Adaptive divergence between ecotypes of the terrestrial garter snake, *Thamnophis elegans*

MOLLIE K. MANIER, CRAIG M. SEYLER, & STEVAN J. ARNOLD

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

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

Short running title: Selection analyses of garter snake ecotypes

Abstract

Because most F_{ST}-Q_{ST} studies are not accompanied by direct analyses of selection on phenotypic traits, diagnosed modes of selection lack confirmation, and the agents of selection are usually not identified. We analysed phenotypic selection in ecotypes of the terrestrial garter snake (*Thamnophis elegans*) using both F_{ST}-Q_{ST} analysis and correlational analysis of selection, as well as indirect and direct observations of selective agents. Differentiation between ecotypes in both scalation (Q_{CT} = 0.18) and coloration (Q_{CT} = 0.21) was about an order of magnitude stronger than differentiation at microsatellite loci (F_{CT} = 0.02). These results suggest that selection has driven ecotypic differentiation, despite the presence of moderate gene flow. Correlational analyses of selection confirm the F_{ST}-Q_{ST} results by revealing stabilising and correlational selection on both scalation and coloration traits. Direct observations of predation and culmen imprints on wild-caught snakes implicate birds as important factors driving adaptive divergence in our study system.

Keywords:

coloration;

correlational selection analysis

ecotypic variation;

F_{ST};

garter snake;

microsatellites;

Q-statistics;

Q_{ST};

scalation;

selection.

Introduction

Adaptive divergence is often a compromise between 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 struggle to adapt to dramatically different selective pressures in the face 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 three techniques that are often used individually but never, to our knowledge, in combination: F_{ST}-Q_{ST} comparison, correlation analysis of selection within a population, and direct and indirect observations of selective agents. Employing these techniques simultaneously in a single study system illuminates their strengths and weaknesses.

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). This approach has been effectively applied to a wide range of taxa, and *a priori* expectations of the F_{ST}-Q_{ST} relationship have been used to both predict modes of selection acting on quantitative traits (Palo et al., 2003; Cano et al., 2004) and to test specific hypotheses about local selection (Baker, 1992; Waldmann & Andersson, 1998; Gomez-Mestre & Tejedo, 2004). Despite these successes, some conspicuous limitations of the F_{ST}-Q_{ST} approach remain: (1) the approach as currently pursued is univariate and does not account for phenotypic or genetic correlations among traits. (2) Because molecular markers only inform on historical patterns of differentiation, contemporary levels of selection cannot be inferred. (3) Modes of selection are only roughly diagnosed and cannot be readily related to concepts of selection surfaces or

adaptive landscapes. (4) The actual agents of selection are not identified. These limitations are rarely accounted for in studies investigating quantitative trait divergence.

The present report addresses some of these limitations by using a combination of techniques to study the microevolutionary processes driving population divergence in closely proximate garter snake populations. We use F_{ST}-Q_{ST} comparison to identify traits involved in the adaptive differentiation of ecotypes. We contrast the utility of F_{ST}-Q_{ST} with the ability of correlational analysis of selection (Lande & Arnold, 1983) to diagnose modes of selection within populations. Finally, we address the inability of both of these techniques to identify selection agents by making direct and indirect observations of predation events.

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 km and in the face of moderate gene flow (Bronikowski & Arnold, 1999; Manier & Arnold 2005). 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. Lakeshore 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 coloration that appear to increase crypticity in rocky lakeshore and grassy meadow environments. To visual predators such as humans, the muted colours of lakeshore snakes (dull yellow or tan stripes on a gray background

colour) tend to match the rocky substrate of the lakeshore, while the meadow snake colour pattern (yellow or orange stripes on a black background) closely resembles dead rushes that litter the shallow meadow substrates. The difference in coloration between lakeshore and meadow ecotypes may be a result of differential selection for crypticity (Kephart, 1981).

We also examined six scale counts for adaptive differences between *T. elegans* ecotypes. Vertebral number (measured using ventral and subcaudal scale counts) can vary between different habitats as a function of a snake's ability to utilise 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, 1982; Kephart & Arnold, 1982), we expect selection for ability to swallow larger prey in lakeshore populations and hence higher scale counts. Both scale counts and coloration have been shown to be under selection in these and other populations of snakes (Arnold, 1988;

Arnold & Bennett, 1988; Brodie, 1992; King, 1993; Lindell et al., 1993), making these traits good candidates for our study.

We used estimates of neutral divergence at microsatellite loci to determine whether coloration 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, while $Q_{ST} < F_{ST}$ suggests stabilizing selection towards the same optimum in different populations (Lande, 1992; Spitze, 1993). Based on our hypothesis of ecotypic differentiation, we expect Q_{ST} estimates to far exceed F_{ST} in ventral scale counts and coloration. We used parallel three-level analyses of variance for microsatellite alleles (to generate *F*-statistics; Excoffier et al., 1992) and for quantitative traits (to generate *Q*-statistics) to test for adaptive differentiation at three different levels: between ecotypes, among populations within ecotypes (local adaptation) and among populations overall (regardless of ecotype). In order to address the F_{ST}-Q_{ST} limitations described above, we also used correlational selection analysis (Lande & Arnold, 1983) to visualize contemporary selection on coloration and scalation in one of our populations and compare these results to the F_{ST}-Q_{ST} analysis. Both kinds of analyses support the hypotheses of adaptive differentiation between ecotypes and local adaptation within ecotypes. Finally, we use direct observations of predation and analysis of culmen impressions on snakes themselves to suggest that avian predators may be the selective agents responsible for this adaptive differentiation.

Materials and Methods

Study sites

Phenotypic and genotypic data were collected from six populations (Table 1) at and around the southeast 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 km long that are uninhabited by *T. elegans*. Distances between populations within an ecotype also ranged from 1.3 to 19.9, while between-ecotype distances ranged from 3.8 to 19.5. Snakes were found in the open and by lifting cover objects and were collected by hand.

Measurement of quantitative traits

Coloration traits were scored on 1268 live *T. elegans*. Population samples ranged from 50–745 individuals. The sample of 745 individuals for Gallatin corresponds to the total sample of captured individuals described in the section on correlational analysis of selection. We scored coloration 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, as 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 www.pantone.com). For statistical analyses, we translated the Pantone colour codes into two different 3-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 gray 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 coloration-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 less than 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 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.

Inheritance of quantitative traits

For coloration traits, heritabilities were assessed using a sample of 325 individuals representing 35 sibships from the Pikes and Wildcat populations. Wildcat is a lakeshore site not included in the selection analyses (see Fig. 1). All coloration scores were made on neonates, no more than one month after birth. Coloration traits appeared to be fully expressed at birth, and no obvious ontogenetic trends were observed when individuals were reared to maturity in the laboratory. Heritabilities (h^2) 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 & McKay, 1996) rather than mother-offspring regression was used to estimate heritabilities, because coloration 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). Estimates of heritability can be inflated by maternal effects (Falconer & McKay, 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 & McKay, 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 inland male and female heritabilities, shown in Table 6 of Arnold & Phillips (1999), estimated from the inland genetic and phenotypic variances 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 sulfate 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 (labeled 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 sec, appropriate annealing temperature for 30 sec and 72 °C for 30 sec, ending with 72°C for 2 min. PCR products were separated using an ABI 3100 capillary electrophoresis genetic analyser and data were visualised using Genotyper 3.7 (ABI Prism).

Population genetic analysis of the microsatellite data is described in Manier & Arnold (2005). 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). 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).

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 corrections for multiple comparisons (Rice, 1989). For these traits the difference between the sexes ranged from 0.2 to 0.4 standard deviations for the colorations 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 genetic variance to estimate variance components and characterise the population structure of both microsatellite and phenotypic traits. For both kinds of traits, variation was partitioned into within-population, among-populations (within ecotypes) and among 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 among 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 among-ecotype and the among-population (within ecotype) components. $F_{CT} = V_a/V$ is the proportion of the total variance that resides among ecotypes. $F_{SC} = V_b /(V_b+V_c)$ is the proportion of the total variance within ecotypes that resides among populations. Standard errors of each *F*-statistic were 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.

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, SAS Institute 2002). 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 among-ecotype mean square over the among-population (within ecotypes) 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 corresponding heritability for that trait. Thus, $V_c = h^2V_w$. The among-ecotype and among-population components of variance were equated with the

corresponding genetic components of variance. To produce results parallel 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 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., due to 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).

Correlational analysis of selection

Survival and coloration traits were assessed in the course of a mark-recapture study at the lakeshore site Gallatin over a five year period. Out of a total of 745 individuals captured along a 2.2 km stretch of shoreline at this site, 141 were recaptured one or more times. Of these, 125 expressed colour in both dorsal and lateral stripes (16 did not express lateral stripes). For each of these snakes, we assessed survival by counting the days that elapsed between the first capture and the last recapture. We subtracted days of inactivity from late summer until early spring (16 August – 14 April) when the snakes were generally inactive or hibernating and so not exposed to predation. Coloration was scored as previously described. Because of the relatively small sample size for each sex, the sexes were pooled for statistical analyses.

We used regression models to measure the intensity of viability selection on coloration traits. Because of the large number of traits (six for each colour coding scheme, plus BKGRD) and the relatively small sample of snakes (N = 125), we tested specific hypotheses that involved only subsets of traits, rather than the simultaneous influence of all possible traits on survival. In light of the general lack of phenotypic correlations between traits within each coding scheme, we first computed linear and quadratic regressions of survival on each trait (e.g., DORRED). Prior to analysis, survival was standardised so that its mean was one and the traits were standardised so that they had zero means and unit variances (Lande & Arnold, 1983). Second, because of observed phenotypic correlations between dorsal and lateral coloration, we tested the hypothesis that coordination between dorsal and lateral stripe colours enhanced survival. For example, to test the hypothesis that selection favored coordination between dorsal and lateral red, we used the following quadratic regression model:

$$survival = 1.0 + \beta_1 z_1 + \beta_2 z_2 + \tfrac{1}{2}\gamma_{11} z_1^{\,2} + \tfrac{1}{2}\gamma_{22} z_2^{\,2} + \gamma_{12} z_1 z_2, \qquad (1)$$

where z_1 is DORRED and z_2 is LATRED, β_1 and β_2 are coefficients of directional selection, γ_{11} and γ_{22} are coefficients of stabilizing selection, and γ_{12} is a coefficient of correlational selection (Lande & Arnold, 1983).

Identification of avian predators

A sample of 46 live *T. elegans* was captured over a three 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 thirty-year period and compared them with culmen impressions on the snakes. The candidate avian predators were Great Blue Heron (*Ardea herodias*), California Gull (*Larus californicus*), American Robin (*Turdus migratorius)*, and Brewer's Blackbird (*Euphagus cyanocephalus*).

Results

Comparison of dorsal and lateral stripes and inheritance of coloration

Dorsal stripes tended to have higher values for saturation than lateral stripes ($P = 0.04$), but dorsal and lateral stripes did not significantly differ in redness. Very few pairs of traits showed significant phenotypic correlations ($P > 0.05$). The few pairs showing significant phenotypic correlations were generally the same trait expressed in dorsal and lateral stripes (e.g., LATRED vs. DORRED), and in such cases the correlations ranged from 0.35-0.77 (e.g., $r = 0.36$, $P < 0.01$ for red, $r = 0.37$, $P < 0.01$ for saturation; N = 125).

Heritability estimates for coloration traits ranged from 0.14-0.80 in females and from 0.01-0.63 in males (Table 2). Sexual averages ranged from 0.08-0.65. 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 coloration (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.

2, 3) suggest relative uniformity among meadow populations, as well as a striking difference in average coloration 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 coloration. Background colour is darker in meadow populations than in lakeshore populations, with lakeshore populations also showing more variation in background colour (Figure 4).

Turning to scalation, the most striking trend apparent in tabulations of population and ecotypic means (Appendix Table 1, 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 male was 0.5 phenotypic standard deviations (SD) for scalation traits and 0.6 SD for coloration 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 coloration 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).

In contrast to results for phenotypic traits, the ANOVA for microsatellite data 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) among ecotypes. The average percentages for males and females for 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 coloration 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 (versus 4% for microsatellites), and the majority of among-population variation resided among 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 *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 4% (± 1% SE) of total variation. Although for individual loci, F_{ST} ranged from only 0.4-6%, this degree of differentiation was statistically significant ($P < 0.01$) 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 2% (± 1% 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 1% (± 0.4% SE), but six out of nine loci showing significant values ($P < 0.05$). In contrast, average population differentiation was seven times greater (Q_{ST}/F_{ST}) for scalation traits and six times greater

for coloration traits. In the case of scalation traits, Q_{ST} ranged from 9–44% and averaged 28% (± 5% SE). In the case of coloration traits, Q_{ST} ranged from 2–71% and averaged 25% (± 5% SE). These statistics identified the scalation traits ILAB, VENT, and SUB and the coloration 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} (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 among 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, while Q_{ST} was 28% for POST, this scalation trait showed no significant differentiation among ecotypes. Among coloration 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 10% (± 5% SE) for scalation traits, and 4% (± 1% SE) for coloration traits. While 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.

Selection on coloration

None of the simple linear or quadratic models relating individual traits to survival yielded statistically significant results. For both colour coding schemes and BKGRD, R^2 was always less than 0.5% for linear models and less than 1% for quadratic models. In contrast, when dorsal and lateral stripe colours were considered pairwise, we found that bivariate stabilizing and correlational selection favored coordination of stripe colours. However, quadratic models of colour coordination yielded statistically significant results only for the cases of dorsal versus lateral red and dorsal versus lateral saturation. In both cases the selection surface corresponded to a positively-inclined ridge (Fig. 5). The effect of selection is to favor intermediate values of stripe colours and to increase the correlation between dorsal and lateral stripe colours.

Birds are probably the selective agents responsible for the observed patterns of selection on stripe colours. The photographs of culmen impressions on seven snakes matched the culmens of two avian predators. Six out of the seven snakes had impressions from Great Blue Herons (*Ardea herodias*), and one of the six had impressions from a Brewer's Blackbird (*Euphagus cyanocephalus*; Appendix Fig. 1). In two instances, birds were directly observed preying on garter snakes near our study site. A Great Blue Heron was observed on the Eagle Lake shoreline, 5 km SW of Gallatin, attacking and then flying off with a large gravid *T. elegans* (C. Cox, pers. comm.). 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.

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). While vertebrate biologists have been relatively slow to use this concept, the existence of ecotypes in sticklebacks (*Gasterosteus*) has been thoroughly documented (Schluter et al., 2004; Baker et al., 2005). 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 coloration 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 (Camin & Ehrlich, 1958; King, 1993a,b; Lawson & King, 1996; Bittner & King, 2003) in which strong selection for crypticity maintains population differentiation in coloration in both species, despite considerable gene flow (Gibson & Falls, 1979; King, 1993b; King & Lawson, 2001; Bittner & King, 2003). These examples, together with our results, suggest that ecotypic variation and other kinds of local adaptation may be relatively common in vertebrates.

Ecotypic variation and local adaptation in *T. elegans*

Differentiation in coloration and scalation traits revealed in the present analyses coincides with ecotypic differences in life history (Bronikowski & Arnold, 1999). Thus, the syndrome of differentiation between ecotypes includes scalation, vertebral numbers, numerous aspects of coloration, as well as growth rate and body size-fecundity relationships. Remarkably, this extensive differentiation occurs over a distance of a few km, between populations connected by moderate gene flow.

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 vertebrate 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 favors 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, 1982; Kephart & Arnold, 1982). Feeding performance trials are needed to determine whether the observed differences in scalation do indeed enhance the ability of lakeshore snakes to eat large fish.

The F_{ST}-Q_{ST} and F_{CT}-Q_{CT} contrasts suggest that diversifying selection is responsible for population differentiation in scalation and coloration characters. For many of the scalation traits and several of the coloration traits population differentiation is five to 10-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 coloration. 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 coloration traits is on the average 10 and four-fold, respectively, more pronounced than what we would expect by drift. Most of the scalation and coloration traits show highly significant Q_{SC}-values, in contrast to F_{SC}- or to Q_{CT}-values. While 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 coloration. Future studies might determine whether subtle differences among meadow sites in vegetation contribute to selection on coloration.

The F_{ST}-Q_{ST} analysis confirms the results of other correlational selection analyses on scalation and coloration traits. 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, while 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 VENT and SUB were not statistically significant. In the present study, bivariate ridges were also revealed by an analysis of selection on coloration (Fig. 2), suggesting that this may be a common mode of selection on both vertebral numbers and coloration traits of garter snakes.

While the F_{ST}-Q_{ST} analysis revealed a statistically significant history of selection on many aspects of scalation and coloration, correlation analyses within populations showed no instances of significant directional selection and only a few instances of significant multivariate stabilizing selection. The absence of directional selection in the correlational analyses may mean that population means are so close to their optima that directional selection is weak or nonexistent. In contrast, the F_{ST}-Q_{ST} analysis succeeded in detecting a history of directional selection that may have arisen from displacement of intermediate optima. Furthermore, the F_{ST}-Q_{ST} analysis may have been more generally successful in detecting selection, because it integrated the effects of selection over the many generations that may have been required for ecotypic differentiation, rather than over a single generation, as in the correlational analysis.

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 coloration 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 below.

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 remarkable 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), primarily unidirectionally from the major source population, Papoose Meadows (Manier & Arnold, 2005). Gallatin and Papoose are only four km apart and are connected by an intermittent stream, Papoose Creek, a known dispersal corridor for *T. elegans* in wet years (Arnold, unpubl. data). Thus, Gallatin and other lakeshore sites have been able to differentiate in the face of persistent migration from Papoose Meadows, and both lakeshore and meadow populations have been able to maintain their ecotypic identities on a very small spatial scale.

Endler (1990) has persuasively argued that spectrographic 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 and by using 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), though the differences may not be huge (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 coloration 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.

Conclusions

As in many other studies (McKay & Latta, 2002; Wong et al., 2003), Q_{ST} greatly exceeds F_{ST} in our study system. This result implies that diversifying selection acting on phenotypic traits has produced a departure 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 reported here and in Arnold (1988) undoubtedly suffered from this problem. In contrast to the F_{ST}-Q_{ST} results, correlational analysis was only able to detect selection on a few traits, presumably the ones experiencing the strongest selection. Nevertheless, despite that modest success, our correlational analyses were able to diagnose modes of selection that could be related to features of the adaptive landscape (Arnold et al., 2001).

Many authors have remarked that correlational analyses do not identify the actual agents of selection. The same limitation plagues F_{ST}-Q_{ST} comparisons. Even a few observations of selective events, as in the present study, can remedy this situation and so direct and illuminate the interpretation of selection analyses. Without such observations, both kinds of analyses stop short of a causal interpretation. Thus, the overall message of our study is that multiple lines of inquiry into the nature and consequences of selection can yield a synthetic overview of the process that no one technique can provide.

Acknowledgements

We are grateful to many Eagle Lake field workers, especially L. Houck. M. Blouin, A. R. Kiester, R. Mauricio, P. McEvoy and anonymous reviewers provided helpful comments on the manuscript. We especially thank S. R. Estes and M. E. Pfrender for encouraging us in this research. We were supported by an EPA STAR Fellowship (U-91552801-5), an NSF GK-12 Teaching Fellowship, and an NSF DDIG (DEB-0309017) to MKM, as well as an NSF grant (DEB-9903934) to SJA and M. E. Pfrender and an NSF grant (DEB-0323379) to SJA and A. M. Bronikowski. CMS was supported by an REU supplement to NSF grant DEB-0323379. The specimens and tissue samples used in this study we collected under the auspices of Scientific Collecting Permits granted by the California State Department of Fish and Wildlife and protocols approved by the IACUC committee at Oregon State University.

Literature Cited

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, Massachusetts.

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. and 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.

Arnold, S. J., Pfrender, M. E. & Jones, A. G. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* **112/113:** 9-32.

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.

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., III. 1992. Correlational selection for color 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.

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 ed.* Prentice Hall, Harlow, U.K.

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

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.

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

Gibson, A. R. & Falls, J. B. 1979. Thermal biology of the common garter snake *Thamnophis sirtalis* (L.) II. The effects of melanism. *Oecologia* **43:** 99-109.

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.

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

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

Jacobs, G. H. 1981. *Comparative Color 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. and A. F. Bennett. 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. 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*. Ph.D. dissertation, Dept. of Biology, University of 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. Color pattern variation in Lake Erie water snakes: inheritance. *Can. J. Zool.* **71:** 1985-1990.

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

King, R. B. & Lawson, R. 2001. 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. 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 .

Manly, B. F. 1997. *Randomization and Monte Carlo Methods in Biology, 2nd ed.* 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, Mass.

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

Munsell Color Company. 1976. *Munsell Book of Color, 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.* Univ. of Texas at Arlington. Software available at www.uoregon.edu/~pphil/software.html.

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.

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

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

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.

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

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

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.

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

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.

Figure Legends

Fig. 1 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.

Fig. 2 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.

Fig. 3 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.

Fig. 4 Frequency histograms for background colour. Gray shades are, left to right, 3, 4, 5, 6, and 7. Other conventions as in FIG. 2.

Fig. 5 Bivariate portrayals of selection on coloration traits. Lighter shading corresponds to higher fitness. Some statistically insignificant terms (P>0.05) were dropped from the full model, given in eq. (1). (a). Bivariate selection on DORRED and LATRED in the Gallatin population. The surface portrays the equation $w = 1.0 + \beta_2 z_2 + \frac{1}{2}\gamma_{11} z_1^2 + \frac{1}{2}\gamma_{22} z_2^2 z + \gamma_{12} z_1 z_2$, where w is relative survival (days), $z_1 =$

DORRED, z_2 = LATRED, β_2 = –885.2 ± 400.4 SE (P = 0.03), γ_{11} = –1843.0 ± 849.4 SE (P = 0.03), γ_{22} = –534.4 ± 635.0 SE (P = 0.40), and γ_{12} = 1694.8 ± 795.3 SE (P = 0.04). (b) Bivariate selection on DORSAT and LATSAT the Gallatin population. The surface portrays the equation $w = 1.0 + \frac{1}{2}\gamma_{11}z_1^2 + \frac{1}{2}\gamma_{22}z_2^2 z + \gamma_{12}z_1 z_2$, where w is relative survival (days), z_1 = DORSAT, z_2 = LATSAT, γ_{11} = -393.8 ± 180.4 SE (P = 0.03), γ_{22} = -471.6 ± 234.2 SE (P = 0.05), and γ_{12} = 345.3 ± 171.6 SE (P = 0.05).

Appendix Fig. 1 Culmen impressions on the ventral surfaces of *T. 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.

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 coloration 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. Number of families indicated by N.

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

	Lakeshore				Meadow				SD		Divergence	
	Male		Female		Male		Female		Male	Female	Male	Female
Trait	N	Mean	N	Mean	N	Mean	N	Mean				
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
Coloration												
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 denotes sample size.

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

Microsat. locus	F_{SC}	*P*	F_{ST}	*P*	F_{CT}	*P*
1	0.02	**	0.02	**	-0.00	ns
2	0.00	ns	0.06	****	0.05	ns
3	0.01	**	0.04	****	0.03	ns
4	-0.01	ns	0.01	ns	0.02	ns
5	0.01	ns	0.00	ns	-0.00	ns
6	0.03	**	0.06	****	0.04	ns
7	0.01	**	0.05	****	0.04	ns
8	0.01	*	0.05	****	0.04	ns
9	0.02	**	0.03	**	0.01	ns
Average	0.01		0.04		0.02	
SE	0.004		0.01		0.01	
Scalation traits	Q_{SC}	*P*	Q_{ST}	*P*	Q_{CT}	*P*
VENT	0.08	****,****	0.31	****,****	0.24	**,ns
SUB	0.16	****,****	0.35	****,****	0.23	*,ns
MID	0.06	ns,****	0.21	****,****	0.15	***,ns
ILAB	0.03	**,****	0.44	****,****	0.42	*,***
SLAB	-0.02	ns,ns	0.09	*,**	0.11	ns,*
POST	0.31	****,****	0.28	*,*	-0.05	ns,ns
Average	0.10		0.28		0.18	
SE	0.05		0.05		0.07	
Coloration traits	Q_{SC}	*P*	Q_{ST}	*P*	Q_{CT}	*P*
DORRED	0.04	****,***	0.13	*,ns	0.09	*,ns
DORGRN	0.05	****,****	0.16	*,ns	0.12	*,ns
DORBLU	0.03	****,****	0.39	**,**	0.37	**,**
LATRED	0.02	****,***	0.39	**,ns	0.36	**,ns
LATGRN	0.12	*,****	0.15	ns,ns	0.04	ns,ns
LATBLU	0.02	****,**	0.13	*,ns	0.10	*,ns
DORHUE	0.02	*,ns	0.02	ns,ns	0.01	ns,ns
DORSAT	0.04	****,****	0.23	*,*	0.20	*,*
DORLT	0.03	****,****	0.26	**,*	0.24	*,*
LATHUE	0.10	**,****	0.21	ns,ns	0.13	ns,ns
LATSAT	0.02	****,***	0.34	**,ns	0.32	**,ns
LATLT	0.01	**,ns	0.06	ns,ns	0.05	ns,ns
BKGRD	0.04	****,****	0.71	***,***	0.70	***,***
Average	0.04		0.25		0.21	
SE	0.01		0.05		0.05	

Q_{SC}, Q_{ST} and Q_{CT} columns show the average of male and female values, *P* columns for scalation and coloration traits show significance levels for separate analyses of females and males, in that order.

Appendix Table 1 Means and standard deviations of phenotypic traits for males and females in each population.

	McCoy*						Mahogany*						Nameless*					
	Males			Females			Males			Females			Males			Females		
Trait	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Scale counts																		
VENT	72	168.6	7.9	115	163.1	6.4	55	167.1	6.4	97	163.7	5.6	40	168.8	5.6	105	164.0	5.4
SUBC	68	82.9	9.5	99	78.9	8.2	48	78.8	10.6	77	76.1	8.1	26	78.1	12.9	46	75.7	7.8
TOTAL	68	251.4	12.5	98	241.5	9.8	48	245.3	15.0	77	238.9	10.6	26	245.2	16.3	46	238.4	11.0
MID	72	19.8	1.2	114	19.9	1.2	56	19.9	1.0	97	19.7	1.1	41	19.5	0.9	109	19.7	1.1
SLAB	72	15.9	0.5	114	16.0	0.6	55	16.0	0.4	100	16.0	0.4	39	15.9	0.4	108	16.0	0.2
ILAB	72	19.8	0.9	115	19.9	0.8	55	19.8	0.5	99	20.0	0.9	39	19.9	0.4	107	20.0	1.0
POST	71	5.9	0.7	114	5.9	0.8	56	6.0	0.4	99	6.1	0.6	39	6.0	0.4	107	6.1	0.2
Colour scores																		
DORRED	23	244.0	2.6	27	244.4	2.2	48	245.1	1.8	65	245.1	1.5	57	244.5	1.6	55	245.1	1.4
DORGREEN	23	191.9	19.0	27	184.8	15.6	48	185.9	14.7	65	182.6	12.7	57	179.4	11.6	55	181.7	11.7
DORBLUE	23	103.8	15.6	27	98.5	13.4	48	100.6	13.8	65	97.8	12.8	57	94.0	10.6	55	97.5	13.2
DORHUE	23	25.7	4.4	27	24.0	3.2	48	23.9	2.8	65	23.3	2.0	57	22.9	2.0	55	23.0	1.4
DORSAT	23	206.9	8.3	27	209.3	6.0	48	210.6	4.8	65	211.2	2.9	57	210.2	3.7	55	211.3	2.0
DORLT	23	163.5	7.5	27	161.1	6.6	48	162.4	6.9	65	161.1	6.6	57	159.0	5.4	55	160.9	6.8
LATRED	23	243.1	2.8	27	244.9	2.9	48	245.5	2.3	65	245.4	3.9	56	245.8	2.2	52	246.2	2.7
LATGREEN	23	217.3	8.7	27	207.7	12.1	48	203.2	13.8	65	197.5	9.4	56	201.7	12.2	52	194.9	7.3
LATBLUE	23	125.8	8.7	27	119.6	9.1	48	116.7	11.0	65	116.2	14.5	56	115.6	9.0	52	111.4	9.7
LATHUE	23	31.5	3.2	27	28.4	4.1	48	27.3	4.0	65	25.3	2.0	56	26.8	3.7	52	24.8	0.9
LATSAT	23	199.5	8.8	27	206.3	9.8	48	208.6	8.6	65	207.9	15.9	56	209.6	8.3	52	211.8	9.4
LATLT	23	173.7	4.7	27	171.4	4.2	48	170.3	5.1	65	169.8	6.0	56	169.8	4.0	52	167.9	4.5
BKGRND	23	6.5	0.5	27	6.3	0.4	48	6.2	0.4	65	6.2	0.4	57	6.0	0.3	55	6.0	0.4

* meadow site; † lakeshore site

Appendix Table 1 Continued.

	Papoose*						Gallatin†						Pikes†					
	Males			Females			Males			Females			Males			Females		
Trait	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Scale counts																		
VENT	62	169.7	5.0	111	165.8	5.5	367	170.3	4.9	373	170.1	4.4	336	173.7	5.2	429	168.9	4.6
SUBC	55	80.9	9.4	93	72.8	11.5	307	84.1	8.6	317	84.0	8.4	330	90.9	5.6	403	80.8	4.3
TOTAL	55	250.7	10.5	93	239.2	12.3	304	254.5	11.5	315	254.0	10.7	323	264.4	7.8	394	249.8	7.0
MID	62	19.4	0.9	110	19.8	1.2	384	20.5	0.8	401	20.4	0.9	343	20.0	1.0	442	20.3	1.0
SLAB	62	16.0	0.4	110	15.9	0.4	386	16.0	0.3	405	16.0	0.4	345	16.0	0.6	442	16.1	0.8
ILAB	62	19.9	0.7	110	19.8	0.8	386	20.3	0.7	405	20.2	0.6	343	20.2	0.8	441	20.4	0.9
POST	61	6.0	0.4	109	6.1	0.5	386	6.2	0.7	405	6.2	0.6	345	5.9	0.7	443	6.0	0.7
Colour scores																		
DORRED	35	243.6	3.2	57	245.2	2.1	356	243.5	5.2	375	243.3	5.3	77	241.4	6.2	79	241.1	6.0
DORGREEN	35	188.3	17.6	57	188.3	13.6	356	195.9	13.8	375	192.0	12.2	77	192.6	9.4	79	190.4	11.2
DORBLUE	35	100.6	14.7	57	104.6	16.6	356	114.7	17.8	375	113.7	19.7	77	112.3	14.6	79	113.2	19.2
DORHUE	35	24.9	3.8	57	23.9	3.2	356	25.2	3.7	375	24.1	2.9	77	24.8	2.0	79	24.2	2.2
DORSAT	35	206.7	9.3	57	210.2	6.1	356	203.4	15.7	375	202.9	16.0	77	198.4	17.9	79	196.9	18.3
DORLT	35	161.7	7.1	57	164.3	8.2	356	168.4	8.8	375	167.8	9.7	77	166.2	7.5	79	166.5	9.4
LATRED	33	244.4	3.2	52	245.4	3.3	305	243.4	5.3	324	243.2	5.4	70	242.7	5.6	69	242.6	5.6
LATGREEN	33	207.4	15.1	52	200.8	11.5	305	201.9	11.9	324	198.7	12.6	70	200.2	8.5	69	196.4	10.0
LATBLUE	33	120.5	13.4	52	119.4	14.4	305	125.2	18.0	324	125.4	21.8	70	124.3	19.1	69	124.1	20.2
LATHUE	33	28.6	4.4	52	26.0	3.3	305	25.8	3.8	324	24.8	3.5	70	25.5	3.2	69	24.4	2.8
LATSAT	33	203.9	12.7	52	207.4	14.1	305	200.1	19.3	324	198.7	21.2	70	197.4	22.0	69	197.7	20.5
LATLT	33	171.6	5.8	52	171.5	6.1	305	173.3	8.2	324	173.2	9.7	70	172.5	8.1	69	172.5	9.3
BKGRND	35	6.1	0.4	56	6.1	0.5	363	4.8	0.7	382	4.7	0.6	77	4.6	0.5	79	4.5	0.6

* meadow site; † lakeshore site

Fig. 1

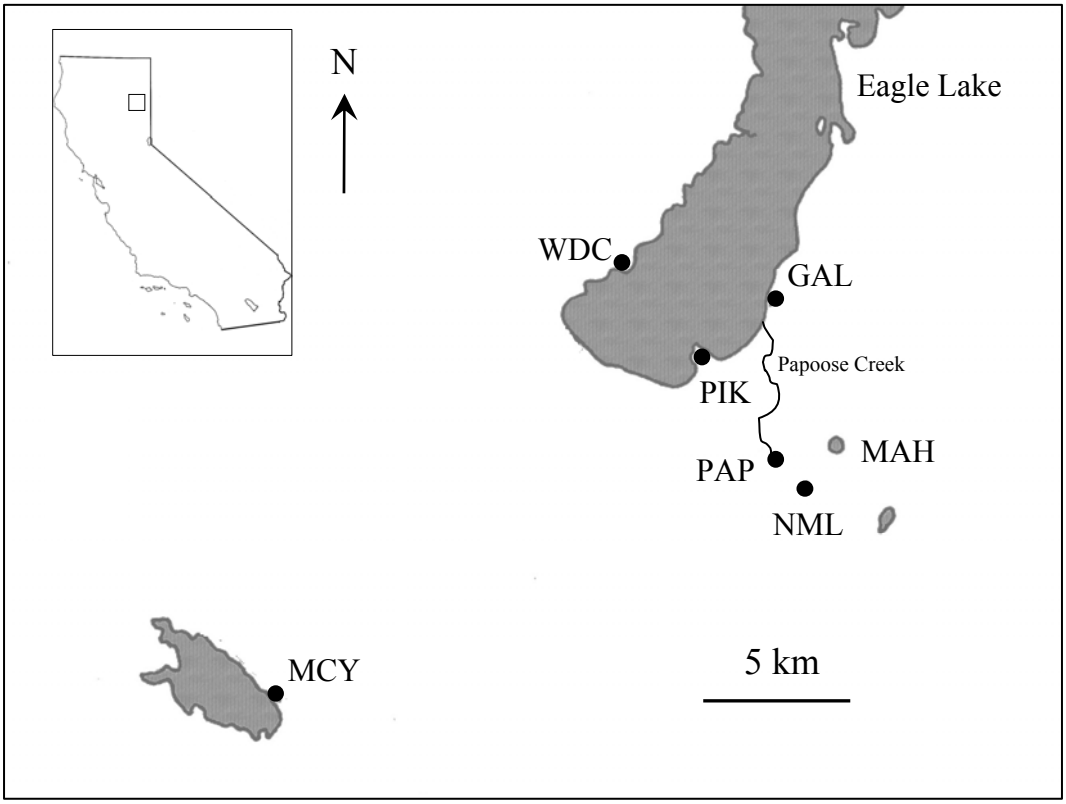


Fig. 2

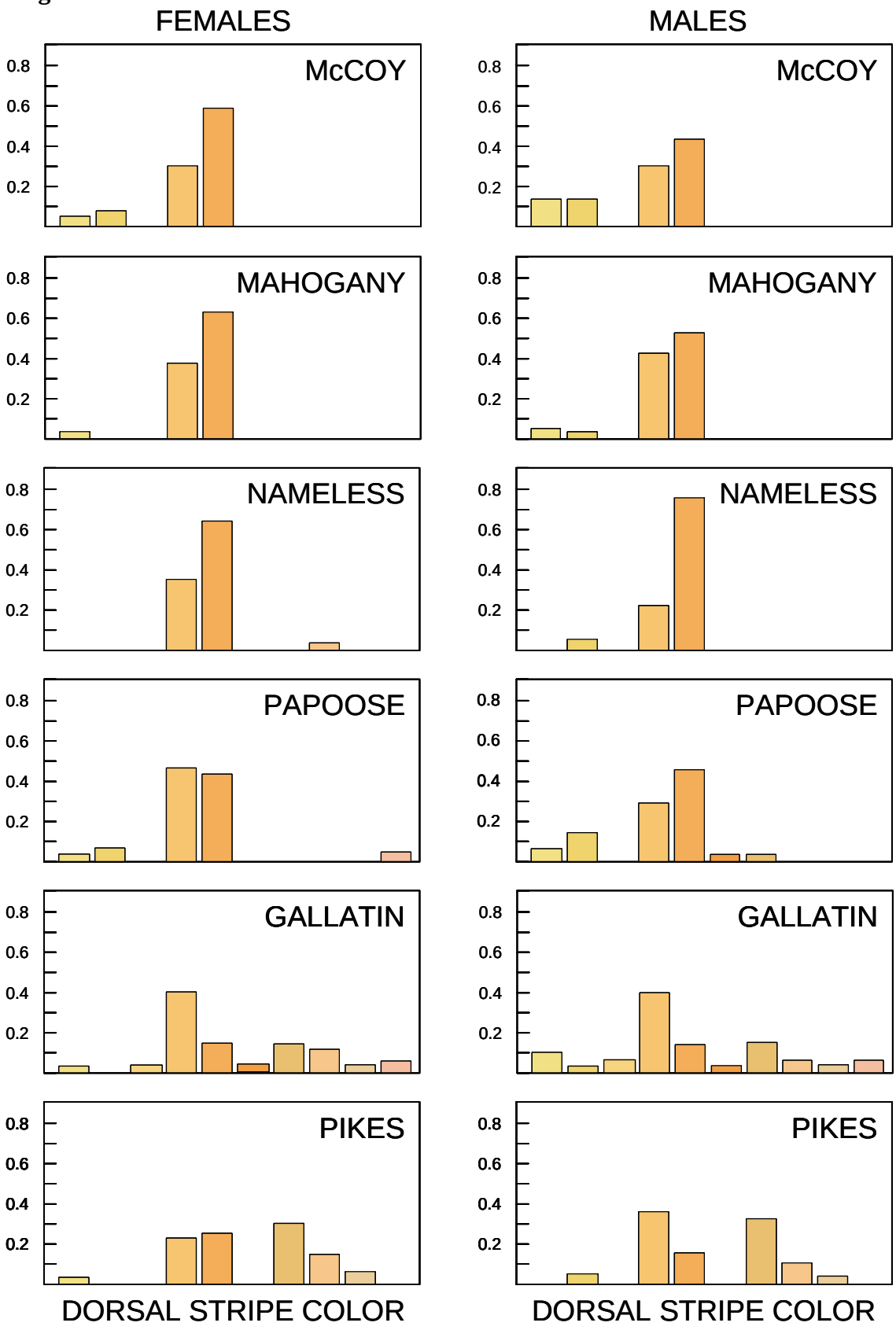

FEMALES
MALES
McCOY
MAHOGANY
NAMELESS
PAPOOSE
GALLATIN
PIKES
0.8
0.6
0.4
0.2
DORSAL STRIPE COLOR
DORSAL STRIPE COLOR

Fig. 3

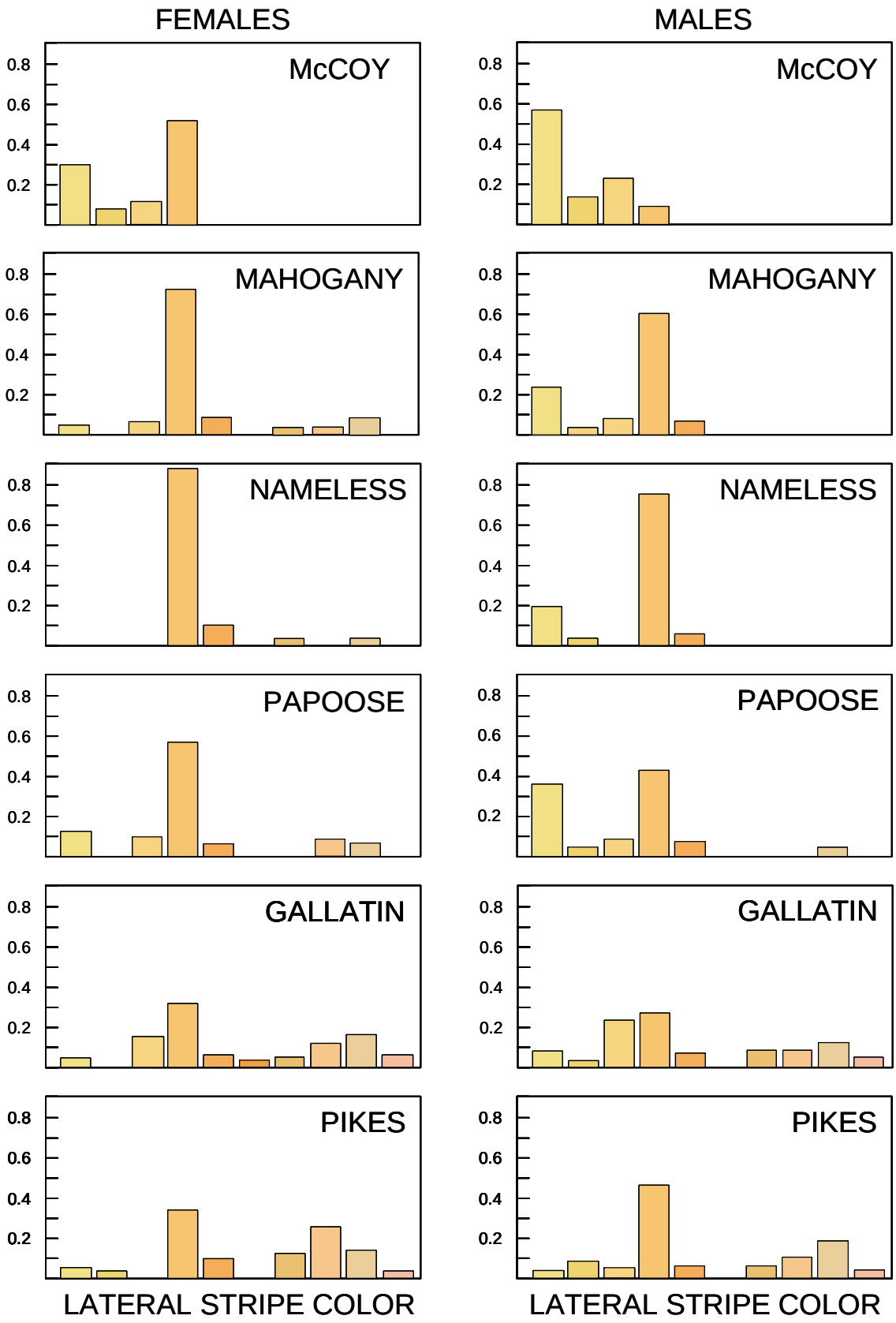

FEMALES
MALES
McCOY
MAHOGANY
NAMELESS
PAPOOSE
GALLATIN
PIKES
0.8
0.6
0.4
0.2
LATERAL STRIPE COLOR
LATERAL STRIPE COLOR

Fig. 4

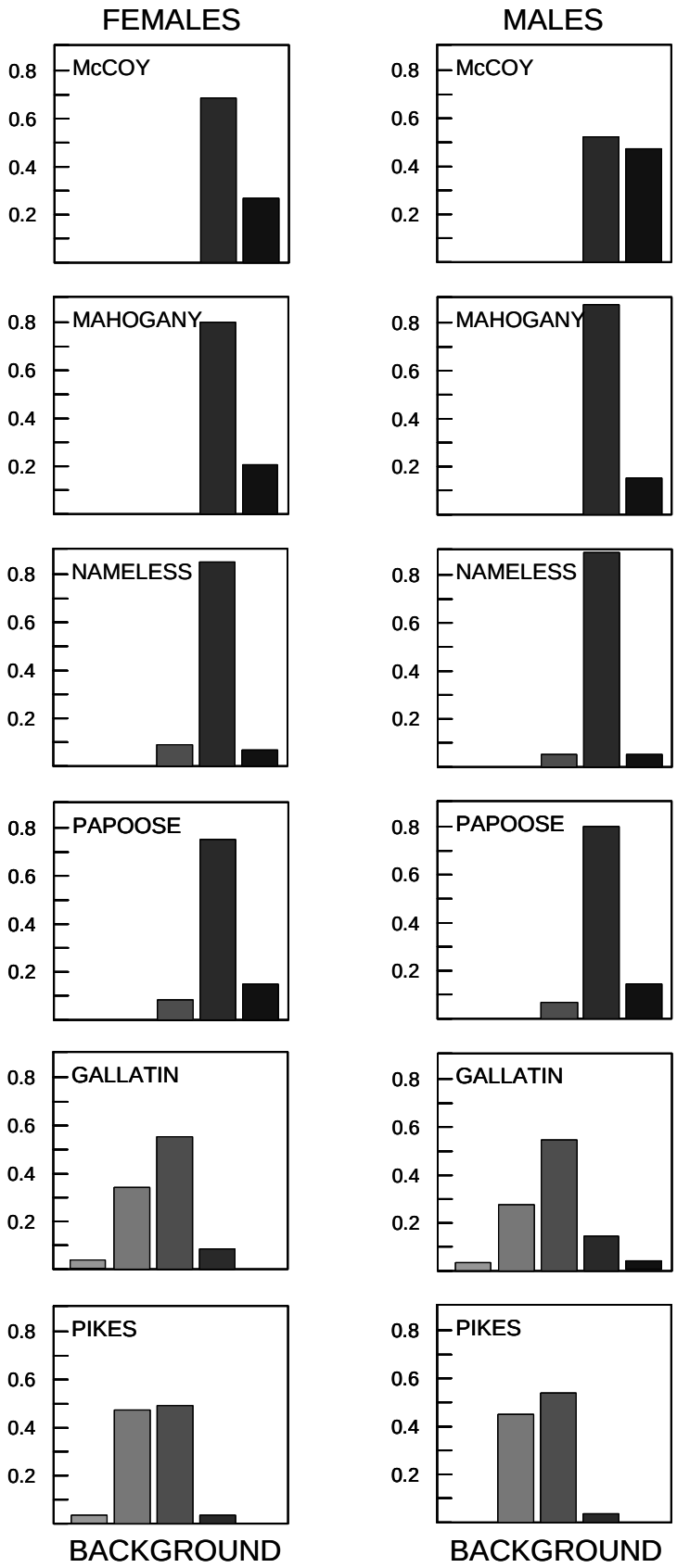
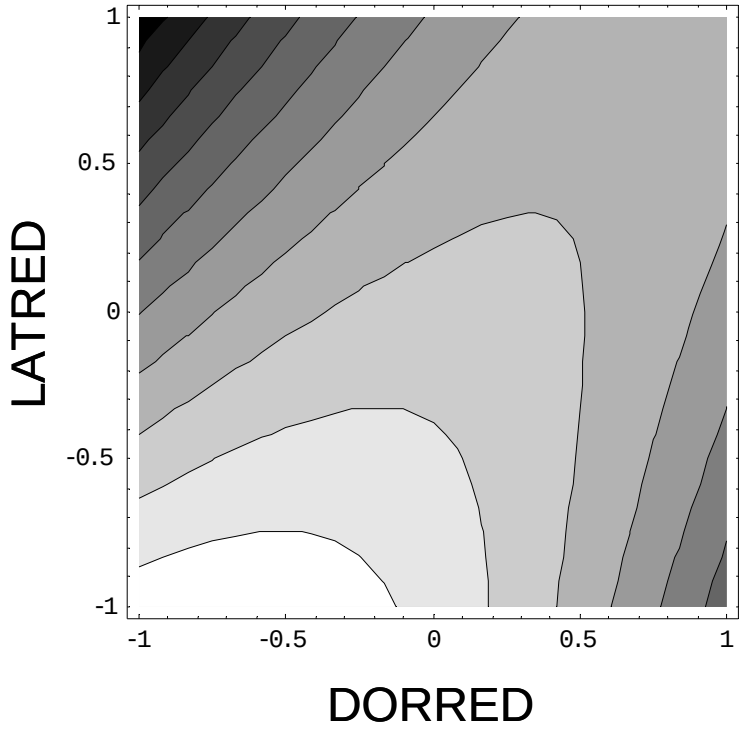
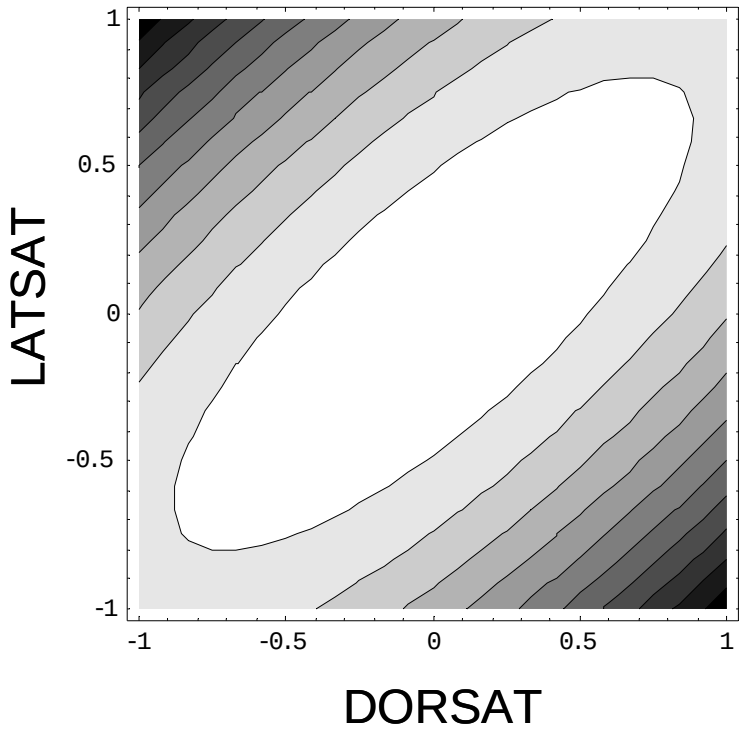

FEMALES
MALES
McCOY
MAHOGANY
NAMELESS
PAPOOSE
GALLATIN
PIKES
0.8
0.6
0.4
0.2
BACKGROUND
BACKGROUND

Fig. 5 (a)


LATRED
DORRED
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1

(b)


LATSAT
DORSAT
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1

Appendix Fig. 1

(a)

(b)