetic differentiation. We compared genetic differentiation and patterns of isolation by distance, based on variation at seven microsatellite loci, among three regions representing two levels of connectivity of high-altitude, nonforested habitats. We found that reduced connectivity of habitats, resulting from more forest cover at high altitudes, was associated with greater genetic differentiation among populations (higher estimated \( F_{ST} \)), a breakdown of isolation by distance, and overall lower levels of inferred gene flow. These observed differences were consistent with expectations based on our knowledge of the movement behaviour of this species and on previous population genetic analyses conducted at the smaller spatial scale. Our results indicate that the role of gene flow may vary among groups of populations depending on the interplay between individual movement and the structure of the surrounding landscape.

Keywords: butterfly, \( F_{ST} \), gene flow, isolation by distance, landscape, microsatellite

Received 27 October 2004; revision received 14 February 2005; accepted 1 March 2005

Introduction

Gene flow is a fundamental microevolutionary force that can determine the potential for genetic differentiation among populations and for local adaptation, and also influences the geographical spread of novel adaptations. Levels of gene flow in natural populations, and therefore its actual role in evolution, have historically been a matter of debate (Fisher 1930; Mayr 1942; Ehrlich & Raven 1969; Wright 1977). The past four decades have seen a proliferation of studies describing the genetic structure of natural populations and the accumulating body of evidence suggests there is no single role of gene flow in differentiation and adaptation, but that its role varies depending on its strength relative to other evolutionary forces and on the geographical distribution of populations (Slatkin 1985). The relative strength of gene flow may be affected by a variety of ecological factors that vary intra- and interspecifically such as dispersal ability, dietary specialization, phenological asynchrony among populations, habitat persistence, population persistence, and spatial structure of habitat within the landscape (Peterson & Denno 1998). Elucidating the factors that affect the levels and spatial extent of gene flow is clearly fundamental in understanding its role in evolution. Furthermore, determining the ecological bases
of variation in levels of gene flow represents a critical link between the ecological and evolutionary dynamics of populations.

The spatial structure of habitat within the landscape and its correlation with levels of gene flow and population differentiation have been investigated extensively, both within species (King 1987; Britten et al. 1995; Hutchison & Templeton 1999) and among species (Caccione & Sbordoni 1987; Shoemaker & Jaenike 1997). Many studies have been conducted within a conservation context, motivated by concern over the potential impacts of habitat fragmentation on genetic variation (Young et al. 1993; VanDongen et al. 1998; Knutsen et al. 2000). However, the importance of the effects of landscape factors on population genetic structure for evolutionary dynamics has also been stressed (Castric et al. 2001; Costello et al. 2003).

Many studies examining the relationship between habitat spatial structure and gene flow have been conducted at a single spatial scale or without background data on how individual movement is affected by landscape elements (but see Michels et al. 2001; Vos et al. 2001; Vucetich et al. 2001), and often both elements are lacking. Thus, it is difficult to assess whether the inferred effects of landscape structure on gene flow are consistent over different spatial scales, and with the effects of landscape structure on movement. More importantly, background data on movement behaviour is essential to establish a link between landscape connectivity, defined by landscape ecologists as ‘the degree to which the landscape facilitates or impedes movement among resource patches’ (Taylor et al. 1993) and gene flow. Landscape connectivity, thus defined, results from the interaction between an organism’s movement behaviour and the spatial structure of the landscape, and cannot be adequately assessed in the absence of movement data (Tischendorf & Fahrig 2000). Furthermore, most studies that compare gene flow among regions varying in landscape structure do not quantify the variation in landscape structure, but use qualitative assessments that are then correlated with inferred levels of gene flow.

Here, we examine how variation in landscape connectivity among three geographical regions affects population differentiation and relative levels of gene flow in the meadow-dwelling butterfly Parnassius smintheus Doubleday, 1847. Because we have previously examined the effects of landscape structure on movement of individuals (Roland et al. 2000) and on genetic differentiation of populations at a smaller spatial scale (Keyghobadi et al. 1999), we are able to determine if the effects of landscape connectivity on gene flow observed in this study are consistent with effects of landscape on movement and fine-scale population genetic differentiation. The ability to compare levels of differentiation and patterns of isolation by distance (IBD) at the large scale to those at a much finer scale (Keyghobadi et al. 1999) is also a significant aid to interpretation of results, because the spatial scale of sampling can have a considerable impact on observed patterns of genetic variation (Peterson & Denno 1998). In addition, we quantify variation in landscape connectivity among the three regions.

In P. smintheus, movement is strongly influenced by intervening landscape, and forested areas are much more resistant to movement than are open meadows (Roland et al. 2000). In this species, restricted dispersal appears to be translated into restricted gene flow. Using highly variable microsatellite DNA markers we have found significant genetic structure, in the form of IBD, at the same spatial scale as the mark–recapture study, with no more than 12 km separating the most distant populations (Keyghobadi et al. 1999). We have also found that, as for dispersal, genetic exchange among local populations is impeded to a greater extent by intervening forests than by intervening, open meadows (Keyghobadi et al. 1999). Intervening forests similarly impede gene flow in Parnassius mnemosyne (Meglécz et al. 1998) and the nymphalid butterfly, Erebia medusa (Schmitt et al. 2000).

In this study, we expand our analyses to a larger spatial scale (up to 58 km separating sites) and compare genetic differentiation of populations and inferred, relative levels of gene flow among three regions varying in the degree of connectivity of the landscape with respect to movement of P. smintheus; that is, in the degree of connectivity of high-altitude, nonforested habitats. In particular, we focus on comparing patterns of IBD among the three regions. A significant correlation between genetic similarity/dissimilarity and geographical distance, referred to as isolation by distance, is predicted by both continuous and stepping-stone models of population structure under conditions of restricted gene flow (Wright 1943). With increasing geographical distance, the influence of gene migration relative to genetic drift declines, such that both the mean and variance of genetic distances among populations increase. Comparison of patterns of IBD is a very useful approach in examining the effects of ecological factors, such as landscape connectivity or dispersal ability, on population genetic structure (Peterson & Denno 1998). This approach allows one to assess the relative influences of gene movement and genetic drift over various geographical distances (Hutchison & Templeton 1999) and is more informative than a single regional estimate of population structure or gene flow.

**Materials and methods**

**Study species**

Parnassius smintheus occurs in mountainous environments in western North America, inhabiting alpine and subalpine meadows. In our study area in the southern Canadian Rocky Mountains, populations are commonly found just at or above treeline (Fownes 1999). Larvae feed exclusively on

© 2005 Blackwell Publishing Ltd, Molecular Ecology, 14, 1897–1909
stonecrop, *Sedum lanceolatum*. Populations here are univoltine and adults are typically in flight from July to September. Although population sizes change from year to year, as in many univoltine butterflies, local populations in our study area do not experience periodic extinctions and therefore do not display true metapopulation dynamics. Using mark–recapture methods, Roland *et al.* (2000) found an average adult dispersal distance of approximately 150 m and the maximum dispersal distance observed was 1729 m.

**Study area**

Our study area included 27 sampling sites in the foothills and front ranges of the Canadian Rocky Mountains in Alberta and covered an extent of approximately $40 \times 135$ km (Fig. 1, Table 1). All sampling sites were meadows, at or above treeline, containing the larval host plant, *S. lanceolatum*.

The sampling sites can be divided among three regions, representing two very different landscape types. Eight sites were located in the region we refer to as East Kananaskis, which represents the hilly foothills landscape, dominated by forests, where open habitats above treeline are relatively small and isolated. Nine and 10 sample sites, respectively, were located in West Kananaskis and Banff, which represent the landscape of the mountain front ranges. This is a mountainous landscape that is less heavily forested than the foothills, where open habitats above treeline are much larger and better connected. Sample sites were similarly spaced in the three regions (Fig. 1), as indicated by overlapping confidence intervals for the mean pairwise distance between sites (mean ± 95%CI: 17.54 ± 4.14 in East Kananaskis, 19.20 ± 3.09 in Banff, and 25.45 ± 5.08 in West Kananaskis).

We have previously conducted a fine-scale study of population genetic structure within a small subregion of East Kananaskis (Keyghobadi *et al.* 1999). Here, genetic samples were collected from 17 sites located on three separate ridges and separated by distances of only 0.15–12 km. One site from each of the three ridges in the fine-scale sample set (sites E, Q, and Z) was shared with the large-scale East Kananaskis sample set (Fig. 1 in Keyghobadi *et al.* 1999; sites E, Q, and Z are, respectively, designated as 1, 2, and 3 in Fig. 1 here).

**Measurement of landscape metrics**

Geographical coordinates (degrees, minutes) of sampling sites were determined from 1:250 000 topographical maps and then used to calculate distances between sites using the *r* Package (Legendre & Vaudor 1991).

We measured a number of landscape metrics that describe the distribution and spatial structure of high-altitude, non-forested areas in each of the three study regions. Such areas

---

**Fig. 1** Locations of sample sites. Sites 1–8 are in East Kananaskis, sites 9–17 are in West Kananaskis, and sites 18–27 are in Banff. The three large rectangles indicate the quadrats used to measure landscape structure for each region. The small, dashed rectangle designates the location of the small-scale study area (Keyghobadi *et al.* 1999; Roland *et al.* 2000). Names of sample sites are given in Table 1.
include *P. smintheus* habitat, meadows in which the butterfly or its host plant are not found (Fownes 1999), as well as bare rock, talus, and scree. These landscape metrics are thus indices of the connectivity of the landscape with respect to potential movement of *P. smintheus*, given their reduced rates of movement through forested areas (Roland et al. 2000).

Although a large number of landscape metrics are available, the essential aspects of landscape structure/pattern can often be represented by a small number of these (Cain et al. 1997) and most metrics are based on a few basic parameters: patch size, patch shape (perimeter–area ratio), and interpatch distance (Hargis et al. 1998). In addition, for binary maps divided into suitable habitat and unsuitable habitat, the most important determinant of spatial pattern is simply the proportion of suitable habitat (Gustafson 1998).

For each region, a rectangular quadrat encompassing all of the sampling sites was delimited with the four edges placed such that they extended 2.5 km past the furthest outlying site in each direction (Fig. 1). This border represents a predicted maximum dispersal distance based on observed maximum movements of 1.5–2 km (Guppy 1986; Roland et al. 2000). Because our sampling sites lay in a northwest–southeast direction, reflecting the directions of the main ridges and valleys in the area, the rectangular quadrats also lay in this direction. Within each quadrat, the areas, perimeters, and interpatch distances of uninterrupted patches of high-altitude (> 2000 m), nonforested habitat were quantified from 1:250 000 topographical maps showing forested areas (Department of Energy, Mines & Resources, Canada, 1991), using image-analysis software (Wilcox et al. 1995). The proportion of each quadrat covered by high-altitude, nonforested habitat (% coverage) and the mean area of individual patches of such habitat (mean patch size) were quantified. We measured patch density as the number of patches of uninterrupted high-altitude, nonforested habitat. Mean nearest neighbour distance was measured as the mean edge-to-edge distance between each patch and its nearest neighbouring patch. This metric should increase as suitable patches become more isolated from each other. Fractal dimension is a measure of the irregularity of the shape of each patch. It is based on the perimeter to area ratio of individual patches (McGarigal & Marks 1995) and was averaged across patches within a region to obtain the mean patch fractal dimension. Values of this metric lie between one and two, and larger numbers indicate more complex patch shapes.

We also used three landscape indices specifically designed to measure habitat fragmentation (Jaeger 2000). These indices characterize the penetration of suitable habitat by unsuitable habitat and are calculated from the distribution function of the sizes of remaining patches of suitable habitat by unsuitable habitat and are calculated from the distribution function of the sizes of remaining patches.
suitable habitat. In our case, habitat suitable for movement is represented as all nonforested areas at high elevation. The degree of landscape division ($D$) is the probability that two randomly chosen points in the landscape are not situated in the same uninterrupted patch of suitable habitat. The splitting index ($S$) is the number of patches of suitable habitat one would have if one divided the total region into patches of equal size such that the degree of landscape division ($D$) remained constant. The effective mesh size ($m$) indicates the area of each of the $S$ resulting equal-sized, hypothetical patches. Values of $D$ and $S$ should increase with greater habitat fragmentation, while $m$ decreases. Jaeger (2000) concluded that these measures are more suitable as indicators of habitat fragmentation than many traditional landscape metrics, particularly because of their insensitivity to the omission or addition of very small residual patches.

In some cases, patches at the edges of a quadrat were actually parts of larger patches that mostly lay outside of the quadrat. These presented a problem for calculating metrics that describe the mean size or shape of patches. If the area of the entire large patch was used, then the boundaries of the analysed region would be extended far beyond the sampled region. However, if only the portion of the patch lying within the quadrat was used, then we would be underestimating mean patch areas. In addition, removing all patches that were not entirely within a quadrat from the analysis would have meant removing some large patches that lay mostly but not completely within the quadrats, and that clearly contributed to the connectivity of the sampled region. To balance these considerations, in calculating mean patch size and mean fractal dimension only those patches for which the majority ($>50\%$) of the area lay within the quadrat were counted and only the portion lying within the quadrat was used. The area of high-altitude open habitat thus excluded, as a proportion of the total area studied, was very similar in all three regions ($3\%$ in all cases).

Sample collection and microsatellite genotyping

Tissue samples from adult butterflies were collected in 1995, 1996, and 1999. Samples were either small wing clippings (approximately $0.15 \text{ cm}^2$) or whole adult butterflies. Non-lethal wing sampling was used in the last 2 years, once we had confirmed that DNA could reliably be extracted and amplified from these tissues. All samples were placed individually in glassine envelopes and stored at $-80$ °C upon return from the field. Genomic DNA was isolated using the QIAGEN Tissue Extraction Kit (QIAGEN). For wing clippings, the entire sample was used. For whole butterflies, approximately $25 \text{ mg}$ of abdominal tissue or thoracic tissue was used for males and females, respectively. Each sample was typed at seven microsatellite loci ($Ps50$, $Ps81$, $Ps85$, $Ps76$, $Ps163$, $Ps165$, and $Ps262$) as described previously (Keyghobadi et al. 1999, 2002), using one fluorescently labelled primer per locus (6-FAM, HEX, or TET). Products of polymerase chain reaction (PCR) amplification were electrophoresed and detected on an Applied Biosystems 373A Automated Sequencer and analysed using GENESCAN and GENOTyper software (Applied Biosystems).

In a previous study (Keyghobadi et al. 1999) we used a panel of four microsatellite loci ($Ps50$, $Ps81$, $Ps85$, and $Ps162$) to describe population genetic structure in this species. Here, to add power to the data set, we have added four more loci ($Ps76$, $Ps163$, $Ps165$, and $Ps262$; Keyghobadi et al. 2002) but have removed $Ps162$ because a small number of individuals from the Banff sites produced more than two bands when typed at this locus.

Data analysis

For each locus at each site, genotype frequencies were initially tested for conformity to Hardy–Weinberg expectations with the testing procedure of the program GENEPOP, version 3.1d (Raymond & Rousset 1995). Because we found excess homozygosity and evidence for null alleles at all loci (see Results), for each locus at each site a maximum-likelihood estimate of the frequency of the null allele was calculated, and the frequencies of all other alleles were simultaneously re-estimated, using the estimation-maximization (EM) algorithm (Yasuda & Kimura 1968) as in Keyghobadi et al. (1999) (calculator at http://wwwbiology.ualberta.ca/jbrzusto/nullallele.html). All further calculations of genetic structure and genetic variation were performed using these estimates of allele frequencies, including that of the null allele.

Separately for each locus, we estimated $F_{ST}$ for each region using the analysis of molecular variance (AMOVA) procedure of the program ARLEQUIN (Schneider et al. 2000), which calculates $F_{ST}$ as described by Weir & Cockerham (1984). Because these calculations were based on allele frequencies and not genotype frequencies, the AMOVA does not include a within-individual level of variance and the analysis assumes random union of gametes within populations. We feel that this is a valid assumption, as various aspects of the behaviour and life history of this species suggest that mating within local populations is probably random: females lay eggs individually and can travel tens of metres between successive ovipositions (Fownes 1999); overwintering occurs in the egg stage and must be associated with a high mortality rate, given that a single female is capable of laying hundreds of eggs (Fownes 1999); larvae are not gregarious; and males patrol areas in search of females and newly emerged virgin females tend to be mated quickly upon eclosure. Given these factors, it is unlikely that a disproportionate number of matings among related individuals normally occurs. For each region, $F_{ST}$ estimates from different loci were combined by weighted averaging.
and confidence intervals were obtained by jackknifing over loci (Weir & Cockerham 1984). The program ARLEQUIN also tests the significance of \( F_{ST} \) for each locus by randomly permuting alleles among populations. We combined the probability values from separate loci using the meta-analytic approach described by Sokal & Rohlf (1995; pp. 794–797), to obtain the probability of significance of the overall \( F_{ST} \) (combined over loci) for each region.

Nei’s standard genetic distance (Nei 1972) was calculated between all pairs of sites within regions (calculator at <http://www.biology.ualberta.ca/jbrzusto>). We chose to use Nei’s standard genetic distance (\( D_s \)) because, in comparing various genetic distance measures for use with microsatellite data, Paetkau et al. (1997) found that \( D_s \) performed very well for detecting isolation by distance at a fine geographical and temporal scale, displaying relatively low variance. The significance of correlations between genetic distance and geographical distance was determined using the Mantel test (Mantel 1967) with 10,000 matrix randomizations, as executed by the program ibd (Bohonak 2002); the ibd program also generates confidence intervals for the coefficients of determination by bootstrapping over data points. The ibd program provided estimates of the slope of the relationship between genetic and geographical distances using reduced major axis (RMA) regression, with jackknifing over populations to obtain confidence intervals for those estimates.

Unbiased estimates of expected heterozygosity for each locus within each site were calculated as 
\[
\hat{H}_e = 1 - \sum p_i^2 \frac{2N}{2N - 1},
\]
where \( p_i \) is the estimated frequency of allele \( i \), and \( N \) is the number of individuals sampled (Nei & Roychoudhury 1974). To test for a difference in mean expected heterozygosity among regions we performed an analysis of variance (ANOVA) with region and locus as factors, followed by Tukey’s post hoc test, using JMP-In software; expected heterozygosity values were arcsine-transformed prior to the analysis. We also tested for differences in allelic richness (Petit et al. 1998) among regions using an ANOVA with region and locus as factors, followed by post hoc tests. Allelic richness corrects for differences in sample size among sites by rarefaction (Petit et al. 1998) and was calculated using FSTAT 2.9.3 (Goudet 1995) and not including a null allele; allelic richness was normalized by ln-transformation prior to ANOVA.

Comparison to genetic structure at a smaller spatial scale

To facilitate comparisons with our previous study of population genetic structure conducted at a very fine scale within East Kananaskis (17 sites separated by up to 12 km only and analysed using four microsatellite loci; Keyghobadi et al. 1999), we re-analysed the samples from the fine-scale analysis using our new panel of seven microsatellite loci. As for sites in the larger-scale study area, estimates of allele frequencies, including a null allele, at all loci were obtained using the EM algorithm. These allele frequencies were used to calculate estimates of population genetic structure and IBD for the fine-scale study area, as described above.

Results

Landscape structure

Most of the landscape metrics indicated a considerably higher level of fragmentation of high-altitude, nonforested areas in the foothills landscape (East Kananaskis) as compared to the front ranges landscape (Banff and West Kananaskis; Table 2). The proportion of area covered by high-altitude, open habitat (\% coverage), probably the most important determinant of landscape spatial pattern (Gustafson 1998), was much lower in East Kananaskis than in the other two regions. Percent coverage of nonforested habitat is a broad, and arguably the most comprehensive,

| Table 2 | Landscape metrics for each of the three large-scale study regions. Metrics describe the extent and spatial structure of patches of high-altitude (> 2000 m), nonforested habitat in a quadrat encompassing all sample sites within a given region (Fig. 1). Coverage (%) indicates the proportion of each quadrat covered by high-altitude, nonforested habitat. Mean patch size is the mean area of uninterrupted patches of high-altitude, nonforested habitat. Patch density is the number of such patches in the quadrat. Mean patch fractal dimension is an average measure of patch shape irregularity based on perimeter to area ratios of individual patches in the quadrat. Mean nearest neighbour distance was measured as the mean edge-to-edge distance between each patch and its nearest neighbouring patch. Landscape division, splitting index, and effective mesh size are indices of habitat fragmentation described by Jaeger (2000). |
|---------|----------------------------------|-----------------|------------------|
|         | East Kananaskis                   | Banff           | West Kananaskis  |
| Coverage (%) | 18.7                           | 54.0            | 52.9             |
| Mean patch size (km²) | 2.7                            | 47.1            | 57.4             |
| Patch density (per km²) | 0.03                          | 0.01            | 0.01             |
| Mean patch fractal dimension | 1.24                          | 1.24            | 1.24             |
| Mean nearest-neighbor distance (m) | 423.9                         | 562.5           | 562.5            |
| Landscape division | 0.99                          | 0.92            | 0.95             |
| Splitting index | 208                            | 12              | 21               |
| Effective mesh size | 3                             | 89              | 59               |
measure of the ease with which *Parnassius smintheus* may move through the landscape given their reduced movement though forests. Mean patch size gives an indication of the areas over which movement may occur before forests are encountered, and was also much smaller in East Kananaskis than in the other regions. Also, measures of habitat fragmentation (Jaeger 2000) indicated considerably higher fragmentation for East Kananaskis than for the other two regions. For example, the effective mesh size was an order of magnitude smaller in East Kananaskis than in Banff or West Kananaskis. Patch shape, measured as mean patch fractal dimension, is an indicator of how likely butterflies are to encounter habitat edges, which may affect their rates of emigration and immigration. This variable was remarkably similar among the three regions. Mean nearest-neighbour distance, which indicates the mean distance butterflies must move over forested habitat to travel from one open patch to the next, was identical in Banff and West Kananaskis, and slightly lower in East Kananaskis. Also, patch density was slightly higher in East Kananaskis than in the other regions. Thus, although patches are considerably smaller in East Kananaskis, they are both more numerous and slightly closer together than the large, uninterrupted patches of open habitat in the other two regions.

Overall, the connectivity of the landscape with respect to movement of *P. smintheus* appeared quite similar in the two front ranges regions (Banff and West Kananaskis). Based on the average size of individual patches of high-altitude, nonforested habitat, West Kananaskis region appeared to have slightly higher connectivity for *P. smintheus* than did the Banff region, although the fragmentation indices suggested slightly higher connectivity in Banff.

**Hardy–Weinberg equilibrium, null alleles, and microsatellite variability**

For the 27 sites in the larger-scale study area, a total of 189 tests of conformity to Hardy–Weinberg proportions were performed. Of these, prior to correction for null alleles, 131 showed significant deviations from expected proportions. Even when the total experiment-wise error rate was controlled by Bonferroni adjustment (α = 0.05/189 = 0.0003), 92 tests were significant. Deviations were observed at all seven loci and consistently involved excess homozygosity.

Homozygote excess at these loci in other samples has been attributed to the occurrence of null alleles (Keyghobadi et al. 1999, 2002), and null alleles are the most probable cause of homozygote excess in this study as well. Mean estimates of the null allele frequency for each locus, obtained using the EM algorithm, were 0.13 at *Ps50*, 0.22 at *Ps76*, 0.30 at *Ps81*, 0.14 at *Ps85*, 0.06 at *Ps163*, 0.34 at *Ps165*, and 0.03 at *Ps262*.

Overall, we detected high microsatellite variability in these populations. Averaged over loci, the mean number of alleles (including a null) within samples ranged from 6.7 to 13.0, and the mean expected heterozygosity ranged from 0.65 to 0.79 (Table 1).

**Genetic structure and $F_{ST}$ estimates**

Genetic differentiation, $F_{ST}$, varied among the three regions and was highest in East Kananaskis, the region displaying the lowest connectivity (Table 3). An intermediate level of differentiation was observed in Banff and the lowest level of differentiation was in West Kananaskis. Estimates of $F_{ST}$ were significantly different between East Kananaskis and both Banff and West Kananaskis, but not between Banff and West Kananaskis.

To determine if a single locus might be having an undue influence on estimates of genetic differentiation, we examined the estimates of $F_{ST}$ for each region (i) by omitting one locus at a time, and (ii) for each locus separately. When omitting one locus at a time (Table 4a), estimated differentiation was consistently highest in East Kananaskis, intermediate in Banff, and lowest in West Kananaskis. When considering each locus separately (Table 4b), six of seven loci indicated that the greatest differentiation occurs in East Kananaskis and five of seven loci indicated that the lowest differentiation occurs in West Kananaskis. Thus, our results were not being driven by the behaviour of any single locus, but appeared quite consistent among loci.

**Isolation by distance**

We observed a significant, positive correlation between genetic distance and geographical distance in West Kananaskis (Table 5; Fig. 2a) and Banff (Fig. 2b), but not in East Kananaskis (Fig. 2c). Within the fine-scale study area, which

### Table 3 Estimates of genetic differentiation ($F_{ST}$) and genetic diversity within each of the large-scale study regions. $H_e$ is the mean expected heterozygosity (Nei & Roychoudhury 1974) averaged over loci and sites. Allelic richness is calculated as in Petit et al. (1998), and is averaged over loci and sites.

<table>
<thead>
<tr>
<th>Region</th>
<th>$F_{ST}$ ($P$)</th>
<th>$F_{ST}$ 95% CI</th>
<th>$H_e$ ($±$ SE)</th>
<th>Allelic richness ($±$ SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Kananaskis (Large scale)</td>
<td>0.038 ($&lt; 0.001$)</td>
<td>0.021–0.056</td>
<td>0.72 ($± 0.02$)</td>
<td>6.20 ($± 0.37$)</td>
</tr>
<tr>
<td>Banff</td>
<td>0.011 ($&lt; 0.001$)</td>
<td>0.003–0.020</td>
<td>0.78 ($± 0.01$)</td>
<td>5.70 ($± 0.20$)</td>
</tr>
<tr>
<td>West Kananaskis</td>
<td>0.0016 ($0.13$)</td>
<td>$−0.009–0.012$</td>
<td>0.76 ($± 0.02$)</td>
<td>6.69 ($± 0.37$)</td>
</tr>
</tbody>
</table>
is located in the East Kananaskis region, we observed a significant positive correlation between $D_s$ and geographical distance for all sample sites (Supplementary material Fig. S1a).

The slope of the IBD relationship, as determined by RMA regression, differed significantly among the three regions (Table 5) and was lowest in East Kananaskis (not significantly different from zero), intermediate in West Kananaskis, and highest in Banff. Also, among the fine-scale study sites in East Kananaskis, the IBD slope was significantly greater than zero.

Subsampling in isolation-by-distance analyses

Sample size may affect the observed strength of an IBD trend (Peterson & Denno 1998). Sample sizes among the three regions were very similar ($n = 8, 9, \text{ and } 10$) so those comparisons should be minimally affected by this factor. However, in comparing IBD between the large and small spatial scale in East Kananaskis ($n = 8$ and $n = 17$, respectively) there is a greater discrepancy in sample size. To control for any effect of sample size in this comparison, we subsampled eight sites from the small scale and performed the IBD analyses on this subsample. The eight sites (E, F, G1, J, M, Q, R, Z in Fig. 1 of Keyghobadi et al. 1999) were chosen to be evenly spaced and to have a minimum of 24 individuals sampled per site. The Mantel test revealed a significant pattern of IBD among these eight sites (Table 5; Supplementary material Fig. S1b), and the IBD slope, determined by RMA regression, was significantly higher than the slope observed for the large-scale analysis in East Kananaskis.

Differences in within-population genetic variation among regions

An ANOVA indicated significant differences in expected heterozygosity (arcsine transformed) among loci (d.f. = 6, $F = 86.50, P < 0.0001$) and among the three regions (d.f. = 2, $F = 5.76, P = 0.0038$). Tukey’s post hoc test (with $\alpha = 0.05$) revealed that, in the among-region comparison, expected heterozygosity in East Kananaskis was significantly lower than in Banff and West Kananaskis, while the latter two did not differ from each other. Mean values of the untransformed variable are given in Table 3. An ANOVA on ln-transformed values of allelic richness also indicated significant differences among loci (d.f. = 6, $F = 83.51, P < 0.0001$) and among the three regions (d.f. = 2, $F = 4.76, P = 0.01$). Post hoc tests ($\alpha = 0.05$) revealed that allelic richness in West Kananaskis was significantly higher than in Banff and East Kananaskis, while the latter two did not differ from each other. Mean values of the untransformed variable are given in Table 3.

Thus, when considering both expected heterozygosity and allelic richness jointly, there appeared to be greater genetic diversity in West Kananaskis as compared to East Kananaskis; however, the position of Banff was unresolved.

### Table 4

<table>
<thead>
<tr>
<th>Location</th>
<th>$F_{ST}$ for all loci</th>
<th>$F_{ST}$ for single locus</th>
<th>$F_{ST}$ for single locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Kananaskis</td>
<td>0.034</td>
<td>0.013</td>
<td>0.003</td>
</tr>
<tr>
<td>Banff</td>
<td>0.041</td>
<td>0.013</td>
<td>-0.0005</td>
</tr>
<tr>
<td>West Kananaskis</td>
<td>0.041</td>
<td>0.011</td>
<td>0.001</td>
</tr>
<tr>
<td>East Kananaskis</td>
<td>0.035</td>
<td>0.011</td>
<td>0.005</td>
</tr>
<tr>
<td>Banff</td>
<td>0.038</td>
<td>0.011</td>
<td>0.0003</td>
</tr>
<tr>
<td>West Kananaskis</td>
<td>0.042</td>
<td>0.009</td>
<td>0.003</td>
</tr>
<tr>
<td>East Kananaskis</td>
<td>0.036</td>
<td>0.011</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

### Table 5

<table>
<thead>
<tr>
<th>Region</th>
<th>Mantel test</th>
<th>Reduced major axis regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r$ ($P$)</td>
<td>$r^2$ (95% CI)</td>
</tr>
<tr>
<td>Large Scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Kananaskis</td>
<td>-0.28 (0.87)</td>
<td>0.08 (0.001−0.34)</td>
</tr>
<tr>
<td>Banff</td>
<td>0.50 (0.03)</td>
<td>0.25 (0.03−0.51)</td>
</tr>
<tr>
<td>West Kananaskis</td>
<td>0.59 (&lt; 0.001)</td>
<td>0.35 (0.10−0.60)</td>
</tr>
<tr>
<td>Small Scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Kananaskis</td>
<td>0.54 (&lt; 0.001)</td>
<td>0.29 (0.16−0.43)</td>
</tr>
<tr>
<td>East Kananaskis (subsample of 8 sites)</td>
<td>0.87 (&lt; 0.001)</td>
<td>0.76 (0.54−0.88)</td>
</tr>
</tbody>
</table>
Discussion

Quantifiable differences in the structure of high-altitude, nonforested habitats among three geographical regions were associated with differences in the degree of population differentiation and inferred levels of gene flow in the alpine meadow-dwelling butterfly, *Parnassius smintheus*. These results were consistent with the effects of intervening forests in restricting movement of individuals (Roland et al. 2000) and in promoting population genetic differentiation at a much smaller spatial scale (Keyghobadi et al. 1999). Genetic differentiation among populations, measured as \( F_{ST} \), was highest in the most fragmented region (East Kananaskis), and the strength of IBD was lowest in this region, where no significant correlation was observed. Comparison to the pattern of IBD observed at a smaller spatial scale indicates that the breakdown of IBD in the most fragmented region is due to reduced levels of gene flow. Also, within-population heterozygosity was significantly lower in the most fragmented region than in the other two regions, while allelic richness was significantly lower in the most fragmented region than in one of the other two regions. Low within-population variation may be associated with population isolation, if genetic drift is not negligible. Overall, our results suggest a positive association between levels of gene flow and landscape connectivity that is observable at different spatial scales and that reflects the movement behaviour of individuals.

Landscape ecologists have focused attention on the importance of landscape spatial pattern in determining various ecological processes (Forman & Godron 1986). Over longer timescales landscape patterns will also influence evolutionary processes, and our results indicate that landscape connectivity influences rates and patterns of gene flow and thereby, levels and spatial distribution of genetic diversity.

Our results are consistent with the majority of studies that have found that increased levels of gene flow are associated with greater continuity of habitats in the landscape (e.g. Caccone & Sbordoni 1987; King 1987; Britten et al. 1995; Shoemaker & Jaenike 1997; VanDongen et al. 1998; Knutsen et al. 2000) and with improved dispersal ability (Peterson & Denno 1998; Bohonak 1999). Other studies have also shown positive associations between movement rates and habitat continuity (Mennechez et al. 2003). Our study is among the few however, that establish the three-way link between landscape, movement, and population genetic structure (Michels et al. 2001; Vos et al. 2001). Although a positive association between gene flow and...
continuity of habitat is common, exceptions to this general pattern do exist (Mossman & Waser 2001). For example, Young et al. (1993) found higher levels of gene flow in fragmented populations of sugar maple, Acer saccharum, than in continuous populations. They suggested that this may have been due to higher wind speed and therefore, greater wind-assisted dispersal of pollen in the open, fragmented landscape. Clearly, ecological factors may sometimes affect gene flow in complex and unanticipated ways.

Our analysis involved comparison of one region of low landscape connectivity with two regions of higher landscape connectivity. The differences we observed in genetic structure between the low- and high-connectivity landscapes are striking. However, the difficulty of replicating across several landscapes, a hurdle shared by all landscape-level analyses (Tischendorf & Fahrig 2000), means that the number of different samples (i.e. landscapes) available to assess a statistical correlation between population genetic parameters and landscape structure is necessarily limited. Additional replicates of different landscapes would be required for a statistical treatment of the correlation between population genetic structure and landscape connectivity.

Lack of isolation by distance in the foothills landscape is due to low levels of gene flow

Lack of a significant pattern of IBD over a particular geographical scale for neutral markers, as observed in the highly fragmented foothills region (East Kananaskis; Fig. 2c), may occur for one of three reasons. First, gene flow may be very low over the distances sampled such that populations are essentially isolated and allele frequencies are determined by drift. Second, gene flow may be very high over the entire distance range such that the sampled region functions effectively as one large population. Third, populations may not be at equilibrium with respect to the forces of gene flow and drift following some historical perturbation. Here, lack of IBD in the foothills region appears to be due to low levels of gene flow over the distances sampled.

Three lines of evidence suggest that the breakdown of IBD in the foothills landscape is due to very low levels of gene flow. First, this interpretation is consistent with the sedentary nature of this species and with the effects of forests in limiting movement (Roland et al. 2000). Second, genetic differentiation ($F_{ST}$) was highest in this region. Finally, and most importantly, we have observed a significant pattern of IBD in the small-scale study area (Supplementary material Fig. S1), which was located within the foothills region (Fig. 1). In general, rates of gene flow should increase as the spatial scale of analysis is reduced. If lack of a significant correlation between genetic distance and geographical distance is due to high gene flow, then decreasing the spatial scale of analysis, as we have done here, should not reveal such a correlation. However, if lack of IBD is due to severely restricted gene flow, then at some smaller spatial scale a pattern of IBD should eventually become evident. Thus, our results suggest that populations in the foothills landscape are quite isolated from each other at the larger spatial scale, but are connected by gene flow at a smaller scale. Peterson & Denno (1998) have suggested that in the many sedentary phytophagous insects for which IBD is not observed, analysis at a smaller spatial scale should reveal significant IBD. Our results substantiate this prediction.

Although we cannot entirely rule out non-equilibrium conditions in the foothills region, this appears an unlikely explanation for the lack of a pattern of IBD at the large scale. Again, the occurrence of IBD at a smaller spatial scale within this region argues against the existence of non-equilibrium conditions there. Our ability to differentiate among competing hypotheses explaining the observed pattern of IBD at a large scale, using the results from a smaller scale, underlines the value of conducting analyses at more than one spatial scale.

Null alleles

We observed significant homozygote excess at all seven loci, which we can attribute to the existence of null alleles. We have previously provided independent evidence for the existence of null alleles at these loci, largely in the form of null homozygous individuals, as well as evidence of instability in the sequences flanking some of these microsatellite repeats (Keyghobadi et al. 1999, 2002). Homozygote excess due to assortative mating of related individuals is unlikely based on various aspects of the life history and mating behaviour of this species, as outlined previously (Keyghobadi et al. 2002). In our small-scale study area, excess homozygosity due to a Wahlund effect, or the erroneous combination of separate populations in a single sample, has previously been ruled out based on the very fine spatial scale of sampling (Keyghobadi et al. 2002). At the larger scale of this study, the possible occurrence of a Wahlund effect cannot be dismissed as easily. However, excess homozygosity within sites, measured as the difference between observed and expected homozygosity, averaged over loci, did not differ between the large-scale study sites (excluding sites E, Q, and Z) and the small-scale study sites (Mann–Whitney U-test: $Z = -0.66$, $P = 0.51$). This suggests that no additional factor other than null alleles is necessary to explain the excess homozygosity observed in the large-scale study sites.

Excess homozygosity and null alleles are not unique to this species, but are commonly observed in butterflies (e.g. Meglécz & Solignac 1998; Williams et al. 2003). Indeed, our estimates of null allele frequencies were comparable to those reported by Harper et al. (2003) for the Adonis blue
butterfly. They observed mean null allele frequencies ranging from 0.03 to 0.2 at five loci, while we observed a mean frequency from 0.03 to 0.22 at five loci and slightly higher frequencies of 0.3 and 0.34 at two loci.

Low estimated values of $F_{ST}$ reflect high variability of microsatellites

In apparent contradiction to the low movement rates of this species (Roland et al. 2000), we obtained low estimates of $F_{ST}$ within all three regions. The low estimated values of $F_{ST}$ observed here are more parsimoniously explained by the high variability and high mutation rates of the molecular markers used. As a result of high variability, estimates of $F_{ST}$ based on microsatellite data tend to be deflated and underestimate the degree of population differentiation (Slatkin 1995; Balloux & Lugon-Moulin 2002). Thus, although the values of $F_{ST}$ that we have measured are useful for comparing levels of population differentiation among the three regions, they cannot reliably be used to obtain absolute estimates of gene flow ($Nm$) using Wright’s (1931) island model. Furthermore, measures of differentiation designed for microsatellite loci and assuming a stepwise-mutation model (e.g. Slatkin 1995) are not appropriate for our data because at several of the $P. smintheus$ microsatellite loci, size variation occurs in the flanking regions (Keyghobadi, unpublished data). However, because we have used the same suite of microsatellite loci in all three regions, comparison of relative $F_{ST}$ values among the regions is still valid and highly informative.

Defining habitat of Parnassius smintheus

Our analysis of landscape structure focused on the relative abundance and distribution of forested and nonforest habitats at high elevations. We thus focused on the effects of forests as barriers to movement in this species and did not distinguish among different types of open, high-elevation habitats. Similarly, Roland et al. (2000) compared movement of the butterfly between meadows and forest habitats, and did not distinguish among different meadows in either their occupancy status or the quality of the habitat for $P. smintheus$. This simplification was necessary in our study because defining habitat for $P. smintheus$ is complex, as many areas containing the host plant do not support populations of the butterfly (Fownes 1999). Thus, determination of habitat and nonhabitat for this species is not readily performed using aerial photographs or biophysical maps, and requires considerable field surveying. However, having established effects of landscape connectivity on population genetic structure and gene flow using this broad approach, it will now be informative to examine in more detail the separate effects of $P. smintheus$ habitat and other high-elevation, nonforested landscape elements.
Britten HB, Brussard PF, Murphy DD, Ehrlich PR (1995) A test for
isolation-by-distance in Central Rocky Mountain and Great
Basin populations of Edith’s checkerspot butterfly (Euphydryas
Caccione A, Sbrondoni V (1987) Molecular evolutionary divergence
Evolution, 41, 1198–1214.
variability on the population genetic structure of the alpine butterfly
Parnassius smintheus (Lepidoptera). Hereditas, 128, 179–180.
smintheus (Lepidoptera). Hereditas, 128, 179–180.
dynamics of the bog fritillary butterfly: comparison of demo-
graphic parameters and dispersal between a continuous and a
highly fragmented landscape. Landscape Ecology, 18, 279–291.
Michels E, Cottenie K, Neys L et al. (2001) Geographical and
genetic distances among zooplankton populations in a set of
interconnected ponds: a plea for using GIS modelling of the
effective geographical distance. Molecular Ecology, 10, 1929–
1938.
Mossman CA, Waser PM (2001) Effects of habitat fragmentation
on population genetic structure in the white-footed mouse
Nei M (1972) Genetic distance between populations. American
Naturalist, 106, 283–292.
Nei M, Roychoudhury AK (1974) Sampling variances of hetero-
Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997)
An empirical evaluation of genetic distance statistics using
microsatellite data from bear (Ursidae) populations. Genetics,
147, 1943–1957.
breadth on patterns of genetic isolation by distance in phyto-
genetics software for exact and ecumenism. Journal of
Heredity, 86, 248–249.
genetics software for exact and ecumenism. Journal of
Heredity, 86, 248–249.
Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN Version 2.000: a
software for population genetics data analysis. Genetics and Bio-
metry Laboratory, Department of Anthropology, University of
Geneva, Switzerland.

© 2005 Blackwell Publishing Ltd, Molecular Ecology, 14, 1897–1909

This work was part of Nusha Keyghobadi’s doctoral thesis. She is currently an assistant professor at Okanagan University College where her research focuses broadly on the population and molecular ecology of insects. Research in Jens Roland’s laboratory focuses on insect population dynamics and landscape ecology. Curtis Strobeck is a population geneticist, with a particular interest in wildlife conservation genetics.