Journal of Biogeography (J. Biogeogr.) (2010)

ORIGINAL ARTICLE

Evidence for selection on coloration in a Panamanian poison frog: a coalescent-based approach

Jason L. Brown[1]†, Martine E. Maan[2,3], Molly E. Cummings[2] and Kyle Summers[1]*

[1]Biology Department, East Carolina University, Greenville, NC 27858, USA, [2]Section of Integrative Biology, University of Texas, Austin, TX 78712, USA, [3]Institute of Ecology and Evolution, University of Bern, Switzerland; and Eawag Centre of Ecology, Evolution and Biogeochemistry, Kastanienbaum, Switzerland

*Correspondence: Kyle Summers, Howell Science N314, Biology Department, East Carolina University, Greenville, NC 27858, USA. E-mail: summersk@ecu.edu
†Present address: Biology Department, Duke University, Durham, NC 27705, USA.

ABSTRACT

Aim The strawberry poison frog, *Oophaga pumilio*, has undergone a remarkable radiation of colour morphs in the Bocas del Toro archipelago in Panama. This species shows extreme variation in colour and pattern between populations that have been geographically isolated for < 10,000 years. While previous research has suggested the involvement of divergent selection, to date no quantitative test has examined this hypothesis.

Location Bocas del Toro archipelago, Panama.

Methods We use a combination of population genetics, phylogeography and phenotypic analyses to test for divergent selection in coloration in *O. pumilio*. Tissue samples of 88 individuals from 15 distinct populations were collected. Using these data, we developed a gene tree using the mitochondrial DNA (mtDNA) d-loop region. Using parameters derived from our mtDNA phylogeny, we predicted the coalescence of a hypothetical nuclear gene underlying coloration. We collected spectral reflectance and body size measurements on 94 individuals from four of the populations and performed a quantitative analysis of phenotypic divergence.

Results The mtDNA d-loop tree revealed considerable polyphyly across populations. Coalescent reconstructions of gene trees within population trees revealed incomplete genotypic sorting among populations. The quantitative analysis of phenotypic divergence revealed complete lineage sorting by colour, but not by body size: populations showed non-overlapping variation in spectral reflectance measures of body coloration, while variation in body size did not separate populations. Simulations of the coalescent using parameter values derived from our empirical analyses demonstrated that the level of sorting among populations seen in colour cannot reasonably be attributed to drift.

Main conclusions These results imply that divergence in colour, but not body size, is occurring at a faster rate than expected under neutral processes. Our study provides the first quantitative support for the claim that strong diversifying selection underlies colour variation in the strawberry poison frog.

Keywords

Aposematism, Bocas del Toro, coalescence, coloration, *Dendrobates*, Dendrobatidae, mate choice, *Oophaga pumilio*, polymorphism, selection.

INTRODUCTION

The evolution and maintenance of colour polymorphisms within species has emerged as an important focus of attention within evolutionary biology (Gray & McKinnon, 2007). Research on such polymorphisms has enhanced our understanding of the evolutionary forces that maintain phenotypic variation in nature (Magurran, 1998; Kapan, 2001; Sinervo

doi:10.1111/j.1365-2699.2009.02260.x

Journal of Biogeography

et al., 2001; Munday *et al.*, 2003) and that contribute to speciation (Jiggins *et al.*, 2001; Seehausen *et al.*, 2003).

One of the most extreme examples of colour polymorphism in vertebrates occurs in the Bocas del Toro archipelago in western Panama, where the strawberry poison frog, *Oophaga pumilio* (Schmidt, 1857), has diverged from a widespread mainland phenotype with a red dorsum and blue legs, into a bewildering array of colours, including orange, yellow, green, blue and black and white morphs, with various forms of melanistic patterning (Fig. 1a; Daly & Myers, 1967; Summers *et al.*, 1997). The evolutionary processes that generated this variation are poorly understood. The biogeography of this polymorphism – with divergent phenotypes mostly restricted to separate islands – suggests a major role for neutral processes. Possibly, the combination of restricted gene flow and reduced population sizes has been sufficient to generate the current distribution of phenotypes, through the random fixation and loss of colour pattern variants across populations.

However, reconstructions of the formation of the Bocas del Toro archipelago suggest that separation of the islands from the mainland began < 10,000 years ago and was not completed until as little as 1000 years ago (Summers *et al.*, 1997; Anderson & Handley, 2002; Wang & Shaffer, 2008). Accordingly, several molecular phylogeographic studies indicate that the different populations of *O. pumilio* in the archipelago are extremely similar genetically (Summers *et al.*, 1997; Hagemann & Pröhl, 2007; Rudh *et al.*, 2007; Wang & Shaffer, 2008). Several authors have therefore argued that the extreme speed of this divergence suggests the involvement of divergent selection on coloration, rather than neutral diversification due to drift (Summers *et al.*, 1997; Rudh *et al.*, 2007; Wang & Shaffer, 2008).

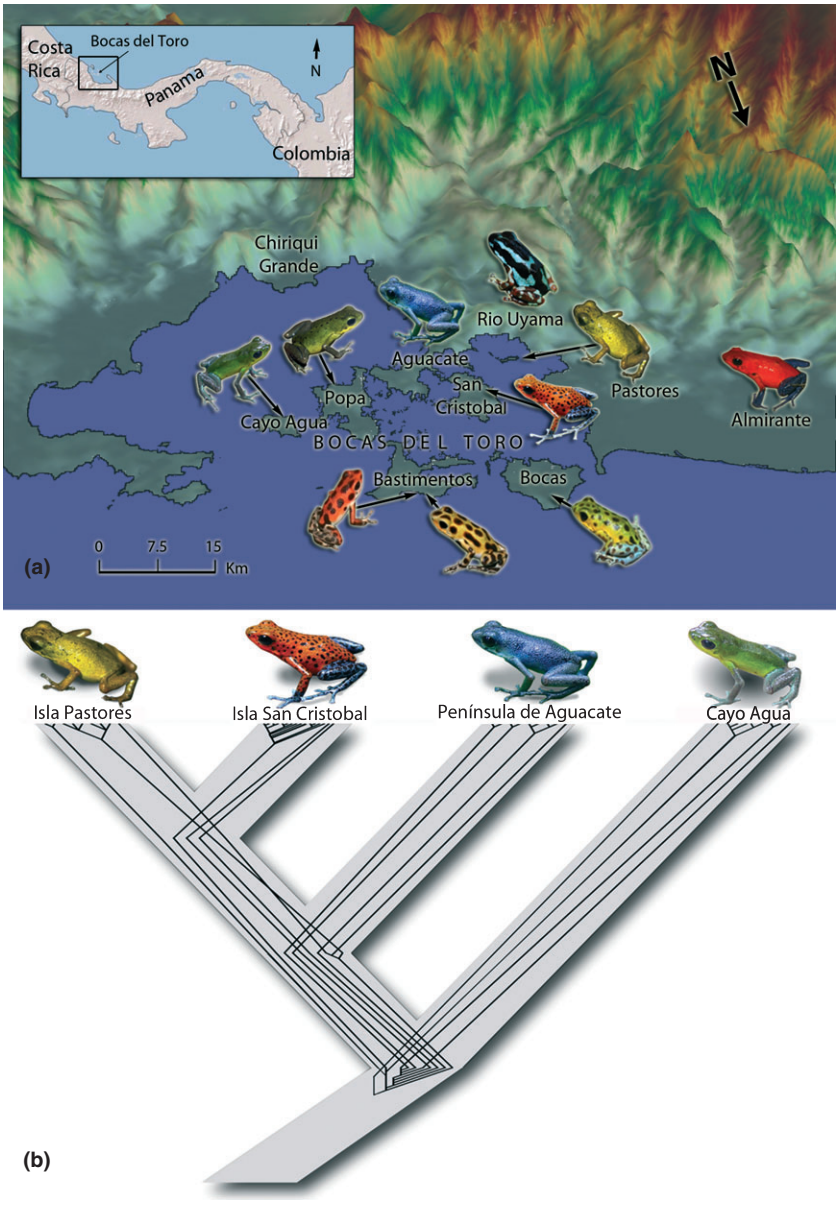

Figure 1 (a) A map of the Bocas del Toro archipelago, Panama, with photos of an individual representative of the focal populations of *Oophaga pumilio* in this study. (b) A 'contained tree' representing the coalescence of mitochondrial DNA (mtDNA) haplotypes for one of the set of most parsimonious gene trees within a population tree for the four populations of *O. pumilio* analysed in this study ('contained tree' topology is from Hagemann & Pröhl, 2007). The lack of reciprocal monophyly of mtDNA haplotypes across populations is consistent across the set of most parsimonious trees.

Adaptive explanations for the colour divergence in *O. pumilio* invoke major roles for natural selection and sexual selection, or some combination thereof (Summers *et al.*, 1997, 1999; Rudh *et al.*, 2007). Since *O. pumilio* coloration is subject to predator selection (Saporito *et al.*, 2007a), spatial variation in predators or habitat features could exert divergent natural selection on coloration. This seems plausible given the evidence for diversifying natural selection in another dendrobatid, *Dendrobates tinctorius*. In this species coloration does not covary with genetic differentiation or geographic distance (Noonan & Gaucher, 2006; Wollenberg *et al.*, 2008), but it is subject to strong selection from avian predators (Noonan & Comeault, 2009). Alternatively, divergence in aposematic coloration may be driven by sexual selection, due to variation among populations in female colour preferences, which in turn may be driven by selection (e.g. sensory drive) or genetic drift. Recent theoretical work supports the possibility of this latter scenario, in which drift in female preference and sexual selection on colour combine in a process of 'coupled drift' (S. J. Tazzyman & Y. Iwasa, unpublished data). The involvement of sexual selection is supported by several studies (Summers *et al.*, 1999; Reynolds & Fitzpatrick, 2007; Maan & Cummings, 2008). This hypothesis also predicts that behavioural isolation between populations precedes the evolution of reproductive incompatibility (e.g. Mendelson, 2003), which is consistent with the observation that frogs from different populations can interbreed and produce viable offspring (Summers *et al.*, 2004).

Strong divergent selection can rapidly drive the fixation of different phenotypic traits in different populations (Kirkpatrick & Ravigné, 2002). In this paper we investigate evidence for an effect of selection on divergence among *O. pumilio* populations by testing whether phenotypic traits are fixed and, if so, whether fixation can be explained by neutral processes or is better explained by selection. We use a rapidly evolving molecular marker to investigate the phylogeographic patterns of divergence across 15 populations of *O. pumilio* in the Bocas del Toro archipelago. In four of these populations, we quantify variation in two phenotypic traits (coloration and body size). To test for selection, we compare the level of reciprocal haplotype monophyly displayed by a putatively neutral molecular marker [i.e. mitochondrial DNA (mtDNA) sequences] to the level expected for a hypothetical gene controlling a specific phenotypic trait of interest (Masta & Maddison, 2002). In our case, the specific trait of interest is coloration, and genetic control of colour and pattern is supported by previous cross-breeding experiments (Summers *et al.*, 2004). This method unites biogeography with phylogenetics and population genetics, providing a novel approach to testing for the action of selection on geographically divergent phenotypes (Masta & Maddison, 2002).

MATERIALS AND METHODS

Frog collection

For genetic analyses, we collected 88 tissue samples (toe tips) from 15 different populations of *O. pumilio* between 1999 and 2001 (see Table 1). We used a tissue sample of *Oophaga speciosus* from Fortuna as an outgroup in the phylogenetic analysis. We used a restricted sample of *O. pumilio* from Cayo Agua ($n = 9$), Isla Pastores ($n = 9$), Isla San Cristóbal ($n = 16$) and the Aguacate Peninsula ($n = 9$) for the analyses of selection. For phenotypic analyses, 94 individuals of *O. pumilio*

Table 1 Populations of the 88 individuals of *Oophaga pumilio* sampled and their corresponding coloration, geographic coordinates of the study sites within Bocas del Toro archipelago, Panama, and GenBank accession numbers of sequences used in this study. *Oophaga speciosa* from Fortuna was used as an outgroup in the phylogenetic analysis.

Population	No. individuals	Primary coloration	GPS coordinates	GenBank accession numbers
Peninsula de Aguacate	9	Blue	9°12.691′ N, 82°12.199′ W	GU138773–GU138781
Almirante	3	Red	9°13.976′ N, 82°22.925′ W	GU138782–GU138784
Bastimentos	11	Orange with spotting	9°21.017′ N, 82°12.728′ W	GU138785–GU138795
Cayo Agua	9	Green	9°8.862′ N, 82°3.175′ W	GU138796–GU138804
Cayo Roldan	1	Green	9°12.953′ N, 82°19.431′ W	GU138805
Chiriqui Grande	3	Red	8°55.913′ N, 82°7.203′ W	GU138806–GU138808
Rio Uyama	2	Black and white	9°10.332′ N, 82°18.775′ W	GU138842, GU138843
Isla Bocas	10	Green with spots	9°22.932′ N, 82°15.822′ W	GU138809–GU138818
Isla Pastores	9	Beige green	9°14.048′ N, 82°20.452′ W	GU138819–GU138823, GU138836–GU138839
Isla Popa	9	Green	9°9.486′ N, 82°8.572′ W	GU138824–GU138832
Ojo de Agua	2	Red	9°18.403′ N, 82°24.746′ W	GU138833, GU138834
Quebrada Pastores	1	Red	9°14.135′ N, 82°20.457′ W	GU138840
Rambala	1	Black and yellow	8°54.368′ N, 82°11.135′ W	GU138841
Isla San Cristobal	16	Red with spots	9°16.355′ N, 82°15.127′ W	GU138844–GU138859
Isla Solarte	2	Orange	9°19.594′ N, 82°12.568′ W	GU138860, GU138861
Oophaga speciosa, Fortuna	1	Red	8°43.0′ N, 82°16.0′ W	GU138835

were captured between 2006 and 2008 from the same four locations, and brought to the Bocas del Toro Field Station of the Smithsonian Tropical Research Institute, Panama (for sample sizes see Fig. 2). Within 24 h after capture, frogs were measured (snout–vent length, to the nearest 0.1 mm) and reflectance spectra collected (see below). After completion of measurements, frogs were returned to the collection sites (except when used for subsequent behavioural studies; see Maan & Cummings, 2008).

Reflectance spectrophotometry and colour metrics

We measured spectral reflectances of *O. pumilio* dorsal and ventral regions using a StellarNet EPP2000C UV-VIS spectrometer and a SL-4 xenon lamp, positioning a R400-7 reflectance fibre with a probe attachment 2.5 mm above the skin. Spectralon white standard measurements were taken between each individual to account for lamp drift. Dorsal reflectance spectra were obtained by averaging measurements of the head, dorsum and lower dorsum (two measurements per region). Ventral reflectance spectra were obtained by averaging two measurements each of the upper throat and belly.

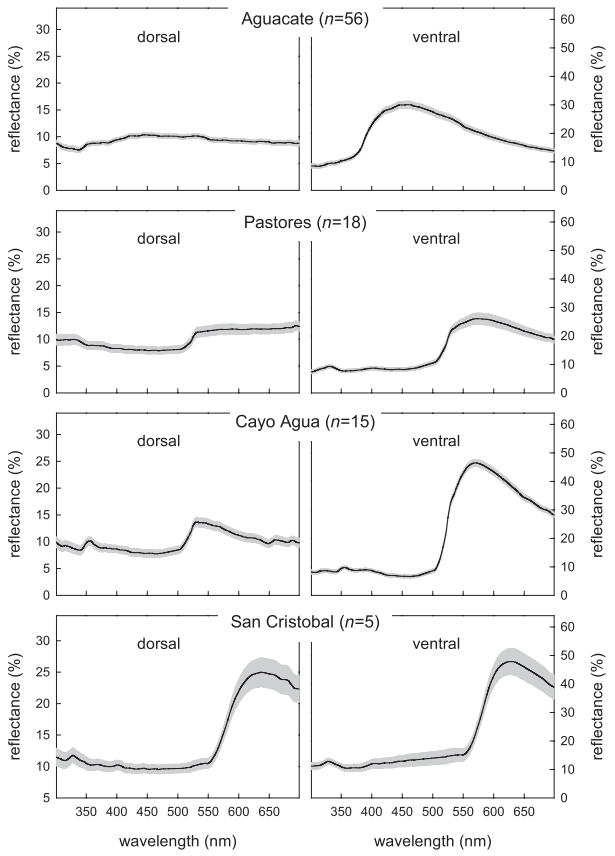

Figure 2 Dorsal and ventral reflectance spectra of the four focal populations of *Oophaga pumilio* from the Bocas del Toro archipelago, Panama: Peninsula de Aguacate, Isla Pastores, Cayo Agua and Isla San Cristobal. Black lines indicate average reflectance across individuals, grey areas are standard errors.

Initial visual inspection of spectral reflectance indicated that populations showed the greatest variation in three wavebands (short, 400–450 nm; middle, 500–525 nm; and long, 540–680 nm). We developed a simple colour metric representing a composite score of the relative reflectance in each of these wavebands, thereby collapsing a continuous variable (see Fig. 2) into a single value that captured the areas of greatest phenotypic variation and enabled statistical comparison across populations. This metric evaluates the change in reflectance in each region (short, middle and long wave) while normalizing each section to baseline reflectance to eliminate differences in total reflectance intensity. For each dorsal and ventral reflectance spectrum, we calculated this colour metric as follows, where $R_{(\lambda)}$ represents the reflectance at each specified wavelength:

$$[(R_{(450)} - R_{(400)})/R_{(400)}] + [(R_{(525)} - R_{(500)})/R_{(500)}] + [(R_{(540)} - R_{(680)})/R_{(540)}].$$

Each individual received a combined colour metric score (dorsal + ventral). While the specific values of these scores are not meaningful in themselves, they allow us to test: (1) whether colour variation between populations exceeds colour variation within populations; and (2) whether the distribution of colour variation is non-overlapping between populations.

Analysis of phenotypic traits

For the analysis of body size variation we included the largest 50% of sampled males and females in each population, to ensure that all individuals were adults. We compared population average values for both body size and colour using ANOVA and *post hoc* pairwise comparisons (Tukey's test). We considered phenotypic traits to be fixed if their distributions were non-overlapping across populations.

DNA sequencing and alignment

Tissue collection and storage, DNA extraction and DNA sequencing followed protocols used by Roberts *et al.* (2006). The final phylogenetic data set included 89 individuals from 15 populations (Table 1 and Fig. 3) and contained 875 bp of mitochondrial d-loop DNA sequence. The d-loop was amplified using the following primer sets (designed for this study): DlpPumi1F(GGTGATTGACTGGAATATTCATGTT), DlpPumi1R(TCTATCTAGCACAATTACCTCATG), DlpPumi2F (GATTGTAATGACTTTATTCTTTCATT), and DlpPumi2R (AATTACTGCATAAATTACATACTATG).

Annealing temperatures were 50 °C for the first six cycles and 55 °C for the following 24 cycles. DNA was sequenced in both directions and consensus DNA sequences were aligned using the Probabilistic Alignment Kit (PRANK; European Bioinformatics Institute, Hinxton, Cambridge, UK; Löytynoja & Goldman, 2005). Because PRANK keeps track of gaps introduced into a multiple sequence alignment, rather than penalizing them, it is expected to

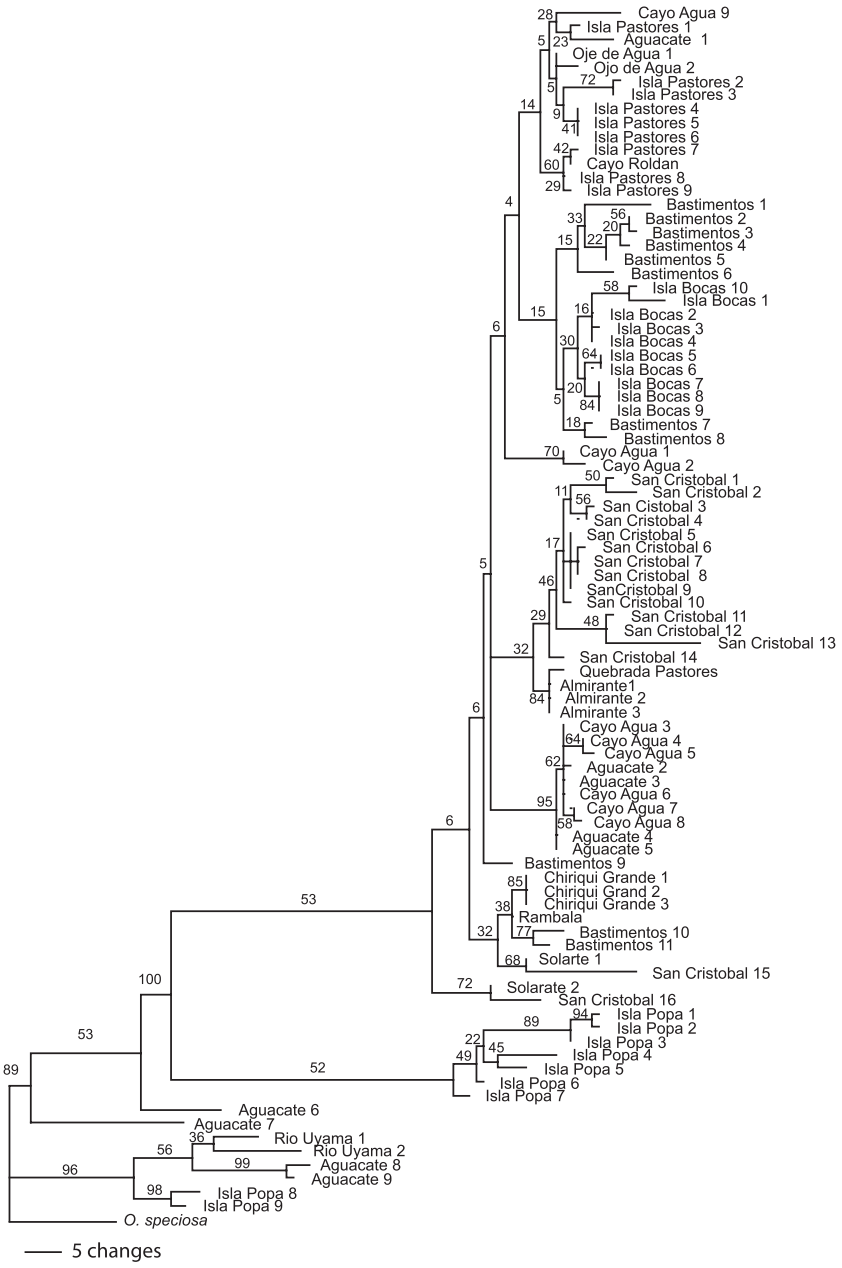

Figure 3 Maximum likelihood estimate of population relationships in the Bocas del Toro *Oophaga pumilio* populations. The estimate was derived using a genetic algorithm approach using GARLI. Numbers above branches represent bootstrap values based on 1000 replicates. The operational taxonomic units represent d-loop sequences from individuals from different populations, as labelled.

better resolve information contained within indel events than other alignment algorithms (i.e. CLUSTALW; Higgins *et al.*, 2005).

Phylogenetic analysis

Phylogenetic analyses were performed using both parsimony criteria in PAUP* (Swofford, 2002) and maximum likelihood (ML) in GARLI 0.951 (Zwickl, 2006) applying a GTR model of nucleotide substitution with gamma distributed rate heterogeneity and a proportion of invariant sites (as suggested by MODELTEST 3.7; Posada & Crandall, 1998). GARLI analyses were repeated 1000 times [each iteration consisted of two independent searches, and of those, the run with the lowest likelihood score (−ln) was stored], and bootstrap values were calculated from the resulting trees.

Testing for selection

We used Tajima's *D*-test to determine whether the sequences are evolving neutrally, and hence were appropriate for our test of selection. This test compares the average level of pairwise sequence divergence with the number of segregating sites for sets of sequences within populations (Tajima, 1989). A negative value for this statistic indicates the action of purifying selection, a selective sweep, population expansion or a complete bottleneck, whereas a positive value is consistent with the action of diversifying selection, popula-

tion mixing or a partial bottleneck (Nielsen, 2005). We used Tajima's *D* to identify the underlying selection trends of the mtDNA d-loop sequence (which serves as our baseline for comparison to evaluate the rate of change in our phenotypic marker). If we uncover indications of purifying selection (negative Tajima's *D*) this will make our test for selection on phenotypic properties using coalescent simulations even more conservative because selection would drive more rapid coalescence in the mtDNA haplotypes within populations. In such a scenario, selection on phenotype (colour) must exhibit even stronger selection (i.e. greater lineage sorting) than that exhibited by the mtDNA to yield a significant difference.

To test for a signal of selection, we selected four different populations: Isla Pastores, Isla San Cristobal, Aguacate Peninsula and Cayo Agua. These populations show distinct dorsal hues (beige-green, red with black spots, blue, and green, respectively). We chose four populations to reduce simulation complexity. The particular populations were chosen because we had good sample sizes for them (both phenotypic and genetic) and because they are representative of the extreme variation that occurs between populations of *O. pumilio* in the Bocas del Toro archipelago. These populations also represent the full range of genetic divergence within *O. pumilio* (see Results).

The test for selection was implemented in MESQUITE version 2.5 (Maddison & Maddison, 2008) and consisted of the following steps (as outlined in Masta & Maddison, 2002).

1. We estimated the phylogenetic relationships from the four selected populations using our mitochondrial d-loop dataset.

2. We then used those trees to calculate the number of steps (*s*) required to reconcile the gene tree with the population tree, which provides a measure of the incompleteness of lineage sorting.

3. These results were compared with simulated mitochondrial gene trees (of the same population and sample sizes), yielding an estimate of the time since divergence.

4. Using those results we simulated the coalescence of a hypothetical nuclear locus that controls phenotype under neutral divergence and compared these results with the actual data.

5. Finally, to test for selection in the context of our observed phylogenetic topology, we estimated the depth of coalescence using the method of Maddison (1997).

Specifically, we are asking whether alleles at a hypothetical gene controlling colour have diverged more rapidly (exhibiting lineage sorting where phenotypic traits are fixed between populations) than a neutral gene would be expected to do. If the estimated coalescence of a neutral gene is shallow (complete lineage sorting) then drift cannot be ruled out as an explanation for phenotypic divergence. If, however, our neutral marker exhibits deep coalescence, or coalescence occurring on a longer time-scale than that of the (hypothetical) colour gene, this would provide statistical support for selection over drift driving phenotypic divergence.

Phylogenetic relationships and calculation of the s *statistic (steps 1 and 2)*

To infer the set of most parsimonious trees for the mtDNA sequences from the four populations ($n = 43$) we used a maximum parsimony analysis (in PAUP*). This yielded a set of 81,300 most parsimonious trees. We then reconstructed these trees as 'contained trees' within a population tree, and calculated the mean value for the *s* statistic (Slatkin & Maddison, 1989), given that population tree topology. Following Masta & Maddison (2002), we used two different tree topologies and two different types of population comparisons for the analyses of the *s* statistic. We used a topology identical to that inferred by Hagemann & Pröhl (2007) in their phylogenetic study using multiple mtDNA markers, and a 'star' topology (i.e. an unresolved polytomy). The use of two topologies ensures that our measure of coalescence is not dependent on a specific population tree topology (given that the 'correct' tree topology is currently under debate). For the population comparisons, we compared all four populations with each other simultaneously (assuming that each population represents fixation of a distinct allele), and we compared one population with all the others combined (assuming only one population is fixed for a distinct allele). In the simultaneous analysis we examined the coalescence of haplotypes within each population at the same time. The second analysis was a two-way comparison where we examined the coalescence of haplotypes within two lineages: one containing one of the focal populations and the other containing all the other individuals (from three populations). This was repeated four times, and each time a different population was designated as the focus for the comparison.

Simulating gene trees and gene coalescence (step 3)

The *s* values obtained in step 2 were compared with those obtained from simulated gene trees (for the same sample size of gene copies) constrained in the same type of population tree under a specific set of conditions (effective population size and branch length). Effective population size was fixed at $N_e = 1000$ across all simulations. Actual N_e-values are likely to be much larger (see Table 2), but because N_e is held constant the value does not influence the difference between rates of coalescence in mitochondrial or nuclear genes, and hence does not influence the significance of our results. The branch lengths were all varied simultaneously and repeatedly until we obtained a value that provided a conservative estimate for the maximum possible branch lengths in the actual tree. This was estimated as the branch length that gave the same value of *s* as the minimum value for the set of most parsimonious trees in *c.* 5% of the simulations. This provides an estimate for the maximum length that the branches can take and still produce a value of *s* consistent with that reconstructed from the set of the most parsimonious trees. For an examination of drift, lengthening the branches represents a conservative approach because it increases the likelihood of coalescence of neutral loci

Table 2 Spatial data on our sampling sites for *Oophaga pumilio* in the Bocas del Toro archipelago, Panama: approximate area, distance and name of the closest body of land, sister population relationships of *O. pumilio* and a maximum estimate of effective population size.

Island/peninsula	Approximate area (km^2)	Name of closest body of land (distance, km)	Sister population(s)* (distance, km)	Gross estimate of maximum effective population size (×1000)†
Isla Bocas‡	65.0	Mainland (1.4)	Bastimentos (2.3)	8000
Isla Bastimentos	54.1	Bocas (2.3)	Bocas (2.3)	6600
Cayo Agua	15.5	Isla Popa (2.0)	Aguacate (10.0)	1900
Isla Popa	57.8	Aguacate (0.8)	A clade containing all islands	7000
Isla Pastores	2.4	Mainland§ (0.8)	Mainland§ (0.8)	300
Isla San Cristobal	39.2	Aguacate (0.6)	Aguacate (0.6)	4800
Aguacate	108.0	n.a.	See Fig. 4	13,100

n.a., not applicable.

*Based on the majority of individuals from the same population.

†Calculated by: [total area/average male territory (16.4 m^2 from Pröhl & Berke, 2001)] × 2 (assuming equal sex ratios). This estimate assumes that each island is homogeneous and entirely hospitable.

‡Also known as 'Isla Colon'.

§Ojo de Agua.

within a particular branch (resulting in genotypic sorting that matches phenotypic sorting).

Simulating the coalescence of a hypothetical nuclear locus (step 4)

Once the maximum value for the branch length was determined, we simulated the coalescence of a hypothetical nuclear locus controlling phenotype. Because the original simulations were done using mtDNA genotypes (which are haploid) we divided the branch lengths by 4 (equivalent to multiplying the population size by 4). Simulations were carried out 10,000 times. We then calculated the distribution of the *s* statistic across the simulated trees. An *s* value of 3 for a four-way comparison represents complete sorting among all populations. For pairwise comparisons of one population against the three others, complete sorting would be indicated by $s = 1$ (indicating that at least one of the populations is fixed for a distinct allele). The significance of the test depends on the frequency distribution of *s* in the nuclear gene tree coalescence simulations. For example, if a value of 3 occurs at a frequency of 5% or more (in the four-way comparison), then we cannot reject the null hypothesis that the divergence of the hypothetical nuclear locus is due to neutral divergence under random genetic drift. In contrast, if a value of 3 occurs < 5% of the time, we can reject the null hypothesis that drift explains the colour variation in present-day strawberry poison frogs.

Testing for selection in the context of observed phylogenetic topology (step 5)

The *s*-statistic is insensitive to the topology of the tree, depending instead on levels of monophyly within populations. In order to replicate this test with a method that is sensitive to the topology of the phylogeny, we carried out the same test described above, but we characterized the coalescent process using Maddison's (1997) measure of 'deep coalescence'. This statistic measures the discordance of a gene tree from a population or species tree. We carried out this test on two distinct topologies, one representing the results of Hagemann & Pröhl (2007), as seen in Fig. 1(b), and another that uses geographic distance for a 'paired' binary tree, with the Pastores and San Cristobal populations as one pair and the Aguacate and Popa populations as the other (data not shown). This test was applied to the four-way comparison only.

RESULTS

Phenotypic variation: colour and body size

Colour variation exhibited no overlap between populations (see Fig. 4a). All four populations were highly significantly different from one another ($F_{3,90} = 478.2$, $P < 0.001$; Tukey *post hoc* comparisons; all pairwise comparisons, $P < 0.001$).

Populations differed significantly in body size (Fig. 4b; $F_{3,58} = 23.69$, $P < 0.001$) and females were significantly larger than males (effect of gender after accounting for population differences: $F_{1,57} = 24.52$, $P < 0.001$). The extent of sexual size dimorphism did not differ significantly between populations ($F_{3,54} = 2.34$, $P = 0.0835$).

Unlike colour variation, size variation within each of the populations overlapped with that of at least one other population (Fig. 4b; female size ranges overlapped for all populations). Aguacate and Pastores frogs did not differ in size (both sexes $P > 0.24$), nor did females from Cayo Agua and San Cristobal ($P = 0.12$). Thus, while colour scores accurately assigned individuals to populations, adult body size did not.

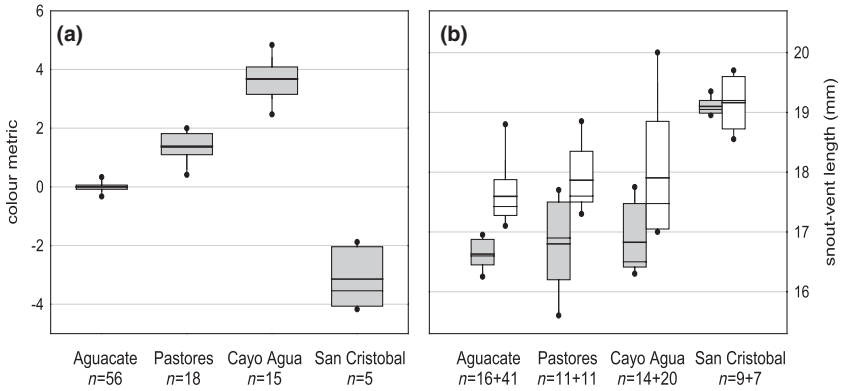

Figure 4 Phenotypic trait distribution across four populations of *Oophaga pumilio* from the Bocas del Toro archipelago, Panama: Aguacate Peninsula, Isla Pastores, Cayo Agua and Isla San Cristobal. Boxes indicate 25th and 75th percentiles intersected by the median (thin line) and mean (thick line). Vertical lines and symbols indicate the data range. (a) Coloration. Summarizing dorsal and ventral coloration in one colour metric, trait values do not overlap between populations. (b) Body size (snout–vent length in mm). Plotted are the 50% largest individuals in each population, separately for males (grey boxes) and females (open boxes). Sample sizes are indicated as *n*(males) + *n*(females). In contrast with coloration, there is considerable size overlap between populations.

Genetic variation

Both the set of most parsimonious trees and the set of ML trees showed substantial evidence of reticulation, or lack of reciprocal monophyly for populations. Figure 3 shows the topology of the ML tree, with bootstrap support values. In order to determine whether this lack of reciprocal monophyly was significant, we tested a most parsimonious tree (from the parsimony analysis) and the tree with the highest likelihood (from the ML analysis) against a tree constrained to be completely reciprocally monophyletic, using (respectively) the Templeton test (Templeton, 1983) and the Shimodaira–Hasegawa test (Shimodaira & Hasegawa, 1999). The tree constrained to be reciprocally monophyletic was significantly longer than the unconstrained tree under parsimony (146 steps longer, $z = -7.37$, $P < 0.001$, Templeton's test). The tree constrained to be reciprocally monophyletic had a significantly lower likelihood than the maximum likelihood tree calculated in GARLI (log likelihood difference = 485.86, $P < 0.001$, Shimodaira–Hasegawa test). The values of Tajima's *D*-test were either not significant or were negative, thus providing no evidence of balancing selection.

For the coalescence analyses (Fig. 1b), we discuss the results from the tree topology identical to the most likely topology in Hagemann & Pröhl (2007), which was also consistent with our phylogenetic analysis (Fig. 3). Figure 1(b) shows a 'contained tree' (a population tree and the gene haplotypes it contains) that shows the coalescence of d-loop haplotypes within the population tree. Comparisons using the alternative ('star') topology gave equivalent results, so we do not present them in detail. The minimum value of *s* calculated across the set of 81,300 most parsimonious trees was 6. Use of the minimum is conservative, as the mean value was substantially higher (10.4). The branch length that yielded a value of 6 for *s* in 5% of the coalescence simulations was $0.9N$ generations, where *N* was set to 1000 (this value was used for all populations, and hence did not influence population differences in branch lengths). Dividing this branch length by 4 (in order to simulate the coalescence of a nuclear gene controlling colour) yielded a branch length of 225 generations. Running the simulation with this value for the branch lengths produced an *s* of 4 for the lowest value, which occurred in fewer than 5% of the simulations. Further, the value $s = 3$, representing complete lineage sorting (or complete reciprocal monophyly, in which every population has its own unique set of haplotypes), did not occur. Together, these results indicate that the observed phenotypic sorting of spectral reflectance (Fig. 4) is highly unlikely ($P < 0.001$) to be the result of neutral coalescence. Similar results were found using Maddison's (1997) measure of deep coalescence. A deep coalescence value of 0 represents complete reciprocal monophyly (lineage sorting); however, our simulations (with the appropriate branch lengths) produced a minimum value of 12. Again, this provides evidence for the action of diversifying selection on colour across these populations. For the pairwise comparisons, each of the four comparisons showed that the probability of achieving the value of *s* expected under reciprocal monophyly (1) was very low ($P < 0.001$). Hence, even if we assume that the colour divergence among these four populations is controlled by the fixation of a single allele, neutral divergence does not provide a satisfactory explanation.

DISCUSSION

Many authors have hypothesized that colour is the target of divergent selection between *O. pumilio* populations in the Bocas del Toro archipelago (Daly & Myers, 1967; Summers *et al.*, 1997; Siddiqi *et al.*, 2004; Hagemann & Pröhl, 2007; Pröhl *et al.*, 2007; Rudh *et al.*, 2007; Wang & Shaffer, 2008). Here we present the first quantitative analysis to support this claim, with evidence for trait fixation that cannot be explained by neutral processes.

Our phylogenetic analysis revealed that lineage sorting is incomplete for the d-loop region sequences used in this study. Our statistical tests of monophyly (Templeton's test and the Shimodaira–Hasegawa test) demonstrated a highly significant lack of reciprocal monophyly of haplotype lineages. This was probably caused by a lack of sufficient time for the normal processes of neutral coalescence to produce gene tree monophyly between populations. Tajima's *D*-statistics were consistent with either neutral evolution or with purifying selection within populations.

The coalescence simulations of the four populations for which we quantified phenotypic variation demonstrated that the probability of fixation of neutral markers is extremely small, rejecting the null hypothesis of neutral evolution leading to the fixation of alternative traits in different populations. In contrast to this pattern of genetic polyphyly, our analysis of spectral reflectance revealed relative uniformity of coloration within populations and complete divergence between populations (Fig. 4a), implying strong divergent selection on this trait. However, there was incomplete sorting by body size (Fig. 4b): we found extensive overlap between populations in both male and female body size, confirming earlier research suggesting gradual evolution of this trait (Pröhl *et al.*, 2007; Rudh *et al.*, 2007). By using objective measures of colour (spectrometry), we were able to identify 'cryptic' divergence: two populations that seem similar to the human eye in hue (Cayo Agua and Isla Pastores; Fig. 1a), exhibit distinct spectral features (Figs 2 & 4) and are probably not the most closely related (Fig. 3). It should be noted, however, that there are other coloration traits that are also distinct between these two populations, specifically leg coloration (sky blue on Cayo Agua; green-brown on Isla Pastores). In addition to limb coloration, our phenotypic analysis ignored population variation in melanistic patterning. This means that we have only captured part of the variation in coloration in this species, and that further analyses could identify additional coloration traits under selection.

While showing that the extreme divergence in *O. pumilio* coloration is likely to be caused by selection, our simulations do not identify the underlying mechanisms. Circumstantial evidence suggests that female choice for dorsal coloration could contribute to divergence (Summers *et al.*, 1997), and recent experimental tests of female colour morph preferences are consistent with that hypothesis (Summers *et al.*, 1999; Reynolds & Fitzpatrick, 2007; Maan & Cummings, 2008). Sexual selection alone would not necessarily explain the initial divergence in coloration among populations, but it is quite possible (even likely) that genetic drift (with or without founder effects) would interact with female colour preferences to trigger divergence (Garcia-Ramos & Kirkpatrick, 1997; Gavrilets, 2003). A recent theoretical study (Tazzyman & Iwasa, 2009) confirms the logical rigour of this scenario in detail for this specific system. This scenario is consistent with the relatively small size of the main islands in the Bocas del Toro archipelago, ranging from 2.4 to 65 km^2 (Table 2).

In a recent study, Boul *et al.* (2007) combined coalescence analysis of frog calls with examinations of female preferences for these calls. The results were fully consistent with the hypothesis that call divergence was driven by sexual selection. In an aposematic organism such as *O. pumilio,* however, we cannot attribute a phylogenetic signal of selection to female mate choice alone. In addition to sexual selection pressures, *O. pumilio* coloration is likely to be subject to natural selection: these frogs are toxic (Daly & Myers, 1967; Saporito *et al.*, 2007b; M. E. Maan & M. E. Cummings, unpublished data) and colour signals probably contribute to avoidance of predation (Saporito *et al.*, 2007a). Population differences in local predator communities, light environments and/or the availability of defensive alkaloids could lead to differences in optimal signal design (Ruxton *et al.*, 2004; Mappes *et al.*, 2005; Ratcliffe & Nydam, 2008).

The substantial lack of reciprocal monophyly observed may be caused either by lack of lineage sorting due to recent separation between these populations or to recent gene flow, or a combination of both. Lack of lineage sorting seems particularly likely to explain most examples, given that the isolation of the islands began < 10,000 years ago, and was not completed until as little as 1000 years ago (Summers *et al.*, 1997; Anderson & Handley, 2002). Moreover, recent research using microsatellite markers indicates substantially reduced gene flow among island populations relative to mainland populations in Costa Rica (Wang & Summers, 2010).

Given the high levels of population polyphyly observed in this study, we urge future researchers of *O. pumilio* phylogenetics to sample multiple individuals from each population, to use both nuclear and mitochondrial genes and to consider coalescence-based methods when estimating species trees. Substantial variation in phylogenetic topologies among studies (Summers *et al.*, 1997; Hagemann & Pröhl, 2007; Rudh *et al.*, 2007; Wang & Shaffer, 2008) suggests that many nuclear and mitochondrial genes have not yet coalesced. The assumption that genes have coalesced may result in erroneous interpretation of biogeographic patterns or evolutionary relationships, even when support levels (e.g. bootstrap values) are high (Liang *et al.*, 2009). On the other hand, if a lack of coalescence is supported by robust sample sizes, it can be exploited to identify selective processes. In particular, when lack of coalescence is associated with different phenotypic trait distributions (e.g. size and colour in the *O. pumilio* system), this provides unique opportunities to examine evolution in action.

ACKNOWLEDGEMENTS

We thank Ricardo Cossio, Angie Estrada, Tiffany Harvey, Sarah Holloway, Ashley Lamb, J. P. Lawrence and Chris Martell for their help with frog collection and measurements. The Smithsonian Tropical Research Institute provided logistical support; we are particularly grateful to Rachel Collin, Gabriel Jacome, Plinio Gondola and Alberto Lawrence of the Bocas del Toro Field Station. We would like to thank Mary

Ramsey, Tara Maginnis, Ryan Wong, Eben Gering, and Laura Crothers for useful comments on early drafts of this manuscript. We are grateful to Marcos Bartelds for providing photographs for Fig. 1(a). This study was financially supported by a Rubicon grant from the Netherlands Foundation for Scientific Research (NWO 825.06.004) and a fieldwork grant from the Association for the Study of Animal Behaviour (ASAB) to M.E.M., a University of Texas Research Award and UT Startup Funding to M.E.C., and a National Geographic Grant (6702-00) to K.S. The Panamanian National Authority for the Environment (ANAM) provided research permission (permit no. SE/A-97-06). This work complied with IACUC protocols UT 04071901 and STRI 200614101606.

REFERENCES

Anderson, R.P. & Handley, C.O., Jr. (2002) Dwarfism in insular sloths: biogeography, selection, and evolutionary rate. *Evolution*, **56,** 1045–1058.

Boul, K.E., Funk, W.C., Darst, C.R., Cannatella, D.C. & Ryan, M.J. (2007) Sexual selection drives speciation in an Amazonian frog. *Proceedings of the Royal Society B: Biological Sciences*, **274,** 399–406.

Daly, J.W. & Myers, C.W. (1967) Toxicity of Panamanian poison frogs (*Dendrobates*) – some biological and chemical aspects. *Science*, **156,** 970–973.

Garcia-Ramos, G. & Kirkpatrick, M. (1997) Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, **51,** 21–28.

Gavrilets, S. (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57,** 2197–2215.

Gray, S.M. & McKinnon, J.S. (2007) Linking color polymorphism maintenance and speciation. *Trends in Ecology and Evolution*, **22,** 71–79.

Hagemann, S. & Pröhl, H. (2007) Mitochondrial paraphyly in a polymorphic poison frog species (Dendrobatidae; *D. pumilio*). *Molecular Phylogenetics and Evolution*, **45,** 740–747.

Higgins, D.G., Blackshields, G. & Wallace, I.M. (2005) Mind the gaps: progress in progressive alignment. *Proceedings of the National Academy of Sciences USA*, **102,** 10411–10412.

Jiggins, C.D., Naisbit, R.E., Coe, R.L. & Mallet, J. (2001) Reproductive isolation caused by colour pattern mimicry. *Nature*, **411,** 302–305.

Kapan, D.D. (2001) Three-butterfly system provides a field test of Müllerian mimicry. *Nature*, **409,** 338–340.

Kirkpatrick, M. & Ravigné, V. (2002) Speciation by natural and sexual selection. *The American Naturalist*, **159,** S22.

Liang, L., Yu, L., Kubatko, L., Pearl, D.K. & Edwards, S.V. (2009) Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution*, **53,** 320–328.

Löytynoja, A. & Goldman, N. (2005) An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences USA*, **102,** 10557–10562.

Maan, M.E. & Cummings, M.E. (2008) Female preferences for aposematic signal components in a polymorphic poison frog. *Evolution*, **62,** 2334–2345.

Maddison, W.P. (1997) Gene trees in species trees. *Systematic Biology*, **46,** 523–536.

Maddison, W.P. & Maddison, D.R. (2008) *Mesquite: a modular system for evolutionary analysis. Version 2.5*. Available at: http://mesquiteproject.org (accessed 1 October 2009).

Magurran, A.E. (1998) Population differentiation without speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **353,** 275–286.

Mappes, J., Marples, N. & Endler, J.A. (2005) The complex business of survival by aposematism. *Trends in Ecology and Evolution*, **20,** 598–603.

Masta, S. & Maddison, W.P. (2002) Sexual selection driving diversification in jumping spiders. *Proceedings of the National Academy of Sciences USA*, **99,** 4442–4447.

Mendelson, T.C. (2003) Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: Etheostoma). *Evolution*, **57,** 317–327.

Munday, P.L., Eyre, P.J. & Jones, G.P. (2003) Ecological mechanisms for coexistence of colour polymorphism in a coral-reef fish: an experimental evaluation. *Oecologia*, **137,** 519–526.

Nielsen, R. (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39,** 197–218.

Noonan, B.P. & Comeault, A.A. (2009) The role of predator selection on polymorphic aposematic poison frogs. *Biology Letters*, **5,** 51–54.

Noonan, B.P. & Gaucher, P. (2006) Refugial isolation and secondary contact in the dyeing poison frog *Dendrobates tinctorius*. *Molecular Ecology*, **15,** 4425–4435.

Posada, D. & Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14,** 817–818.

Pröhl, H. & Berke, O. (2001) Spatial distributions of male and female strawberry poison frogs and their relation to female reproductive resources. *Oecologia*, **29,** 534–542.

Pröhl, H., Hagemann, S., Karsch, J. & Hoebel, G. (2007) Geographic variation in male sexual signals in strawberry poison frogs (*Dendrobates pumilio*). *Ethology*, **113,** 825–837.

Ratcliffe, J.M. & Nydam, M.L. (2008) Multimodal warning signals for a multiple predator world. *Nature*, **455,** 96–100.

Reynolds, R.G. & Fitzpatrick, B.M. (2007) Assortative mating in poison-dart frogs based on an ecologically important trait. *Evolution*, **61,** 2253–2259.

Roberts, J.L., Brown, J.L., von May, R., Arizabal, W., Schulte, R. & Summers, K. (2006) Genetic divergence and speciation in lowland and montane Peruvian poison frogs. *Molecular Phylogenetics and Evolution*, **41,** 149–164.

Rudh, A., Rogell, B. & Höglund, J. (2007) Non-gradual variation in colour morphs of the strawberry poison frog *Dendrobates pumilio*: genetic and geographical isolation suggest a role for selection in maintaining polymorphism. *Molecular Ecology*, **16,** 4284–4294.

Ruxton, G.D., Sherrat, T.N. & Speed, M. (2004) *Avoiding attack: the evolutionary ecology of crypsis, warning signals and mimicry*. Oxford University Press, Oxford.

Saporito, R.A., Zuercher, R., Roberts, M., Gerow, K.G. & Donnelly, M.A. (2007a) Experimental evidence for aposematism in the dendrobatid poison frog *Oophaga pumilio*. *Copeia*, **2007,** 1006–1011.

Saporito, R.A., Donnelly, M.A., Jain, P., Garraffo, H.M., Spande, T.F. & Daly, J.W. (2007b) Spatial and temporal patterns of alkaloid variation in the poison frog *Oophaga pumilio* in Costa Rica and Panama over 30 years. *Toxicon*, **50,** 757–778.

Seehausen, O., Allender, C.J., Knight, M.E., Turner, G.F. & Maclean, N. (2003) Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *Proceedings of the National Academy of Sciences USA*, **100,** 14074–14079.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16,** 1114–1116.

Siddiqi, A., Cronin, T.W., Loew, E.R., Vorobyev, M. & Summers, K. (2004) Interspecific and intraspecific views of color signals in the strawberry poison frog *Dendrobates pumilio*. *Journal of Experimental Biology*, **207,** 2471–2485.

Sinervo, B., Bleay, C. & Adamopoulou, C. (2001) Social causes of correlational selection and the resolution of a heritable throat color polymorphism in a lizard. *Evolution*, **55,** 2040–2052.

Slatkin, M. & Maddison, W.P. (1989) A cladistic measure of gene flow inferred from phylogenies of alleles. *Genetics*, **123,** 603–613.

Summers, K., Bermingham, E., Weigt, L., McCafferty, S. & Dahlstrom, L. (1997) Phenotypic and genetic divergence in three species of dart-poison frogs with contrasting parental behavior. *Journal of Heredity*, **88,** 8–13.

Summers, K., Symula, R., Clough, M. & Cronin, T.W. (1999) Visual mate choice in poison frogs. *Proceedings of the Royal Society B: Biological Sciences*, **266,** 1–5.

Summers, K., Cronin, T. & Kennedy, T. (2004) Cross-breeding of distinct color morphs of the strawberry poison frog (*Dendrobates pumilio*) from the Bocas del Toro Archipelago, Panama. *Journal of Herpetology*, **38,** 1–8.

Swofford, D.L. (2002) *PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4*. Sinauer Associates, Sunderland, MA.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123,** 585–595.

Tazzyman, S.J. & Iwasa, Y. (2009) Sexual selection can increase the effect of random genetic drift - a quantative model of polymorphism in *O. pumilio*, the strawberry poison-dart frog. *Evolution*, doi: 10.1111/j.1558-5646.2009.00923.x.

Templeton, A.R. (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution*, **37,** 221–244.

Wang, I.J. & Shaffer, H.B. (2008) Rapid color evolution in an aposematic species: a phylogenetic analysis of color variation in the strikingly polymorphic strawberry poison-dart frog. *Evolution*, **62,** 2742–2759.

Wang, I.J. & Summers, K. (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, **19,** 447–458.

Wollenberg, K.C., Lötters, S., Mora-Ferrer, C. & Veith, M. (2008) Disentangling composite colour patterns in a poison frog species. *Biological Journal of the Linnean Society*, **93,** 433–444.

Zwickl, D.J. (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. PhD Thesis, University of Texas at Austin.

BIOSKETCH

Jason L. Brown is a NSF Postdoctoral Fellow at Duke University. His current research interests include behavioural ecology, phylogeography, the integration of GIS and phylogenetics, and molecular systematics of poison frogs (family Dendrobatidae).

Author contributions: J.L.B., M.E.M., M.E.C. and K.S. wrote the manuscript; J.L.B. performed phylogenetic analyses and DNA sequencing; K.S. collected genetic samples and performed coalescent analysis; M.E.M. and M.E.C. performed morphological measurements and analyses of morphological data.

Editor: Brett Riddle