Adaptive divergence within and between ecotypes of the terrestrial garter snake, *Thamnophis elegans*, assessed with F_{ST}-Q_{ST} comparisons

M. K. MANIER, C. M. SEYLER & S. J. ARNOLD

Department of Zoology, Oregon State University, Corvallis, OR, USA

Abstract

Populations of the terrestrial garter snake (*Thamnophis elegans*) around Eagle Lake in California exhibit dramatic ecotypic differentiation in life history, colouration and morphology across distances as small as a few kilometres. We assayed the role of selection in ecotypic differentiation in *T. elegans* using F_{ST}-Q_{ST} analysis and identified selective agents using direct and indirect observations. We extended the conventional implementation of the F_{ST}-Q_{ST} approach by using three-level analyses of genetic and phenotypic variance to assess the role of selection in differentiating populations both within and between ecotypes. These results suggest that selection has driven differentiation between as well as within ecotypes, and in the presence of moderate to high gene flow. Our findings are discussed in the context of previous correlational selection analyses which revealed stabilizing and correlational selection for some of the traits examined.

Introduction

Adaptive divergence is often a compromise between the opposing forces of selection and gene flow. The tension between these forces is especially strong in the case of ecotypic differentiation, in which populations in close proximity may become locally adapted to different selective pressures in the presence of persistent gene flow. An ongoing challenge is to establish the degree and scale of ecotypic differentiation, empirical questions that must be tackled on a taxon by taxon basis. We pursue an analysis of ecotypic differentiation using F_{ST}-Q_{ST} comparison and test a hypothesis of divergence within and between ecotypes. We further identify candidate selective agents using direct and indirect observations of predation events.

The statistical comparison of population differentiation at quantitative traits (Q_{ST}) and neutral molecular markers (F_{ST}) provides a powerful test for the role of selection in phenotypic divergence (Lande, 1992; Spitze, 1993; Merilä & Crnokrak, 2001; McKay & Latta, 2002). Departures from neutral expectations can be used to distinguish

Correspondence: Mollie K. Manier, Hopkins Marine Station, Oceanview Blvd., Pacific Grove, CA 93950, USA.

Tel.: +1 831 655 6210; fax: + 1 831 655 6215;

e-mail: manier@stanford.edu

between a history dominated by diversifying ($Q_{ST} > F_{ST}$) as opposed to stabilizing selection ($Q_{ST} < F_{ST}$). In applications in a wide range of taxa, *a priori* expectations of the F_{ST}-Q_{ST} relationship have been used both to detect and assay modes of selection acting on quantitative traits (Palo *et al.*, 2003; Wong *et al.*, 2003; Cano *et al.*, 2004) and to test specific hypotheses about local selection (Baker, 1992; Waldmann & Andersson, 1998; Gomez-Mestre & Tejedo, 2004). Here, we test the power of the F_{ST}-Q_{ST} approach by asking whether it can resolve the role of selection in promoting differentiation among populations within ecotypes, as well as between ecotypes. The success of this test may indicate whether the F_{ST}-Q_{ST} approach will generally be able to assess the role of selection in truly subtle instances of phenotypic differentiation.

In our study system, populations of the terrestrial garter snake (*Thamnophis elegans*) in the Eagle Lake basin of Lassen Co., California show ecotypic differentiation on a scale of several kilometres and in the presence of gene flow high enough to override drift (up to 1.8 effective migrants per generation) at neutral microsatellite markers (Bronikowski & Arnold, 1999; Manier & Arnold, 2005; Sparkman *et al.*, 2007). Previous research has documented ecotypic differences in reproduction, growth and survival between populations along the rocky

shoreline of Eagle Lake and those inhabiting the densely vegetated surrounding meadows (Bronikowski & Arnold, 1999). The life history differences constitute a syndrome that may be driven by higher predation rates at the lakeshore. These populations grow faster, reproduce at an earlier age and have larger litters but suffer higher adult mortality than meadow populations. Common garden experiments have demonstrated a genetic basis for the difference in growth rate (Bronikowski, 2000).

Here, we focus on differences in colouration that appear to increase cryptic in rocky lakeshore and grassy meadow environments. To visual predators, the muted colours of lakeshore snakes (dull yellow or tan stripes on a grey background colour) tend to match the rocky substrate of the lakeshore, whereas the meadow snake colour pattern (yellow or orange stripes on a black background) closely resembles dead rushes that litter the shallow meadow substrates (Fig. 1). The difference in colouration between lakeshore and meadow ecotypes may be a result of differential selection for cryptic (Kephart, 1981).

We also examined six scale counts for adaptive differences between *T. elegans* ecotypes. Vertebral num-

ber (measured using ventral and subcaudal scale counts) can vary between different habitats as a function of a snake's ability to utilize substrate irregularities or 'push-points' for propulsion during locomotion (Jayne, 1988; Gasc *et al.*, 1989; Kelley *et al.*, 1997). In thickly vegetated habitats with a higher density of push-points, *T. elegans* and other snake populations have fewer vertebrae, whereas rocky habitats that have fewer push-points support populations with more vertebrae (Klauber, 1941; Kelley *et al.*, 1997; Arnold & Phillips, 1999). Because lakeshore habitats provide fewer push-points than meadow habitats, we expected to see more body and tail vertebrae in lakeshore than in meadow *T. elegans*. The other scalation traits are likely to reflect a snake's ability to ingest large prey, with high values for these traits promoting extended cranial kinesis (infralabial, supralabial and postocular scale counts) and midsection elasticity (midbody scale count). Because diet studies indicate that lakeshore snakes generally eat larger prey items (fish) than meadow snakes (anuran larvae, leeches; Kephart & Arnold, 1982; Kephart, 1982), we expect selection for ability to swallow larger prey in lakeshore populations and hence higher scale counts. Both scale counts and colouration have been shown to be under selection in these and other populations of snakes (Arnold, 1988; Arnold & Bennett, 1988; Brodie, 1992; King, 1993a; Lindell *et al.*, 1993), making these traits good candidates for our study.

We also ask whether selection has played a role in promoting the more subtle differences in colouration and morphology that appear to characterize populations within each of the two ecotypes. Meadow populations in our system, for example, appear to differ subtly in colouration, such that some populations have a higher incidence of yellow as opposed to orange dorsal stripes. Likewise, meadow populations occupy habitats that differ slightly in water depth, seasonal patterns of drying, phenology and composition of vegetation, as well as in prey availability (Kephart, 1982; Manier & Arnold, 2006). Similar small differences in snake colouration and habitat characteristics can be seen from one lakeshore population to the next. One of our goals is to assess the statistical reality of these impressions of morphological differentiation within ecotypes and to determine whether they might represent responses to selection.

We used estimates of neutral divergence at microsatellite loci to determine whether colouration and scalation traits have experienced diversifying selection, especially between the two ecotypes. We can reject neutrality as an explanation of population differentiation in quantitative traits if $Q_{ST} \neq F_{ST}$. Thus, $Q_{ST} > F_{ST}$ suggests diversifying selection, whereas $Q_{ST} < F_{ST}$ suggests stabilizing selection towards the same optimum in different populations (Lande, 1992; Spitze, 1993). Because each ecotype is represented by multiple populations, we can also determine the importance of differentiation within and between ecotypes to F_{ST} and Q_{ST}. Based on our
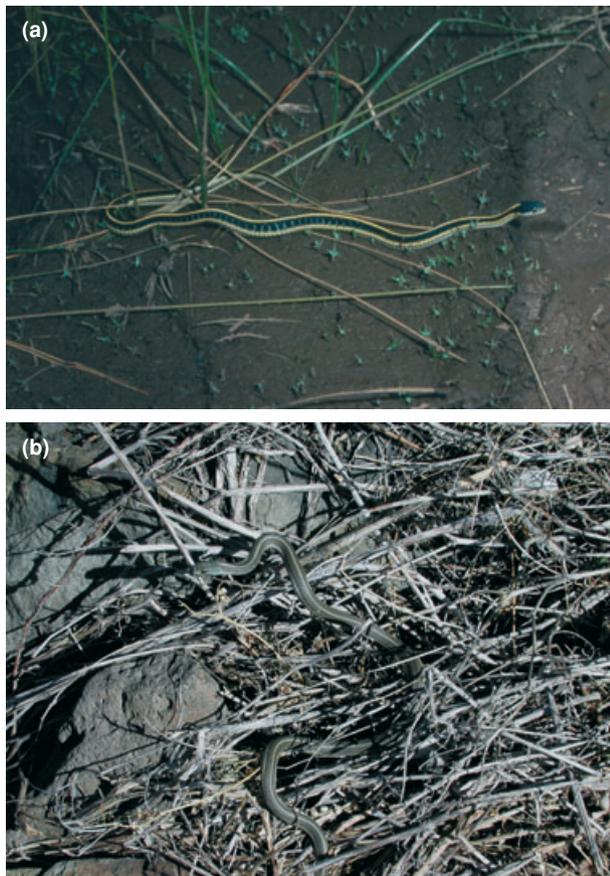
Fig. 1 Examples of *Thamnophis elegans* ecotypes. (a) Meadow ecotype. (b) Lakeshore ecotype.

hypothesis of ecotypic differentiation, we expect Q_{ST} estimates to far exceed F_{ST} in ventral scale counts and colouration. We expect that this population differentiation is largely a consequence of strong divergence between ecotypes ($Q_{CT} \gg F_{CT}$), with some contribution because of slight divergence among populations within ecotypes ($Q_{SC} > F_{SC}$). We used parallel, three-level analysis of variance for microsatellite alleles and for quantitative traits to test these hypotheses. We also present evidence from direct observations of predation and analysis of culmen impressions on snakes that implicate avian predators as the selective agents responsible for this adaptive differentiation.

Materials and methods

Study sites

Phenotypic and genotypic data were collected from six populations (Table 1) in the vicinity of the south-east corner of Eagle Lake in Lassen Co., California (Fig. 1). Eagle Lake is California's second-largest natural lake and supports populations of garter snakes at intervals along its extensive shoreline. These lakeshore populations are separated by stretches of shoreline a few to many kilometres long that are uninhabited by *T. elegans*. Distances between populations within an ecotype also ranged from 1.3 to 19.9 km, whereas between-ecotype distances ranged from 3.8 to 19.5 km. Euclidean distances were used among all sites, except the two lakeshore populations, for which shoreline distance was used. Snakes were found in the open and by lifting cover objects and were collected by hand.

Measurement of quantitative traits

Colouration traits were scored on 1268 live *T. elegans*. Population samples ranged from 50 to 745 individuals. We scored colouration traits by matching the dorsal and lateral stripes and the background area between the stripes to colour standards under diffuse, natural light. One person (SJA) did all colour scoring in the field, and snake colours were scored at mid-body under uniform lighting at mid-morning. Predation is most likely at this time of day, when snakes emerge from nocturnal refugia to sun themselves.

Dorsal and lateral stripes were matched to 10 different Pantone® colour swatches, ranging from yellow to orange to tan to pink (127, 128, 134, 135, 136, 137, 141, 148, 155, 162; out of 36 tested; available at http://www.pantone.com). For statistical analyses, we translated the Pantone colour codes into two different three-element colour schemes, HSL (hue, saturation, lightness) and RGB (red, green, blue), which were then evaluated for their ability to detect selection. The RGB system should be more biologically relevant, as it approximates the types of vertebrate wavelength receptors (Jacobs, 1981; Chen & Goldsmith, 1986; Jane & Bowmaker, 1988). HSL, on the other hand, is derived from Munsell codes (Munsell Colour Company, 1976), which classifies colour into groups based on human perception. For background, we used a 5-point Kodak grey scale. Sampling bias of certain colour patterns is unlikely, because snakes were not always identified against the dominant background, as when found under cover objects. Furthermore, any sampling bias would most likely result in an underestimate of population divergence, because collectors would preferentially capture snakes with colouration-background mismatch. Variable abbreviations represent all combinations of stripe and colour component; for example, saturation of the dorsal stripe was DORSAT (e.g. Table 2). Background colour (BKGRD) was quantified as various degrees of darkness, with higher numbers corresponding to a darker background.

The following six scale counts were made on 2251 preserved and live specimens: number of ventral scales on the body (VENT); number of subcaudal scales (SUB); total number of infralabial (ILAB); supralabial (SLAB); postocular (POST) scales on the left and right sides; and number of dorsal scale rows at midbody (MID), as described by Arnold & Phillips (1999). VENT and SUB correspond, respectively, to the numbers of body and tail vertebrae (Alexander & Gans, 1966; Voris, 1975). Missing values comprised < 4% of the dataset, three-fourths of which were attributed to missing tail tips (SUB). Sex was determined by eversion of hemipenes. Scale counts do not change during the ontogeny of an individual. Experimental studies of the developmental effect of

Table 1 Names, abbreviations, ecotype and sample sizes of study populations for scalation and colour variables (for males and females) and microsatellite markers.

Population name	Abbreviation	Ecotype	Scalation		Colour		Microsat
			M	F	M	F	
Gallatin Shoreline*	GAL	lakeshore	387	406	363	382	56
Mahogany Lake	MAH	meadow	56	100	48	65	91
McCoy Flat Res.	MCY	meadow	72	116	23	27	16
Nameless Meadow*	NML	meadow	41	109	57	55	29
Papoose Meadows	PAP	meadow	62	111	36	56	140
Pikes Point	PIK	lakeshore	346	445	77	79	48

*Study site name is informal only, not an official geographic place name.

Table 2 Heritability estimates for colouration traits.

	BKGRD	DOR HUE	DOR SAT	DOR LT	LAT HUE	LAT SAT	LAT LT	DOR RED	DOR GRN	DOR BLU	LAT RED	LAT GRN	LAT BLU
Females													
h^2	0.80	0.73	0.21	0.22	0.31	0.17	0.25	0.32	0.35	0.16	0.14	0.26	0.24
SE	0.18	0.29	0.10	0.14	0.18	0.18	0.19	0.13	0.14	0.13	0.18	0.24	0.17
P	0.001	0.001	0.017	0.057	0.052	0.169	0.101	0.002	0.008	0.117	0.223	0.143	0.087
N	32	32	32	32	30	30	30	32	32	32	30	30	30
Males													
h^2	0.48	0.58	0.54	0.37	0.62	0.04	0.28	0.63	0.60	0.35	0.01	0.55	0.25
SE	0.15	0.11	0.14	0.16	0.08	0.16	0.19	0.14	0.15	0.17	0.16	0.12	0.18
P	< 0.001	< 0.001	< 0.001	0.007	< 0.001	0.421	0.056	< 0.001	< 0.001	0.014	0.507	< 0.001	0.082
N	35	35	35	35	35	35	35	35	35	35	35	35	35
Average													
h^2	0.64	0.65	0.38	0.30	0.46	0.10	0.27	0.48	0.48	0.26	0.07	0.40	0.25
SE	0.17	0.20	0.12	0.15	0.13	0.17	0.19	0.14	0.15	0.15	0.17	0.18	0.18

Significance levels (*P*) and standard errors (SE) were assessed by boot-strapping over the family structure.

N, number of families.

Inheritance of quantitative traits

temperature on scalation traits in *T. elegans* (Arnold & Peterson, 2002) indicate that environmental differences in temperature are unlikely to account for population and ecotypic differences in scalation.

The heritabilities of colouration traits were assessed using a sample of 325 individuals representing 35 sibships from the Pikes and Wildcat populations. Wildcat is a lakeshore site (see Fig. 2) excluded from the microsatellite analysis because of low sample size. All colouration scores were made on neonates, no more than 1 month after birth. Colouration traits are fully expressed at birth, and no obvious ontogenetic trends were observed when indi-

Fig. 2 Map showing locations of meadow (MCY, PAP, NML, MAH) and lakeshore (PIK, GAL) sites. An additional lakeshore site, WDC, was used to generate heritabilities for scale counts.

viduals were reared to maturity in the laboratory. Heritabilities (h^2), genetic covariances, and environmental covariances were estimated by treating the sibships as unrelated sets of fullsibs using software (H2BOOT; Phillips, 1998) available at a website maintained by P.C. Phillips (http://www.uoregon.edu/~pphil/software.html). Fullsib ANOVA (Falconer & MacKay, 1996) rather than mother–offspring regression was used to estimate heritabilities, because colouration scores for mothers were missing for over a third of the sibships. Multiple paternity, known to occur in garter snakes (Garner & Larsen, 2005), probably had a negligible effect on our estimates (Arnold & Phillips, 1999). Even in the presence of multiple paternity, most littermates are, on average, nearly full-sibs (Schwartz *et al.*, 1989). Estimates of heritability can be inflated by maternal effects (Falconer & MacKay, 1996), but we have shown that scalation traits are buffered against the most conspicuous type of maternal influence, viz., maternal temperature during development (Arnold & Peterson, 2002). Nevertheless, these fullsib estimates should be viewed as upper bounds on narrow sense heritability, because they may be inflated to an unknown degree by dominance variance and a common family environment during gestation (Arnold, 1981; Falconer & MacKay, 1996). Standard errors of heritability were estimated and tests of the hypothesis that $h^2 \geq 0$ were conducted using 1000 bootstrap samples in H2BOOT. Heritabilities were estimated separately for males (184 individuals in 35 sibships) and females (141 individuals in 32 sibships), and the average of the separate estimates was used in the Q_{ST} analyses.

Estimates of heritability for the scale count traits were taken from Arnold & Phillips (1999). Those estimates are based on mother–offspring regressions using a sample of 102 mothers and 911 offspring from Pikes Point and the population at Wildcat Point. We used the
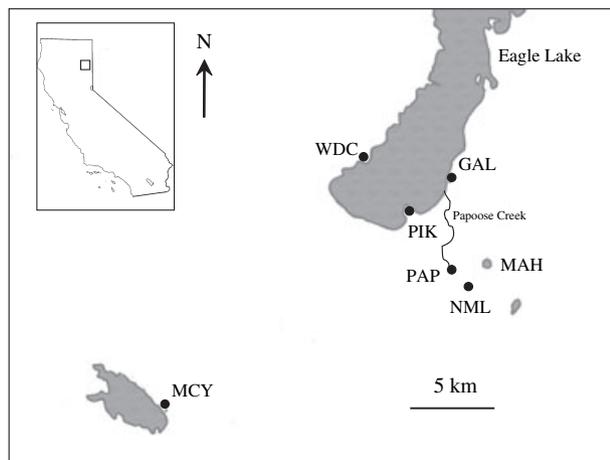
average of these male and female heritabilities, shown in Table 6 of Arnold & Phillips (1999), estimated from the genetic and phenotypic variances from this population given in Tables 3 and 4 of Arnold & Phillips (1999).

Microsatellite analysis of population structure

Nine microsatellite loci were scored for a total 380 individuals representing the six study populations. Population samples ranged from 16 to 140 individuals (Table 1). These samples generally do not overlap with those used for phenotypic traits. A tail tip or piece of ventral scale was clipped and stored in Drierite®, an anhydrous calcium sulphate desiccant. Whole genomic DNA was extracted using sodium dodecyl sulphate-proteinase K digestion followed by a standard phenol–chloroform extraction, NaCl purification and isopropanol precipitation. DNA was PCR amplified in a 12.5 μL reaction with 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 0.1% Triton X-100, 0.2 mM each of dNTPs, 1.5 mM MgCl$_2$, 0.48 μM forward (labelled with fluorescent ABI dye) and reverse primer, and 0.3 U *Taq* DNA polymerase. PCR profiles consisted of 94 °C for 2 min followed by 36 cycles of 94 °C for 30 s, appropriate annealing temperature for 30 s and 72 °C for 30 s, ending with 72 °C for 2 min. PCR products were separated using an ABI 3100 capillary electrophoresis genetic analyser and data were visualized using GENOTYPER 3.7 (ABI Prism).

Population genetic analysis of the microsatellite data is described in Manier & Arnold (2005). All microsatellite data were combined for both sexes. These analyses included exact tests for departure from Hardy–Weinberg equilibrium (Guo & Thompson, 1992; Markov chain parameters: 5000 dememorizations; 500 000 steps per chain) calculated in ARLEQUIN v. 2.000 (Schneider *et al.*, 2000) and tests for linkage disequilibrium (Slatkin & Excoffier, 1996; Markov chain parameters: 5000 dememorizations, 1000 batches, 5000 iterations per batch), performed in GENEPOP (Raymond & Rousset, 1995). Significance levels were adjusted for multiple comparisons using two methods, the sequential Bonferroni correction (Rice, 1989) and false discovery rate control (FDR; Benjamini & Hochberg, 1995), both of which yielded the same results. Only Mahogany Lake was out of Hardy–Weinberg equilibrium (at one locus), and we found no evidence for linkage disequilibrium (Manier & Arnold, 2005). Number of alleles and observed and expected heterozygosities in each population and over all populations were calculated in GENEPOP (Raymond & Rousset, 1995). Global estimates of F_{ST} were calculated using AMOVA (Excoffier *et al.*, 1992) in ARLEQUIN v. 2.000, and significance was assessed after 16 000 permutations. Standard errors of F_{ST} were estimated by jackknifing in FSTAT v. 2.9.3.2 (Goudet, 2001).

Population differentiation in quantitative traits

Differences in phenotypic traits between males and females were calculated for scale counts and colour scores using ANOVA in SAS v. 9.1 (SAS Institute, 2002). Sample sizes are shown in Table 1. In a preliminary analysis, all populations were pooled to assess sexual dimorphism over the entire study area. That analysis revealed that two scale counts (VENT and SUB) and four colour scores (DORGRN, DORHUE, LATGRN and LATHUE) were sexually dimorphic ($P < 0.0001$), after sequential Bonferroni and FDR control for multiple comparisons. For these traits, the difference between the sexes ranged from 0.2 to 0.4 standard deviations for the colourations traits to 0.7 standard deviations for the scalation traits. Females had more muted colours, and males had more ventral and subcaudal scales. Consequently, we analysed males and females separately.

We used a three-level partitioning of variance to estimate variance components and characterize the population structure of both microsatellite and phenotypic traits. For both kinds of traits, variation was partitioned into within-population, among-population (within ecotypes) and between ecotype components of variance by ANOVA. Variance components were calculated by equating observed mean squares with their expectations. For microsatellite traits, an analysis of molecular variance (AMOVA) was conducted in ARLEQUIN v. 2.000, applying results given by Excoffier *et al.* (1992). Thus, we partitioned the total variance in repeat number at each locus into three parts, $V = V_a + V_b + V_c$, where V_a is the between-ecotype component of variance, V_b is the among-population (within ecotype) component of variance and V_c is the within-population component of variance. F-statistics were computed from these descriptive components of variance (Excoffier *et al.*, 1992). $F_{ST} = (V_a + V_b)/V$ is the proportion of the total variance that resides among populations, including both the between-ecotype and the among-population (within ecotype) components. $F_{CT} = V_a/V$ is the proportion of the total variance that resides between ecotypes. $F_{SC} = V_b/(V_b + V_c)$ is the proportion of the total variance within ecotypes that resides among populations. Statistical significance of each F-statistic was computed in ARLEQUIN v. 2.000 by permutation analysis (Excoffier *et al.*, 1992). Global and population pairwise estimates of F_{ST} were also calculated using AMOVA in ARLEQUIN v. 2.000. Significance was assessed after 16 000 permutations for global estimates and 3000 permutations for pairwise estimates, using the default settings in ARLEQUIN.

For phenotypic traits, descriptive components of variance were estimated separately for each sex from a three-level nested ANOVA, computed using RANDOM in PROC GLM in SAS v. 9.1. The statistical significance of V_b was evaluated by testing the among-population within-ecotype mean square over the error (within-population)

mean square. The statistical significance of V_a was evaluated by testing the between-ecotype mean square over the among-population (within-ecotype) mean square. The statistical significance of $(V_a + V_b)$ was conducted by computing separate two-level ANOVAS (in which the ecotype identities of populations were ignored) and testing the among-population mean square over the error mean square. In all cases Type III sums of squares were used. The within-population component of genetic variance, V_c, was estimated by multiplying the observed within-population component of variance, V_w, by the corresponding heritability for that trait. Thus, $V_c = h^2 V_w$. The between-ecotype and among-population components of variance were equated with the corresponding genetic components of variance. To produce results analogous to the *F*-statistics for the microsatellite loci, *Q*-statistics (Spitze, 1993) were computed from these genetic components of variance using results given by Wright (1943) and Lande (1992) for quantitative traits. Thus,

$$V = V_a + V_b + 2V_c$$

$$Q_{ST} = (V_a + V_b)/V$$

$$Q_{CT} = V_a/V$$

$$Q_{SC} = V_b/(V_b + 2V_c)$$

These statistics have the same interpretations as the corresponding *F*-statistics. Notice that because estimates of heritability algebraically affect only a portion of the denominator in estimates of Q_{ST}, Q_{CT} and Q_{SC}, the value of heritability (and hence any error in its estimation) has relatively little effect on these *Q*-statistics. Notice, too, that an overestimate of h^2 (e.g. because of dominance or maternal effects) will lead to an underestimate of *Q*-statistics, and hence a conservative comparison with *F*-statistics. A Mantel test (Mantel, 1967; Mantel & Valand, 1970; Manly, 1997), implemented in ARLEQUIN v. 2.000 was used to assess the correlation between pairwise estimates of F_{ST} and Q_{ST} for each trait (significance over 10 000 permutations).

Identification of avian predators

A sample of 46 live *T. elegans* was captured over a 3-day period in June 2004 at the Gallatin field site and examined for bird culmen marks on their ventral surfaces. Seven of these snakes (15%) had culmen impressions that could unambiguously be attributed to avian attacks. To identify the avian predators responsible for these impressions, we photographed the culmens of a series of candidate avian predators that have been regularly observed at Gallatin over a 30-year period and compared them with culmen impressions on the snakes. The candidate avian predators were Great Blue Heron (*Ardea herodias*), Ring-billed Gull (*Larus delawarensis*), American Robin (*Turdus migratorius)* and Brewer's Blackbird (*Euphagus cyanocephalus*).

Results

Inheritance of colouration

Heritability estimates for colouration traits ranged from 0.14 to 0.80 in females and from 0.01 to 0.63 in males (Table 2). Sexual averages ranged from 0.08 to 0.65. Focusing on the sexual averages, the highest heritabilities were for DORHUE (0.65) and BKGRD (0.64) and the lowest were for LATRED (0.08) and LATSAT (0.10). Samples sizes were large enough to bound point estimates of heritability that were above about 0.32 away from zero at the 0.05 level. The alternative scoring schemes for colouration (RGB and HSL) showed comparable averages (0.32 and 0.36, respectively) and ranges for heritability. No pairs of traits showed significant genetic or environmental correlation.

Population differentiation in molecular and quantitative traits

Phenotypic traits showed subtle differences among populations within ecotypes but pronounced differences between ecotypes. Histograms of the PANTONE colour scores (Figs 3 and 4) suggest relative uniformity among meadow populations, as well as a striking difference in average colouration between lakeshore and meadow populations. The modal dorsal stripe colour in meadow populations is bright orange, but in lakeshore populations it tends to be tan. Dorsal and lateral stripe colours are more variable in lakeshore populations and include colours (e.g. brown and pink) that are not present in meadow populations. In all populations the colours of lateral stripes tend to be less bright than dorsal stripes. Similar trends are seen in background colouration. Background colour is darker in meadow populations than in lakeshore populations, with lakeshore populations also showing more variation in background colour (Fig. 5).

Turning to scalation, the most striking trend apparent in tabulations of population and ecotypic means (*Appendix S1* in *Supplementary material*, Table 3) is that lakeshore populations show higher average counts for all traits in both males and females. Average divergence between ecotypes for both females and males was 0.5 phenotypic standard deviations (SD) for scalation traits and 0.6 SD for colouration traits (Table 3). VENT and SUB showed the strongest divergence among scalation traits (0.6–1.0 SD) and BKGRD showed the largest divergence of colouration traits (2.2–2.3 SD). Statistical analyses reveal that all of these apparent trends are highly significant, and many traits show significant differentiation among populations within ecotypes (Table 4).
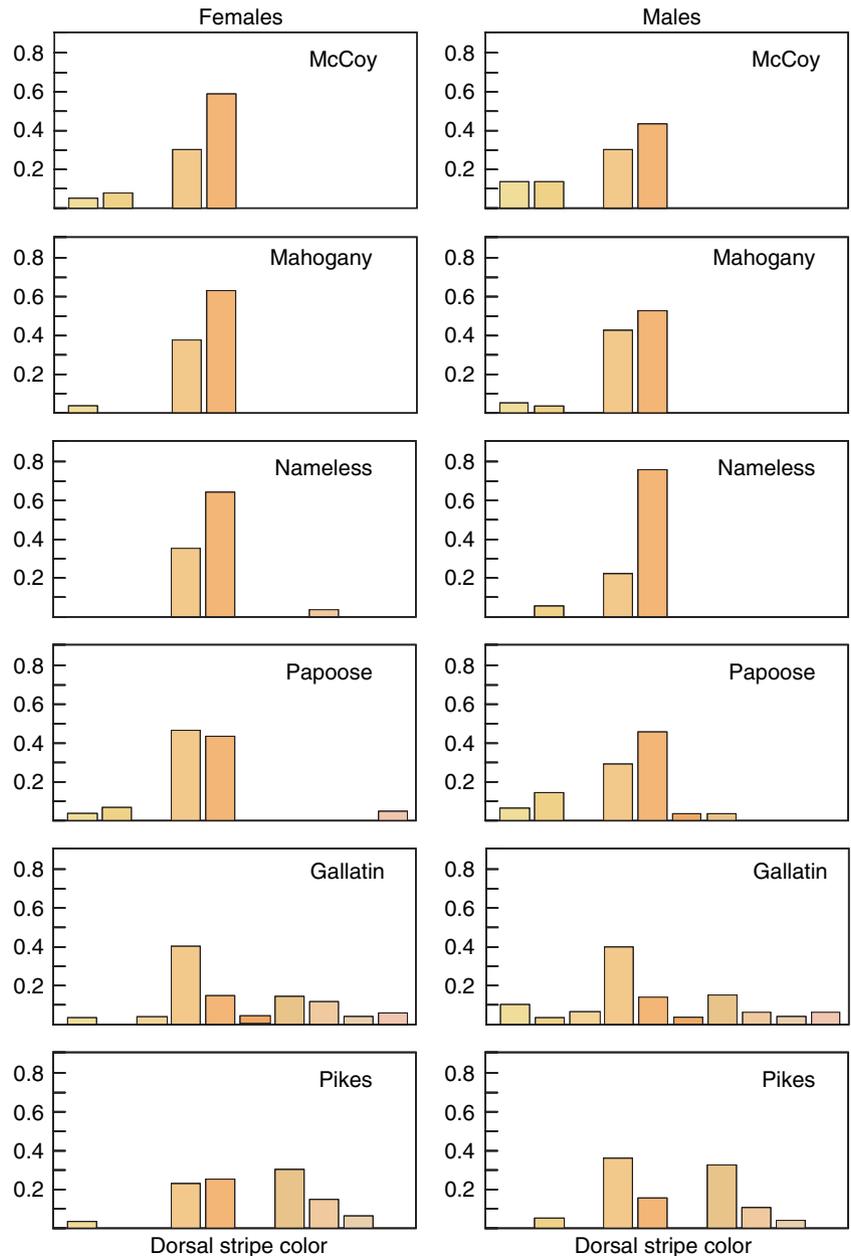
Fig. 3 Frequency histograms for dorsal stripe colours. Colours are, from left to right, Pantone 127, 128, 134, 135, 136, 137, 141, 148, 155 and 162. For Gallatin only, the 127 bin includes rare instances of Pantone 106, 120 and 121. The *y*-axis represents frequency. Localities at the top are furthest from Eagle Lake, such that the first four histograms correspond to meadow populations, and the last two to lakeshore populations.

In contrast to results for phenotypic traits, the AMOVA for microsatellite data (sexes combined) revealed that the overwhelming proportion of variance was within populations. Thus, the averages across nine loci were 96% (± 0.6% SE) within populations, 1% (± 0.3% SE) among populations within ecotypes and 3% (± 0.6% SE) between ecotypes. The average percentages for males and females across all scalation traits are, respectively, 58% (± 7% SE) within populations, 15% (± 8% SE) among populations within ecotypes and 28% (± 10% SE) among ecotypes. The average percentages for males and females for colouration traits are, respectively, 64% (± 6% SE) within populations, 6% (± 1.6% SE) among

populations within ecotypes and 30% (± 7% SE) among ecotypes. (These summary figures are from ANOVAs whose *P*-values are given in Table 4 but are otherwise not reported in this article. Within-population variance components were converted to genetic components using heritability estimates.) Thus, for the two sets of phenotypic traits, on the order of 40% of variation resided among populations (vs. 4% for microsatellites), and the majority of among-population variation resided between ecotypes. This discrepancy between microsatellite and phenotypic traits in among-population differentiation suggests that strong diversifying selection has acted on the phenotypic traits. We computed *F*- and
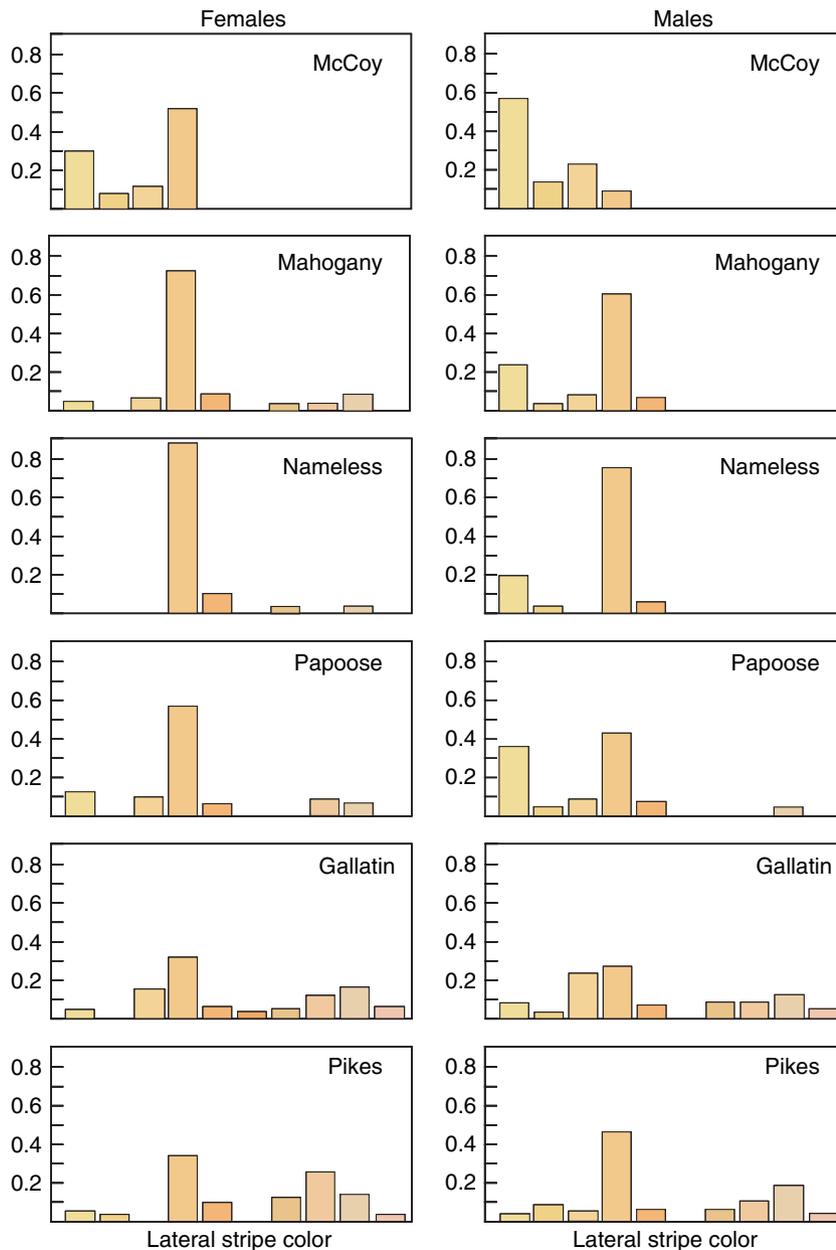
Fig. 4 Frequency histograms for lateral stripe colours. For Gallatin only, the 127 bin includes rare instances of Pantone 113, 120 and 121. Other conventions as in Fig. 2.

Q-statistics to conduct a more rigorous test of this selection hypothesis.

Fixation statistics revealed major differences between microsatellite and phenotypic traits in population differentiation (Table 4). In the case of the microsatellites, population differentiation (F_{ST}) accounted for an average of only 0.04 (± 0.01 SE) of total variation. Although for individual loci, F_{ST} ranged from only 0.004 to 0.06, this degree of differentiation was statistically significant (at the $\alpha = 0.01$ level) for seven out of nine loci. Virtually none of this proportion of among-population differentiation could be attributed to differences between ecotypes. On the average, F_{CT} was only 0.02 (± 0.01 SE) and

F_{CT} values were not significant for any of the nine loci. We were able to detect significant variation among populations within ecotypes. F_{SC} averaged only 0.01 (± 0.004 SE), but six out of nine loci showing significant values (at the $\alpha = 0.05$). The large variation in population sample size (16–140) may impact estimates of population subdivision, so we repeated the AMOVA using a dataset with more equalized population sizes (16–30 individuals; described in Manier & Arnold, 2005). These results were virtually identical for F_{SC} and F_{CT}, and similar for F_{ST} (0.03 instead of 0.04).

In contrast, average population differentiation was seven times greater (Q_{ST}/F_{ST}) for scalation traits and six
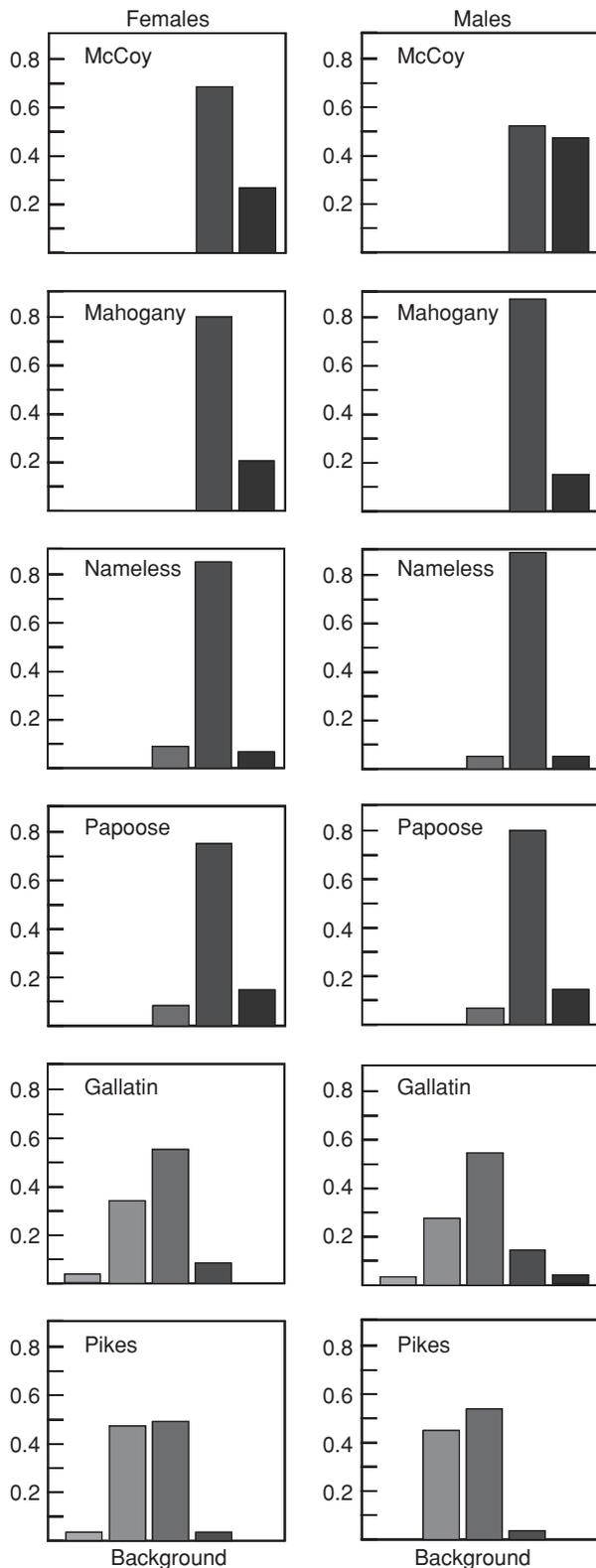
Fig. 5 Frequency histograms for background colour. Grey shades are, left to right, 3, 4, 5, 6 and 7. Other conventions as in Fig. 2.

times greater for colouration traits. In the case of scalation traits, Q_{ST} ranged from 0.09 to 0.44 and averaged 0.28 (± 0.05 SE). In the case of colouration traits, Q_{ST} ranged from 0.02 to 0.71 and averaged 0.25 (± 0.05 SE). These statistics identified the scalation traits ILAB, VENT and SUB, and the colouration traits BKGRD, DORBLU, LATRED and LATSAT as traits that have experienced especially strong diversifying selection. The extreme case was BKGRD with a value for Q_{ST} (0.71) nearly 18 times greater than the microsatellite average. Most of the among-population differentiation (Q_{ST}) in phenotypic traits could be attributed to differences between ecotypes (Q_{CT}). In particular, the seven traits just highlighted as having high values for Q_{ST}, also showed large values for Q_{CT} that were statistically significant in one or both sexes. It is also apparent that some phenotypic traits have experienced no or only weak diversifying selection. Thus, although Q_{ST} was 0.28 for POST, this scalation trait showed no significant differentiation between ecotypes. Among colouration traits, DORHUE, LATGRN, LATHUE and LATLT showed statistically insignificant values for both Q_{ST} and Q_{CT}. Finally, differentiation among populations within ecotypes (Q_{SC}) averaged 0.10 (± 0.05 SE) for scalation traits, and 0.04 (± 0.01 SE) for colouration traits. Although small, these percentages are, respectively, ten and four times greater than the microsatellite average.

Mantel tests of pairwise F_{ST} and Q_{ST} matrices showed no evidence of correlated patterns of population differentiation. F_{ST} and Q_{ST} were not significantly correlated for any phenotypic trait. Pearson correlation coefficients varied from −0.237 to 0.302 for males and −0.283 to 0.290 for females.

Putative agents of selection

Birds are probably the selective agents responsible for the observed patterns of selection on stripe colours. The photographs of culmen impressions on seven snakes from the Gallatin population matched the culmens of two avian predators. Six out of the seven snakes had impressions from Great Blue Herons (*Ardea herodias*), and one had impressions from a Brewer's Blackbird (*Euphagus cyanocephalus*; *Appendix S2*). In two instances, birds were directly observed preying on garter snakes near this study site. A Great Blue Heron was observed on the Eagle Lake shoreline, 5 km south-west of Gallatin, attacking and flying off with a large gravid *T. elegans* (C. Cox, personal communication). Jayne & Bennett (1990) observed an American Robin (*Turdus migratorius*) capturing and flying off with a juvenile *T. sirtalis* at a site 15 km from Gallatin. Although these results implicate birds as agents of selection in our study system (especially on colouration), we have not demonstrated that they are actually responsible for the selection we detected. Such a demonstration would require experimental manipulation of avian predators in conjunction with measurement of selection.

Table 3 Divergence between lakeshore and meadow ecotypes in phenotypic traits.

Trait	Lakeshore				Meadow				SD		Divergence	
	Male		Female		Male		Female					
	N	Mean	N	Mean	N	Mean	N	Mean	Male	Female	Male	Female
Scalation												
VENT	703	171.9	802	169.5	229	168.6	428	164.2	5.4	5.0	0.6	1.0
SUB	637	87.6	720	82.2	197	80.7	315	76.0	8.0	7.4	0.9	0.8
MID	727	20.3	843	20.4	231	19.7	430	19.8	1.0	1.0	0.6	0.6
SLAB	731	16.0	847	16.1	228	15.9	432	16.0	0.5	0.6	0.2	0.2
ILAB	729	20.2	846	20.3	228	19.8	431	19.9	0.7	0.8	0.6	0.5
POST	731	6.1	848	6.1	227	5.9	429	6.0	0.6	0.7	0.3	0.1
Average											0.5	0.5
Colouration												
DORRED	433	243.1	163	242.9	454	244.4	204	245.0	4.8	4.6	0.3	0.5
DORGRN	433	195.3	163	191.7	454	185.0	204	184.3	13.7	12.4	0.8	0.6
DORBLU	433	114.3	163	113.6	454	98.7	204	99.8	16.3	18.1	1.0	0.8
DORHUE	433	25.1	163	24.1	454	24.0	204	23.5	3.4	2.7	0.3	0.2
DORSAT	433	202.5	163	201.9	454	209.1	204	210.7	14.1	13.9	0.5	0.6
DORLT	433	168.0	163	167.6	454	161.2	204	162.0	8.1	9.0	0.8	0.6
LATRED	375	243.3	160	243.1	393	245.0	196	245.5	4.7	4.8	0.4	0.5
LATGRN	375	201.6	160	198.3	393	205.6	196	199.1	11.8	11.5	0.3	0.1
LATBLU	375	125.0	160	125.1	393	118.4	196	116.2	16.3	19.1	0.4	0.5
LATHUE	375	25.8	160	24.7	393	28.0	196	25.8	3.8	3.2	0.6	0.3
LATSAT	375	199.6	160	198.5	393	206.7	196	208.6	17.5	18.9	0.4	0.5
LATLT	375	173.1	160	173.1	393	170.9	196	170.0	7.4	8.5	0.3	0.4
BKGRD	440	4.8	163	4.7	461	6.1	203	6.1	0.6	0.6	2.2	2.3
Average											0.6	0.6

Divergence is expressed in units of average within-population phenotypic standard deviation (SD).

N, sample size.

Discussion

The phenomenon of ecotypic variation

Ecotypic variation refers to a repeated spatial pattern of population differentiation that coincides with particular environmental variables (Mayr, 1963). In the botanical literature such spatial coincidence has long been interpreted as evidence for local adaptation to ecological features (Turesson, 1922). With the notable exception of fish systems, such as sticklebacks (e.g. Moodie, 1972; Schluter *et al.*, 2004; Baker *et al.*, 2005), vertebrate biologists have been relatively slow to use this concept. Aside from our work on *T. elegans*, a few examples of ecotypic variation have been reported in snakes. Fox (1951) reported ecotypic variation in *Thamnophis atratus* in the Eel River drainage of northern California that parallels the colouration differentiation in our study system.

Colour pattern differences between island and mainland water snakes (*Nerodia sipedon*) and garter snakes (*T. sirtalis*) in Lake Erie have been viewed as an equilibrium between selection and migration in which strong selection for crypticity maintains population differentiation in colouration in both species in the presence of considerable gene flow (Camin & Ehrlich, 1958; King, 1993a,b; Lawson & King, 1996; Hendry *et al.*, 2001; King & Lawson, 2001a,b; Bittner & King, 2003; Ray & King, 2006). These examples, together with our results, demonstrate that ecotypic variation and other kinds of local adaptation are relatively common in vertebrates.

Ecotypic variation and local adaptation in *T. elegans*

The historical importance of selection revealed by the F_{ST}-Q_{ST} analysis reinforces the results of correlational selection analyses on vertebral numbers in these and related garter snakes. A previous study showed that the effect of vertebral numbers on growth rate in the Gallatin population can be portrayed as a bivariate ridge (Arnold, 1988). In other words, both VENT and SUB experience stabilizing selection, whereas the ridge's positive slope reflects correlational selection on VENT and SUB. In a study of the effect of vertebral numbers on locomotory performance in a closely related species (*T. radix*), Arnold & Bennett (1988) found significant positive correlational selection on VENT and SUB, although coefficients of stabilizing selection on these two traits were not statistically significant. Thus, a goal of future work will be to determine whether selectively driven differentiation, as assayed by our F_{ST}-Q_{ST} analysis, occurs along the selective ridges that were diagnosed in the correlational studies

Table 4 Hierarchical analysis of population structure for microsatellite loci, scalation traits and colouration traits.

Microsatellites	F_{SC}	P	F_{ST}	P	F_{CT}	P
1	0.02	0.005	0.02 (0.01)	0.003	−0.00	0.46
2	0.00	0.20	0.06 (0.02)	< 0.0001	0.05	0.07
3	0.01	0.002	0.04 (0.01)	< 0.0001	0.03	0.07
4	−0.01	0.95	0.01 (0.01)	0.24	0.02	0.07
5	0.01	0.13	0.00 (0.01)	0.14	−0.00	0.46
6	0.03	0.008	0.06 (0.03)	< 0.0001	0.04	0.20
7	0.01	0.01	0.05 (0.02)	< 0.0001	0.04	0.06
8	0.01	0.03	0.05 (0.02)	< 0.0001	0.04	0.07
9	0.02	0.006	0.03 (0.03)	0.002	0.01	0.26
Average	0.01 (0.004)		0.04 (0.01)		0.02 (0.01)	
Scalation	Q_{SC}	P	Q_{ST}	P	Q_{CT}	P
VENT	0.08	< 0.0001, < 0.0001	0.31	< 0.0001, < 0.0001	0.24	0.003, 0.13
SUB	0.16	< 0.0001, < 0.0001	0.35	< 0.0001, < 0.0001	0.23	0.04, 0.13
MID	0.06	0.44, < 0.0001	0.21	< 0.0001, < 0.0001	0.15	0.001, 0.12
ILAB	0.03	0.004, 0.62	0.44	< 0.0001, < 0.0001	0.42	0.01, 0.001
SLAB	−0.02	0.47, 0.78	0.09	0.02, 0.005	0.11	0.07, 0.01
POST	0.31	< 0.0001, < 0.0001	0.28	0.01, 0.01	−0.05	0.37, 0.52
Average	0.10 (0.05)		0.28 (0.05)		0.18 (0.07)	
Colouration	Q_{SC}	P	Q_{ST}	P	Q_{CT}	P
DORRED	0.04	0.004, 0.008	0.13	< 0.0001, 0.0002	0.09	0.03, > 0.05
DORGRN	0.05	0.03, 0.0003	0.16	< 0.0001, < 0.0001	0.12	0.02, > 0.05
DORBLU	0.03	> 0.05, 0.05	0.39	< 0.0001, < 0.0001	0.37	0.003, 0.006
LATRED	0.02	> 0.05, > 0.05	0.39	< 0.0001, 0.001	0.36	0.002, > 0.05
LATGRN	0.12	< 0.0001, < 0.0001	0.15	0.02, < 0.0001	0.04	> 0.05, > 0.05
LATBLU	0.02	> 0.05, > 0.05	0.13	< 0.0001, 0.004	0.10	0.03, > 0.05
DORHUE	0.02	> 0.05, 0.007	0.02	0.03, > 0.05	0.01	> 0.05, > 0.05
DORSAT	0.04	0.01, 0.04	0.23	< 0.0001, < 0.0001	0.20	0.01, 0.03
DORLT	0.03	> 0.05, 0.02	0.26	< 0.0001, < 0.0001	0.24	0.01, 0.02
LATHUE	0.10	< 0.0001, < 0.0001	0.21	< 0.0001, < 0.0001	0.13	> 0.05, > 0.05
LATSAT	0.02	> 0.05, > 0.05	0.34	< 0.0001, 0.0004	0.32	0.002, > 0.05
LATLT	0.01	> 0.05, > 0.05	0.06	0.002, > 0.05	0.05	> 0.05, > 0.05
BKGRD	0.04	0.002, 0.0003	0.71	< 0.0001, < 0.0001	0.70	0.0002, 0.0006
Average	0.04 (0.01)		0.25 (0.05)		0.21 (0.05)	

Q_{SC}, Q_{ST} and Q_{CT} columns show the average of male and female values, P columns for scalation and colouration traits show significance levels for separate analyses of females and males, respectively. SE values are given in parentheses.

of selection. Although the correlational analysis revealed an intermediate optimum for VENT and SUB in the Gallatin lakeshore population (Arnold, 1988), we have not conducted a parallel study to check for an intermediate optimum for lower counts in meadow populations. Finally, it might be illuminating to compare a correlational analysis of selection on the colouration traits with the F_{ST}-Q_{ST} results.

It should be noted that the F_{ST}-Q_{ST} analysis does not account for correlated responses to selection that arise from genetic correlations among traits (Lande, 1979). The result that $Q_{ST} > F_{ST}$ may suggest diversifying selection on the trait in question or on a genetically correlated trait, producing an illusion of selection on the trait in question. Although genetic correlations among scalation traits (Arnold & Phillips, 1999) and among colouration traits appear to be relatively weak or nonexistent, they may have helped to produce correlated responses to selection. We return to the ambiguity induced by genetic

correlations in F_{ST}-Q_{ST} analysis in our concluding remarks.

The population differentiation that we assessed with Q-statistics is not likely to represent environmental effects or sampling artefacts. Laboratory rearing of individuals from multiple populations of both ecotypes in the course of conducting a cross between inland and coastal populations of *T. elegans* (Arnold, 1988) showed that individual, population and ecotypic differences in scalation and colouration were congenital and ontogenetically stable. Furthermore, F_1 progenies in this cross showed maternal effects for birth weight but not for scalation or colouration. Temperature effects during gestation in viviparous snakes were once thought to affect vertebral numbers and the other kinds of scalation traits used in our study (Fox, 1948). More extensive experimentation, however, has shown that these traits are extremely well buffered against temperature effects (Arnold & Peterson, 2002). Thus, the population differentiation that we

observed probably does not represent direct environmental effects. Likewise, bias and sampling errors in heritability estimates probably had little effect on our results and conclusions. In principle, the actual value of heritability can fluctuate through time or vary from population to population (Falconer & MacKay, 1996). Analytical models and simulation studies show, however, that heritability and other polygenic genetic parameters can achieve stable equilibria, especially in large populations with persistent selection regimes (Lande, 1976; Jones *et al.*, 2003, 2004). Although the persistence of selection has not been resolved in our study system, the effective sizes of our populations (a few to several hundred individuals; Manier & Arnold, 2005) would help to dampen fluctuations. Furthermore, comparative studies indicate that the heritabilities of scalation traits are relatively stable in garter snakes (Dohm & Garland, 1993; Arnold & Phillips, 1999). With respect to sampling issues, our use of full-sib estimates of heritability could have led to an upward bias. However, because of the way that heritability enters the calculation of Q-statistics, this potential bias would have produced conservative estimates of population differentiation. Furthermore, Q-statistics are relatively insensitive to the value of h^2 and hence to errors in its estimation. For example, in our data, a value of 0.48 for heritability of VENT in females, in conjunction with estimates of V_a, V_b and V_c, produced an estimate of 0.39 for Q_{ST}. Suppose this value of heritability was in error by 100%, so that the true value was 0.24, then the corresponding estimate for Q_{ST} would have been 0.56. In other words, even an extreme error in heritability would have produced only a 30% error in Q_{ST}. Such an error is small, when we consider that in our study, Q_{ST} was seven times greater than F_{ST} for scalation traits and six times greater for colouration traits.

Differentiation in colouration and scalation traits revealed in the present analyses coincides with ecotypic differences in life history (Bronikowski & Arnold, 1999; Sparkman *et al.*, 2007). Thus, the syndrome of differentiation between ecotypes includes scalation, vertebral numbers, numerous aspects of colouration, as well as growth rate and body size–fecundity relationships. Remarkably, this extensive differentiation occurs over a distance of a few kilometres, between populations connected by moderate to high gene flow (Manier & Arnold, 2005).

As expected, vertebral number showed evidence of diversifying selection between ecotypes, with lakeshore snakes having more vertebrae than meadow snakes. This trend corresponds with a difference in push-point density at these sites (lower on the lakeshore, higher in meadow sites). A similar association was observed in a comparison of coastal and inland populations of *T. elegans*, but because juvenile snakes with more vertebrae crawl faster at all push-point densities, the biomechanical basis for these associations remains unclear (Kelley *et al.*, 1997; Arnold & Phillips, 1999).

The direction of differentiation in other scale counts is consistent with the hypothesis that selection favours the ability to ingest large prey in lakeshore habitats. Studies of stomach contents reveal that lakeshore populations feed on fish and leeches, whereas the meadow populations feed primarily on anuran larvae and leeches (Kephart & Arnold, 1982; Kephart, 1982). The lakeshore diet includes occasional ingestion of large fish, but feeding performance trials are needed to determine whether the observed differences in scalation do indeed enhance the ability of lakeshore snakes to eat large prey.

The F_{ST}-Q_{ST} and F_{CT}-Q_{CT} contrasts suggest that diversifying selection is responsible for population differentiation in scalation and colouration characters. For many of the scalation traits and several of the colouration traits, population differentiation is five- to ten-fold more pronounced than would be expected under a drift-migration balance. Furthermore, most of the population differentiation in phenotypic traits coincides with ecotypic differences. The most likely explanation for these results is local adaptation to lakeshore and meadow habitats in scalation and colouration. Thus, *T. elegans* in the Eagle Lake basin have adapted to lakeshore habitats by evolving more body and tail vertebrae, more scale rows at midbody, more infralabial and supralabial scales, lighter background colour, and bluer dorsal stripes. Concomitantly, adaptation in all these traits has occurred in the opposite direction in meadow habitats.

The F_{SC}-Q_{SC} contrasts suggest that diversifying selection is also responsible for population differentiation within each of the two ecotypes. This level of differentiation in scalation and colouration traits is, on average, ten- and four-fold, respectively, more pronounced than what we would expect by drift. Most of the scalation and colouration traits show highly significant Q_{SC}-values, in contrast to F_{SC}- or to Q_{CT}-values. Although it should be noted that we have considerably more statistical power for Q_{SC} than for Q_{ST} or Q_{CT}, it nevertheless appears that subtle local adaptation within lakeshore and meadow habitats has involved numerous aspects of scalation and colouration. Future studies might determine whether small-scale differences among meadow sites in vegetation and other habitat characteristics contribute to selection on colouration. Our results also suggest that adaptation to local conditions may occur on a much smaller spatial scale than is typically assayed in snakes and other terrestrial vertebrates (but see Hoekstra *et al.*, 2004).

Local adaptation in our study system, revealed by Q_{SC}, Q_{ST} and Q_{CT}, is especially remarkable given the close proximity of populations (1.3–19.9 km apart) and moderate levels of gene flow among them. The Gallatin shoreline and the Papoose Meadows populations present a case in point illustrating this pattern. An earlier study established that substantial gene flow occurs among 20 populations in our study system (F_{ST} = 0.024; average N_em = 0.4), primarily unidirectionally from the major source population, Papoose Meadows (average N_em from

Papoose = 1.4; Manier & Arnold, 2005). Gallatin and Papoose are only 4 km apart and are connected by an intermittent stream, Papoose Creek, which is a known dispersal corridor for *T. elegans* in wet years (S.J. Arnold, unpublished data). Thus, Gallatin and other lakeshore sites have been able to differentiate in the presence of persistent migration from Papoose Meadows, and both lakeshore and meadow populations have maintained their ecotypic identities on a small spatial scale.

Endler (1990) has persuasively argued that spectro-graphic measurement of colour is preferable to matching with colour standards. Because spectrographic measurements are time-consuming, we opted for colour matching to maximize sample sizes. We circumvented some of Endler's objections to subjective matching by using one person to score colours under a uniform condition of lighting that coincided with the most likely time of heavy predation (mid-morning). Nevertheless, the human visual system undoubtedly differs from that of birds (Chen & Goldsmith, 1986; Jane & Bowmaker, 1988), although the differences may not be large (Ali & Klyne, 1985; Chen & Goldsmith, 1986). Although stripe colours clearly experience selection in our study system, the details of the selection results might be different if a bird-based rather than a human-based scoring scheme had been employed. For this reason, our selection results for colouration should be viewed with caution. Nevertheless, the RGB and the HSL scoring systems gave comparable results, even though the RGB scheme probably better approximates avian colour vision. This consistency between the two coding schemes suggests that the overall picture of selection may be robust to scoring method.

Concluding remarks

As in many other studies (Merilä & Crnokrak, 2001; McKay & Latta, 2002), Q_{ST} greatly exceeds F_{ST} in our study system. This result implies that diversifying selection acting on phenotypic traits has produced departures from neutral expectations. Although this interpretation is correct *per se*, it is subject to some qualifications. Differences among traits in Q_{ST} can reflect differences in both inheritance and selection. For example, a trait may show $Q_{ST} > F_{ST}$, not because it has been a target of selection, but because it has responded to selection that has acted on one or more genetically correlated traits. Although a truly multivariate F_{ST}-Q_{ST} comparison has not yet been devised, the direct role of selection in differentiation can be diagnosed with multivariate retrospective analyses that use the *G*-matrix for a set of traits (Lande, 1979; Jones *et al.*, 2004).

Studies of selection within a generation do not suffer from Q_{ST}'s problem of confounding selection with response to selection. Consequently, multivariate analyses of ongoing selection (Lande & Arnold, 1983; Schluter & Nychka, 1994) are a better vehicle than F_{ST}-Q_{ST}

analysis for identifying the actual targets of phenotypic selection. Statistical power is, however, a serious concern in these approaches, especially when stabilizing selection is weak and the phenotypic mean is close to an intermediate optimum (Hersch & Phillips, 2004). The analyses of selection discussed here (Arnold & Bennett, 1988; Arnold, 1988) undoubtedly suffered from this problem.

F_{ST}-Q_{ST} comparisons are limited in their inability to identify the actual agents of selection. Even a few observations of selective events, as in the present study, can help remedy this situation and so direct and illuminate the interpretation of selection analyses. Thus, the overall message of our study is that multiple levels of inquiry into the nature and consequences of selection can yield a synthetic overview of the process that a standard F_{ST}-Q_{ST} comparison alone cannot provide.

References

Alexander, A.A. & Gans, C. 1966. The pattern of dermal-vertebral correlation in snakes and amphisbaenians. *Zool. Meded (Leiden)* **41**: 171–190.

Ali, M.A. & Klyne, M.A. 1985. *Vision in Vertebrates*. Plenum Press, New York.

Arnold, S.J. 1981. Behavioral variation in natural populations. I. Phenotypic, genetic and environmental correlations between chemoreceptive responses to prey in the garter snake, *Thamnophis elegans. Evolution* **35**: 489–509.

Arnold, S.J. 1988. *Quantitative Genetics and Selection in Natural Populations: Microevolution of Vertebral Numbers in the Garter Snake Thamnophis elegans*. Second International Conference on Quantitative Genetics. Sinauer, Sunderland, MA, USA.

Arnold, S.J. & Bennett, A.F. 1988. Behavioural variation in natural populations. V. Morphological correlates of locomotion in the garter snake (*Thamnophis radix*). *Biol. J. Linn. Soc.* **34**: 175–190.

Arnold, S.J. & Peterson, C.R. 2002. A model for optimal reaction norms: the case of the pregnant garter snake and her temperature sensitive embryos. *Am. Nat.* **160**: 306–316.

Arnold, S.J. & Phillips, P.C. 1999. Hierarchical comparison of genetic variance-covariance matrices. II. Coastal-inland divergence in the garter snake, *Thamnophis elegans*. *Evolution* **53**: 1516–1527.

Baker, A.J. 1992. Genetic and morphometric divergence in ancestral European and descendent New Zealand populations of chaffinches (*Fringilla coelebs*). *Evolution* **46**: 1784–1800.

Baker, J., Cresko, W., Foster, S. & Heins, D. 2005. Life history differentiation of benthic and limnetic ecotypes in a polytypic population of threespine stickleback, *Gasterosteus aculeatus*. *Evol. Ecol. Res.* **7**: 121–131.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**: 289–300.

Bittner, T.D. & King, R.B. 2003. Gene flow and melanism in garter snakes revisited: a comparison of molecular markers and island vs. coalescent models. *Biol. J. Linn. Soc.* **79**: 389–399.

Brodie, E.D. Jr. 1992. Correlational selection for colour pattern and antipredator behavior in the garter snake *Thamnophis ordinoides*. *Evolution* **46**: 1284–1298.

Bronikowski, A.M. 2000. Experimental evidence for the adaptive evolution of growth rate in the garter snake (*Thamnophis elegans*). *Evolution* **54**: 1760–1767.

Bronikowski, A.M. & Arnold, S.J. 1999. The evolutionary ecology of life history variation in the garter snake *Thamnophis elegans*. *Ecology* **80**: 2314–2325.

Camin, J.H. & Ehrlich, P.R. 1958. Natural selection in water snakes (*Natrix sipedon* L.) on islands in Lake Erie. *Evolution* **12**: 504–511.

Cano, J.M., Laurila, A., Palo, J. & Merilä, J. 2004. Population differentiation in G matrix structure due to natural selection in *Rana temporaria*. *Evolution* **58**: 2013–2020.

Chen, D.-M. & Goldsmith, T.M. 1986. Four spectral classes of cone in the retinas of birds. *J. Comp. Physiol. A* **159**: 473–479.

Dohm, M.R. & Garland, T.J. Jr. 1993. Quantitative genetics of scale counts in the garter snake *Thamnophis sirtalis*. *Copeia* **1993**: 987–1002.

Endler, J. 1990. On the measurement and classification of colour in studies of animal colour patterns. *Biol. J. Linn. Soc.* **41**: 315–352.

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

Falconer, D.S. & MacKay, T.F.C. 1996. *Introduction to Quantitative Genetics*, 4th edn. Prentice Hall, Harlow, UK.

Fox, W. 1948. Effect of temperature on development of scutellation in the garter snake, *Thamnophis elegans atratus*. *Copeia* **1948**: 252–262.

Fox, W. 1951. Relationships among garter snakes of the *Thamnophis elegans* Rassenkreis. *Univ. Calif. Publ. Zool.* **50**: 485–530.

Garner, T.W.J. & Larsen, K.W. 2005. Multiple paternity in the western terrestrial garter snake, *Thamnophis elegans*. *Can. J. Zool.* **83**: 656–663.

Gasc, J.-P., Cattaert, D., Chasserat, C. & Clarac, F. 1989. Propulsive action of a snake pushing against a single site: its combined analysis. *J. Morphol.* **201**: 315–329.

Gomez-Mestre, I. & Tejedo, M. 2004. Contrasting patterns of quantitative and neutral genetic variation in locally adapted populations of the natterjack toad, *Bufo calamita*. *Evolution* **58**: 5343–5352.

Goudet, J. 2001. *FSTAT, A Program to Estimate and Test Gene Diversities and Fixation Indices* (version 2.9.3). Available at: http://www2.unil.ch/popgen/softwares/fstat.htm

Guo, S.W. & Thompson, E.A. 1992. Performing the exact test for Hardy-Weinberg proportions for multiple alleles. *Biometrics* **48**: 361–372.

Hendry, A.P., Day, T. & Taylor, E.B. 2001. Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution* **55**: 459–466.

Hersch, E.I. & Phillips, P.C. 2004. Power and potential bias in field studies of natural selection. *Evolution* **58**: 479–485.

Hoekstra, H.E., Drumm, K.E. & Nachman, M.W. 2004. Ecological genetics of adaptive colour polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution* **58**: 1329–1341.

Jacobs, G.H. 1981. *Comparative Colour Vision*. Academic Press, New York.

Jane, S.D. & Bowmaker, J.K. 1988. Tetrachromatic colour vision in the duck (*Anas platyrhynchos* L.): microspectrophotometry of visual pigments and oil droplets. *J. Comp. Physiol. A* **162**: 225–235.

Jayne, B.C. 1988. Muscular mechanisms of snake locomotion: an electromyographic study of lateral undulation of the Florida Banded Water Snake (*Nerodia fasciata*) and the Yellow Rat Snake (*Elaphe obsoleta*). *J. Morphol.* **197**: 159–181.

Jayne, B.C. & Bennett, A.F. 1990. Selection on locomotor performance capacity in a natural population of garter snakes. *Evolution* **44**: 1204–1229.

Jones, A.G., Arnold, S.J. & Bürger, R. 2003. Stability of the G-matrix in a population experiencing stabilizing selection, pleiotropic mutation, and genetic drift. *Evolution* **57**: 1747–1760.

Jones, A.G., Arnold, S.J. & Bürger, R. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution* **58**: 1639–1654.

Kelley, K.C., Arnold, S.J. & Gladstone, J. 1997. The effects of substrate and vertebral number on locomotion in the garter snake, *Thamnophis elegans*. *Funct. Ecol.* **11**: 189–198.

Kephart, D.G. 1981. *Population Ecology and Population Structure of Thamnophis elegans and Thamnophis sirtalis*. PhD dissertation, Department of Biology, University of Chicago, Chicago.

Kephart, D.G. 1982. Microgeographic variation in the diets of garter snakes. *Oecologia* **52**: 287–291.

Kephart, D.G. & Arnold, S.J. 1982. Garter snake diets in a fluctuating environment: a seven-year study. *Ecology* **63**: 1232–1236.

King, R.B. 1993a. Colour-pattern variation in lake Erie water snakes: prediction and measurement of natural selection. *Evolution* **47**: 1819–1833.

King, R.B. 1993b. Colour pattern variation in Lake Erie water snakes: inheritance. *Can. J. Zool.* **71**: 1985–1990.

King, R.B. & Lawson, R. 2001a. Colour-pattern variation in Lake Eire water snakes: the role of gene flow. *Evolution* **49**: 885–896.

King, R.B. & Lawson, R. 2001b. Patterns of population subdivision and gene flow in three sympatric natricine snakes. *Copeia* **2001**: 602–614.

Klauber, L.M. 1941. The correlation between scalation and life zones in San Diego county snakes. *Bull. Zool. Soc. San Diego* **17**: 73–79.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**: 314–334.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**: 402–416.

Lande, R. 1992. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution* **46**: 381–389.

Lande, R. & Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.

Lawson, R. & King, R.B. 1996. Gene flow and melanism in Lake Erie garter snake populations. *Biol. J. Linn. Soc.* **59**: 1–19.

Lindell, L.E., Forsman, A. & Merilä, J. 1993. Variation in number of ventral scales in snakes: effects on body size, growth rate and survival in the adder *Vipera berus*. *J. Zool.* **230**: 101–115.

Manier, M.K. & Arnold, S.J. 2005. Population genetic analysis identifies source-sink dynamics for two sympatric garter snake species (*Thamnophis elegans* and *T. sirtalis*). *Mol. Ecol.* **14**: 3965–3976.

Manier, M.K. & Arnold, S.J. 2006. Ecological correlates of population genetic structure: a comparative approach using a vertebrate metacommunity. *Proc. Roy. Soc. Lond. B* **273**: 3001–3009.

Manly, B.F. 1997. *Randomization and Monte Carlo Methods in Biology*, 2nd edn. Chapman and Hall, New York.

Mantel, N. 1967. The detection of disease clustering and generalized regression approach. *Cancer Res.* **27**: 209–220.

Mantel, N. & Valand, R.S. 1970. A technique of nonparametric multivariate analysis. *Biometrics* **26**: 547–558.

Mayr, E. 1963. *Animal Species and Evolution*. Belknap Press of Harvard University Press, Cambridge, MA, USA.

McKay, J.K. & Latta, R.G. 2002. Adaptive population divergence: markers, QTL and traits. *Trends Ecol. Evol.* **17**: 285–291.

Merilä, J. & Crnokrak, P. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* **14**: 892–903.

Moodie, G.E.E. 1972. Predation, natural selection and adaptation in an unusual threespine stickleback. *Heredity* **28**: 155–167.

Munsell Colour Company. 1976. *Munsell Book of Colour, Glossy Finish Collection*. Munsell/Macbeth/Kollmorgen Corp, Baltimore.

Palo, J.U., O'Hara, R.B., Laugen, A.T., Laurila, A., Primmer, C.R. & Merilä, J. 2003. Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Mol. Ecol.* **12**: 1963–1978.

Phillips, P.C. 1998. *H2boot: Bootstrap Estimates and Tests of Quantitative Genetic Data*. University of Texas, Arlington, Software available at: http://www.uoregon.edu/~pphil/software.html.

Ray, J.M. & King, R.B. 2006. The temporal and spatial scale of microevolution: fine-scale colour pattern variation in the Lake Erie watersnake, *Nerodia sipedon insularum*. *Evol. Ecol. Res.* **8**: 915–928.

Raymond, M. & Rousset, F. 1995. GENEPOP Version 1.2: population genetics software for exact tests and ecumenicism. *J. Hered.* **86**: 248–249.

Rice, W.R. 1989. Analysing tables of statistical tests. *Evolution* **43**: 223–225.

SAS Institute. 2002. *SAS/STAT Software User's Guide*, Rel. 9.1. SAS Institute, Inc., Cary, NC, USA.

Schluter, D. & Nychka, D. 1994. Exploring fitness surfaces. *Am. Nat.* **143**: 597–616.

Schluter, D., Clifford, E.A., Nemethy, M. & McKinnon, J.S. 2004. Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**: 809–822.

Schneider, S., Roessli, D. & Excoffier, L. 2000. *Arlequin ver 2.000: A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schwartz, J.M., McCracken, G.F. & Burghardt, G.M. 1989. Multiple paternity in wild populations of the garter snake *Thamnophis sirtalis*. *Behav. Ecol. Sociobiol.* **25**: 269–274.

Slatkin, M. & Excoffier, L. 1996. Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity* **76**: 377–383.

Sparkman, A.M., Arnold, S.J. & Bronikowski, A.M. 2007. An empirical test of evolutionary theories for reproductive senescence and reproductive effort in the garter snake *Thamnophis elegans*. *Proc. Roy. Soc. Lond. B.* **274**: 943–950.

Spitze, K. 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozyme variation. *Genetics* **135**: 367–374.

Turesson, G. 1922. The genotypical response of the plant species to the habitat. *Hereditas* **3**: 211–350.

Voris, H.K. 1975. Dermal scale-vertebra relationships in sea snakes (Hydrophidae). *Copeia* **1975**: 746–755.

Waldmann, P. & Andersson, S. 1998. Comparison of quantitative genetic variation and allozyme diversity within and between populations of *Scabiosa canescens* and *S. columbaria*. *Heredity* **81**: 79–86.

Wong, A., Smith, M.L. & Forbes, M.R. 2003. Differentiation between subpopulations of a polychromatic damselfly with respect to morph frequencies, but not neutral genetic markers. *Mol. Ecol.* **12**: 3505–3513.

Wright, S. 1943. An analysis of local variability of flower colour in *Linanthus parryae*. *Genetics* **28**: 139–156.

Supplementary material

The following supplementary material is available for this article:

Appendix S1 Means and standard deviations of phenotypic traits for males and females in each population.

Appendix S2 Culmen impressions on the ventral surfaces of *Thamnophis elegans* from the Gallatin population. (a) Brewer's Blackbird culmen impression. (b) Great Blue Heron culmen impression. Scale units at bottom of each photograph are mm. Arrows indicate lateral edges of culmen imprint.

This material is available as part of the online article from http://www.blackwell-synergy.com/doi/abs/10.1111/j.1420-9101.2007.01401.x

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.