

UNCORRECTED PROOF

Testing for selection on color and pattern in a mimetic radiation

Justin YEAGER^{1,2}, Jason L. BROWN^{1,3}, Victor MORALES⁴, Molly CUMMINGS⁵, Kyle SUMMERS^{1*}

¹ Department of Biology, East Carolina University, Greenville, NC, 27858

² Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA 70118

³ Department of Biology, Duke University, Durham, NC 27708

⁴ Universidad de Ricardo Palma, Lima, Peru

⁵ Department of Integrative Biology, University of Texas, Austin, Texas 78712

Abstract In this paper, we analyze variation in spectral reflectance and color pattern among populations to demonstrate dramatic divergence between four distinct morphs of the mimic poison frog *Ranitomeya imitator*. We also analyze genetic divergence in d-loop mtDNA sequences between populations. We then use coalescent-based simulations to demonstrate that the high levels of observed phenotypic divergence are not consistent with levels of genetic divergence expected under neutral drift among populations, implying an important role for selection in driving divergence between these populations [*Current Zoology* 58 (): – , 2012].

Keywords Mullerian Mimicry, Natural Selection, Population Divergence, Coalescent

Müllerian mimicry occurs when one unpalatable species mimics another. This provides an advantage to both species, in that the cost of training naive predators to avoid the shared color pattern is reduced as it is shared between the species (Müller, 1879). There is evidence for Müllerian mimicry in a variety of taxa, including butterflies (Mallet and Joron, 1999), snakes (Sanders et al., 2005) and birds (Dumbacher and Fleischer, 2001). In some cases, populations of a single species have evolved to resemble multiple model species, often in geographically distinct areas, resulting in a “mimetic radiation”. One such radiation likely occurred in Northern Peru, involving the mimic poison frog *Ranitomeya imitator*. Local populations bear a strong resemblance to a co-occurring, congeneric species in four separate areas (Symula et al., 2001; Yeager, 2009). Each of these congeneric species are distinct in color pattern. Phylogeographic analyses indicate that the different populations of *R. imitator* are members of a southern clade of small poison frogs that diverged in the Peruvian province of San Martín relatively recently (Symula et al., 2001; 2003). In most cases, genetic divergence between putative model species substantially exceeds that between distinct morphs of *R. imitator* (Symula et al., 2001; 2003; Brown et al., 2011), suggesting that *R. imitator* is the mimic rather than the model (but see Chouteau et al., 2011). Because the models and mimics in this system are all toxic (Daly et al., 1987), this system likely represents an example of Müllerian mimicry (Symula et al., 2001).

Received Aug. 30, 2011; accepted Nov. 25, 2011.

*Corresponding author. E-mail: summersk@ecu.edu

© 2012 *Current Zoology*

A key component of the mimetic radiation hypothesis is that selection via risk of predator attack drove divergence among distinct morphs of the putative mimic, *R. imitator*. This hypothesis can be evaluated experimentally in the field or laboratory (e.g. Mallet and Barton, 1989), but can also be investigated in a historical context, using evolutionary genetic methods. Recently, Chouteau and Angers (2011) tested the hypothesis of mimicry in the field experimentally, using clay models. These experiments demonstrated that predators impose reciprocal selection pressures across two morphs of *R. imitator*, attacking the non-local color pattern morph when clay models of both morphs were presented in the same area. Here, we use molecular evolutionary methods to test the hypothesis of divergent selection under mimicry. To do this, we use a recently developed method that compares the divergence of a neutrally-evolving gene to that of a hypothetical nuclear gene or genes underlying divergence in color pattern, using simulations of the coalescent (Masta and Maddison, 2002; Brown et al., 2010). Using these simulations we are able to provide statistical support for the action of selection promoting color pattern divergence, in this case driven by mimicry.

1 Materials and Methods

1.1 Sampling for Color and Pattern Analyses

Ranitomeya imitator individuals from 11 locations throughout the range of *R. imitator* were sampled in 2008. Locations in Peru were as follows: Tarapoto, Sauce, Cainarachi Valley, Shapaja/Chumilla, Curiyacu, Santa Rosa, Chipesa, Ricardo Palma, Pongo de Cainarachi, Aguas Termales (all in San Martin province) and Varadero (Loreto province): see Figure 1. Individuals from these populations represented one of four morphs (Fig. 2). Sampling consisted of dorsal and ventral photographs, spectral reflectance and tissue collection in the form of toe clips.

Individuals were measured and both the dorsum and venter were photographed with a Canon 10d DSLR camera with a 100mm macro lens and a macro ring flash, in a standard position. Spectral reflectance was measured in a darkened area (tent) from 12 body regions (6 dorsal, 4 ventral and 2 dorsal leg) using an Ocean Optics USB 4000 spectrometer and a LS-1 tungsten halogen light source. A non-reflective black probe tip was used to maintain a fixed 3mm distance and a 45° angle from the frogs dorsum to account for glare from wet amphibian skin. White standards were taken every other frog using an Ocean Optics WS-1-SL white reflectance standard to account for lamp drift. Six readings were taken per frog by dividing the body in half vertically and in thirds laterally into six parts. Ventral coloration was measured using the same general regions (throat, venter between axilla, center of vent).

1.2 Color Pattern Analyses

Pattern was analyzed by using ImageJ64 software (Rasbund, 1997) to confirm significant variation among populations in elements of pattern (spots, stripes, bands) which promote mimetic similarity among populations. Stripes extending anterior-posteriorly were classified as ‘stripes’ and those that were perpendicular, extending dorsal-ventrally, as ‘bands’. All dorsal pattern markings were first counted and those that were clearly visible were measured for length, width and area for six populations (Tarapoto, Cainarachi Valley, Callanayacu, Chazuta, Sauce and Pongo). Using the tracing tool the black regions were traced and the tracings smoothed, and the area calculated. Averages were taken for length, width and area of spots for each individual. These means were then used in a discriminant function analysis using SPSS v17 to determine if pattern elements used for color pattern convergence differ between mimetic populations. We also investigated the total size (snout-vent length) as it relates to color pattern. While there are marginally significant differences in snout-vent length differences among these mimetic populations (unpubl. data), this does not directly translate into pattern element differences. For example, a larger frog does not have larger spots than a smaller one.

1.3 Spectral Analyses

Variation in morph “color”, independent of pattern, was evaluated via spectral reflectance of dorsal regions. Spectral reflectances were grouped by morph and included the following populations (in parentheses): Varadero (Varadero), Striped (Pongo, Ricardo Palma, Aguas Termales), Banded (Sauce, Curiyacu, Santa Rosa) and Spotted (Tarapoto, Cainarachi Valley). These groupings represent the main mimetic forms of *R. imitator*. Because the four morph colors vary in terms of three color categories or hues (blue, green, and yellow-orange; see Fig 2), we evaluated variation in morph spectral reflectance using a color metric that combined the variation across these three color regions into a discrete quantitative variable. Each reflectance measurement was converted into a single color metric that captured variation in three bandwidths (400–450 nm ‘blue’ or short-, 500-525nm ‘green’ or middle-, and 540-680nm ‘yellow-orange’ or long-wavelength portions of the spectrum) as in Brown et al (2010):

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

where $R_{(\lambda)}$ represents the reflectance at each specified wavelength. Reducing the continuous variable of a spectrum to a discrete unit (color metric) allowed us to evaluate whether ‘color’ differences across morphs were statistically significant and therefore suggestive that the phenotypic trait was fixed. Morph color metrics were then evaluated via ANOVA and TUKEY HSD for post-hoc comparisons.

1.4 Coalescent Simulations

Tissue Extraction and DNA Amplification Tissue samples (toe tips) from 103 individuals from populations showing distinctly different phenotypes were collected between 2006–2008 (Tarapoto, Cainarachi Valley, Shapaja/Chumilla (spotted morph); Sauce, Curiyacu, Santa Rosa, Chipesa (banded morph); Ricardo Palma, Pongo de Cainarachi, Aguas Termales (striped morph), and Varadero), and preserved in vials with 20% DMSO saturated with NaCl. Note that in this study we focus on an analysis of populations that show distinct phenotypes. There are populations that show intermediate phenotypes, which will be analyzed elsewhere. The presence of intermediate phenotypes in some populations does not affect our conclusions concerning the action of selection on the populations analyzed here.

DNA was extracted with the Qiagen Dneasy Tissue Kit (Qiagen Inc, Valencia, CA, USA) following previously published protocols (e.g. Roberts et al. 2006). Amplification of DNA was done with polymerase chain reactions (PCR). Gel electrophoresis confirmed successful amplification and samples were cleaned using ExoSap-it™ using standard protocols. Purified PCR products were then added for sequencing reactions and cleaned again using hydrated Sephadex™ in Millipore™ plates prior to drying and sequencing in an ABI 3130 automated sequencer. Because of the large size, the d-loop region was amplified and sequenced in two parts for both forward and reverse directions. Sequences were aligned using MUSCLE (Edgar 2004, Fig. S1). The final phylogenetic data set consisted of 103 individuals from 12 populations representing four broadly defined phenotypes (spotted, striped, banded and Varadero). See Brown et al. (2010) for complete descriptions of the molecular genetic methods.

Phylogenetic Analysis Phylogenetic analyses were performed with Parsimony using PAUP (Swofford, 2002) and with Maximum Likelihood (ML) using GARLI 0.951 (Zwickl, 2006). Parsimony analysis was carried out in PAUP (Swofford, 2002) using TBR branch-swapping and 2000 random addition replicates. For the maximum likelihood analyses we applied a GTR model of nucleotide substitution with gamma distributed rate heterogeneity and a proportion of invariant sites (based on the results of an analysis of the fit of different models to the data using Modeltest 3.7; Posada and Crandall, 1998). Analyses were repeated 500 times and bootstrap values were calculated from the resulting trees.

Testing for Selection To determine whether the sequences are evolving neutrally we used Tajima’s D test. This test compares the average level of pairwise sequence divergence for sets of sequences within populations with the number of segregating sites (Tajima, 1989). Negative values for this statistic indicate 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, population mixing or a partial bottleneck (Nielsen, 2005). We used Tajima's *D* to investigate the pattern of selection on the mitochondrial DNA sequences (which serves as our baseline for comparison to evaluate the rate of change in our hypothetical nuclear sequences). If this analysis uncovers elements of purifying selection (e.g. a Negative Tajima's *D*), this will make our test for selection on nuclear genes using simulations of the coalescent more conservative as it would drive more rapid coalescence in the mtDNA haplotypes within populations. In such a scenario, selection on phenotype (pattern) must be even stronger, producing greater lineage sorting than that exhibited by the mtDNA, to detect a significant difference.

Both the parsimony tree and the ML tree showed evidence of extensive reticulation, or lack of reciprocal monophyly among haplotypes from different populations (see below). In order to determine whether the degree of reticulation was significantly higher than that expected due to random variation from a tree showing reciprocal monophyly, we tested a most parsimonious tree (from the parsimony analysis) or 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 and Hasegawa, 1999).

The coalescent simulation test for selection was implemented in MESQUITE version 2.6 (Maddison and Maddison, 2008) and consisted of the following steps (following Masta and Maddison, 2002): 1) We estimated the phylogenetic relationships for the haplotypes from the four morphs, and then 2) used those trees to calculate the number of steps required to constrain the gene haplotypes (haploid genotypes) within each morph [this measure is called the *s* statistic (Maddison and Slatkin, 1991)]. 3) These results were compared to simulated gene trees (of the same population size, phylogenetic structure and sample size), which provides an estimate of time since divergence. Using those results, 4) we simulated the coalescence of a hypothetical nuclear locus that controls phenotype under neutral divergence and compared these results to our observed data.

Essentially, we are asking whether alleles at a hypothetical gene controlling color pattern have diverged more rapidly (exhibiting sorting among lineages so that phenotypic traits are fixed between populations) than a hypothetical nuclear gene evolving neutrally. If the estimated coalescence of a neutral gene exhibits shallow (rapid) convergence and complete lineage sorting in a high proportion of the simulations, then drift cannot be ruled out as a possible source of this divergence. If, however, our neutral gene exhibits coalescence occurring on a longer timescale than the observed sorting of the phenotypic trait (color pattern, in this case), this would support the hypothesis that selection rather than drift is responsible for the level of phenotypic divergence observed.

To infer the set of most parsimonious trees for the mtDNA sequences from the four morphs ($n = 103$), we used a maximum parsimony analysis (in PAUP). This yielded a set of 1096 most parsimonious trees. We then reconstructed these trees as "contained trees" within a morph tree (a type of population tree, with each morph representing a distinct population, made up of the adjacent group of populations representing that morph), and calculated the mean value for the *s* statistic (Slatkin and Maddison, 1989), given that morph tree topology. The *s* statistic reconstructs (using parsimony) the number of steps required to reconcile the gene haplotypes with the containing morph tree (so that the haplotypes in each morph form a monophyletic group), and hence provides a measure of the incompleteness of lineage sorting.

Following the Masta and Maddison (2002) approach, we used two different tree topologies and two different types of comparisons for the analyses of the *s* statistic. We used a topology derived from our maximum likelihood phylogenetic analysis, and a "star" topology (i.e. an unresolved polytomy). Population comparisons were done by first comparing all four morphs to each other simultaneously (assuming that each morph represents fixation of a distinct allele), and then comparing one morph with all others combined (assuming only one morph is fixed for a distinct allele).

For these pairwise comparisons, we analyzed each of the morphs as the focal point of the comparison in turn, in four separate analyses.

The values of s obtained from our phylogenetic analyses (step two) were compared against the values of s obtained from simulating gene trees (for the same sample size of gene copies), constrained in the same type of morph tree, under a specific set of conditions (i.e. a specific combination of effective population size and branch length). Effective population size was held constant (at $N_e = 10000$) across all simulations. This number may not be completely accurate, but it is held constant between all analyses and hence does not bias the results. The branch lengths were varied repeatedly until a value was obtained 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 minimum value of s across the set of most parsimonious trees in approximately 5% of the simulations. This provides an estimate for the maximum length that branches in the simulated trees can have 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 more conservative approach as it increases the likelihood of coalescence of neutral genes within a particular branch (resulting in genotypic sorting that matches phenotypic sorting).

Upon determination of the maximum value for the branch length, 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 four (quadrupling the population would have an equivalent effect). Simulations were replicated 10,000 times. The simulated trees were then used to calculate the distribution of the s statistic among the simulations. 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 $s = 1$ (making the more conservative assumption that only one of the populations is fixed for a distinct allele). The significance of the test is dictated by the frequency distribution of s in the simulated dataset. For example, a value of 3 occurring at a frequency of 5% or more (for the four way comparison), would mean we were unable to reject the null hypothesis that the divergence of the hypothetical nuclear locus is due to neutral divergence under random genetic drift. In contrast, if the value of 3 occurs less than 5% of the time, we can reject the null hypothesis that drift and not selection explains the divergence in color pattern among morphs in *R. imitator*.

2 Results

2.1 Image J Comparison

Color pattern differences between the mimetic morphs of *R. imitator* are shown in Fig. 3. Analysis of the specific pattern (spot) characteristics (length, width and area) revealed distinct differences in pattern attributes between populations. Pattern length and area explained the highest proportions of the variability in pattern. These characteristics were significantly different ($P < 0.001$) between the populations sampled demonstrating significant pattern differences between geographic populations (Table 1 and 2).

2.2 Spectral Reflectance

The mean spectral reflectances of the four color morphs are shown in Fig. 4. Color metric analyses showed that the four population reflectances are statistically distinct (ANOVA, $F = 265.49$, $df = 3$, $n=378$, $p < 0.001$), and post-hoc pairwise comparisons showed that each morph was distinct from each other (Spotted-Banded, $P < 0.001$; Striped-Banded, $P < 0.001$; Varadero-Banded $P = 0.037$, Striped-Spotted $P < 0.001$; Varadero-Spotted $P < 0.001$; Varadero-Striped $p < 0.001$; see Fig. 5 for mean and 95 % confidence intervals for each morph).

2.3 Coalescent Simulations

Both the set of most parsimonious trees and the set of maximum likelihood trees showed substantial evidence of reticulation, or lack of reciprocal monophyly across populations. The tree constrained to be reciprocally monophyletic was significantly longer than an unconstrained tree under parsimony (45 steps longer, $t = 4.77$, $P < 0.001$, Kishino-Hasegawa Test). Fig. 6 shows the maximum likelihood tree topology for individuals from populations with the distinct color patterns spotted, banded, striped and Varadero, showing the interdigitation of the haplotypes of these individuals from different populations with distinct color pattern morphs. The tree constrained to be reciprocally monophyletic had a significantly lower likelihood than the maximum likelihood tree calculated in GARLI (log likelihood difference = 250.55, $P < 0.001$, Shimodaira-Hasegawa Test). The values of Tajima's D test were either not significant or negative, providing no evidence of diversifying selection.

For the coalescent analyses, we discuss the results from the star tree topology (i.e. four-way polytomy), as the results were equivalent to those using alternate topologies (data not shown). Fig. 7 shows a "contained tree" that illustrates the coalescence of d-loop haplotypes within the population tree (a bifurcating tree is shown to make the lack of lineage sorting more visible). The minimum value of s calculated across the set of 1096 most parsimonious trees was 15 (mean = 18.33). Use of the minimum is conservative, as the mean value was substantially higher. The branch length that yielded a value of 15 for s in 5% of the coalescence simulations was $0.32N_e$ generations, where N_e was set to 10000 (this value was used for all populations, and hence does not influence population differences in branch length). Dividing this branch length by four (in order to simulate the coalescence of a nuclear gene controlling color pattern) yielded a branch length of 800 generations. Running the simulation with this value for the branch lengths produced an s of 17 for the lowest value, which occurred in less than 5% of the simulations. Thus, the value $s = 3$, representing complete lineage sorting (or complete reciprocal monophyly), did not occur. Together, these results indicate that the observed phenotypic sorting of color pattern is highly unlikely ($P < 0.001$) to be the result of neutral drift under the coalescent. This result held for each population individually as well as all populations together (Table 3).

3 Discussion

The phylogenetic analyses presented here show incomplete lineage sorting for the d-loop region sequences of *R. imitator*. Statistical tests of monophyly (Templeton's test and the Shimodaira-Hasegawa test) demonstrate that this lack of reciprocal monophyly is highly significant. This lack of reciprocal monophyly is likely due to a lack of sufficient time for the normal processes of natural coalescence to sort the haplotypes and produce gene tree monophyly, though it could also be the result of gene flow between populations. The results of the Tajima's D statistic for each population were consistent with either neutral evolution or the presence of purifying selection within populations, and the latter possibility is conservative with respect to our test.

The result of the coalescent simulations demonstrates that the divergence in pattern is highly unlikely to be the result of neutral coalescence of the loci controlling color pattern in *R. imitator*, implying that diversifying selection has played a significant role in the divergence of color pattern phenotypes. It is, of course, possible that the lack of coalescence observed in the mtDNA haplotypes is due to gene flow rather than lack of coalescence, but this would not reduce the strength of our conclusions concerning selection. On the contrary, if gene flow is causing the degree of mtDNA haplotype entanglement (reticulation) observed in these phenotypically distinct populations, then selection on the genes controlling color pattern must have been extremely potent in order to produce the level of distinctiveness observed.

In summary, our genetic analyses as a whole support the hypothesis that the striking divergence seen in color pattern and color reflectance in *R. imitator* is likely to be the result of strong selection promoting phenotypic divergence among

these populations. The four phenotypically distinct populations analyzed using coalescent simulations demonstrated that the probability for the fixation of neutral markers is extremely low, allowing us to reject the null hypothesis of neutral evolution. In direct contrast with the pattern of genetic polyphyly or lack of reciprocal monophyly, sorting of pattern and color reflectances in geographically distinct phenotypes is complete in the populations analyzed.

The tests performed here do not allow us to specifically identify the mechanism of selection driving the differences in color pattern. However, in geographic regions where populations of *R. imitator* are within the range of congeneric species their conformance to the phenotype of the related species is likely the result of predation pressures that promoted Müllerian mimicry (Symula et al., 2001, 2003; Yeager, 2009). Recent experimental work has demonstrated reciprocal predation pressure favoring the local morph between two areas occupied by two distinct morphs (spotted versus striped) in this region of northern Peru (Chouteau and Angers, 2011). By converging on a shared phenotype, both *R. imitator* and the sympatric species it resembles likely benefit by sharing the costs of naïve predator education and both theoretically experience reduced predation due to this common phenotype. It is likely that as *R. imitator* colonized areas previously inhabited by related species strong selection by predators drove the convergence of phenotype to the model species present (Symula et al., 2001, 2003; Yeager, 2009; but see Chouteau et al., 2011 for discussion). Overall, the cumulative evidence from this and previous research supports the predictions of the mimetic radiation hypothesis.

Acknowledgments We are grateful to C. Aguilar, J. Cordova, and K. Siu Ting for their help at the Museo de Historia Natural, San Marcos, Peru, and K. Ramirez for help at the Ministry of Natural Resources. We thank E. Twomey for advice and discussion. This research was funded by grants from the National Science Foundation (IOB-0544010), the National Geographic Society (7658-04), and an East Carolina University Research and Development Grant (2006). Research permits were obtained from the Ministry of Natural Resources (DGFES) in Lima, Peru (Authorizations No. 050-2006-INRENA-IFFS-DCB, No. 067-2007-INRENA-IFFS-DCB, No. 005-2008-INRENA-IFFS-DCB, CITES 11076).

References

- Brown JL, Twomey E, Amézquita A., Barbosa de Souza M, Caldwell JP et al., 2011. A taxonomic revision of the neotropical poison frog genus *Ranitomeya* (Amphibia: Dendrobatidae). *Zootaxa* 3083: 1–120.
- Brown JL, Maan ME, Cummings ME, Summers K, 2010. Evidence for selection on coloration in a Panamanian poison frog: A coalescent-based approach. *J. Biogeog.* 37: 891–901.
- Chouteau M, Summers K, Morales V, Angers B, 2011. Divergence in Mullerian mimicry: The case of the poison dart frogs of northern Peru revisited. *Biol. Lett.* 7:796–800.
- Chouteau M, Angers B, 2011. The role of predation in maintaining the geographic structure of aposematic signals. *American Naturalist* 178: 810–817.
- Clough M, Summers K, 2000. Phylogenetic systematics and biogeography of the poison frogs: evidence from mitochondrial DNA sequences. *Biol. J. Linn. Soc.* 70: 515–540.
- Daly JW, Myers CW, Whittaker N, 1987. Further classification of skin alkaloids from neotropical poison frogs (Dendrobatidae), with a general survey of toxic/noxious substances in the Amphibia. *Toxicon* 25: 1023–1095.
- Dumbacher JP, Fleischer RC, 2001. Phylogenetic evidence for colour pattern convergence in toxic pitohuis: Mullerian mimicry in birds? *Proc. Roy. Soc. Lond. B* 268: 1971–1976.
- Edgar RC, 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nuc. Acids Res.* 32: 1792–1797.
- Kishino H, Hasegawa, M, 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* 29: 170–179.
- Maddison WP, Maddison DR, 2008. Mesquite: A modular system for evolutionary analysis. Version 2.5 <http://mesquiteproject.org>
- Maddison WP, Slatkin M, 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution* 45: 1184–1197.

- Mallet J, Barton NH, 1989 Strong natural selection in a warning-colour hybrid zone. *Evolution* 43: 421–431.
- Mallet J, Joron, M. 1999 Evolution of diversity in warning colour and mimicry: Polymorphisms, shifting balance, and speciation. *Ann. Rev. Ecol. Syst.* 30: 201–233.
- Masta S, Maddison WP, 2002. Sexual selection driving diversification in jumping spiders. *Proc. Natl Acad. Sci. USA* 99: 4442–4447.
- Müller F, 1879. *Ituna* and *Thyridia*: A remarkable case of mimicry in butterflies. *Trans. Entomol. Soc. Lond.* 1879: xx–xxix.
- Nielsen R, 2005. Molecular signatures of natural selection. *Ann. Rev. Genet.* 39: 197–218
- Posada D, Crandall KA, 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
- Rasband WS, ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, <http://imagej.nih.gov/ij/>, 1997–2011
- Roberts JL, Brown JL, von May R, Arizabal W, Schulte R et al., 2006. Genetic divergence and speciation in lowland and montane Peruvian poison frogs. *Mol. Phylogenet. Evol.* 41: 149–164.
- Sanders KL, Malhotra A, Thorpe RS, 2006. Evidence for a Mullerian mimetic radiation in Asian pitvipers. *Proc. Roy. Soc. Lond. B* 273: 1135–1141.
- Shimodaira H, Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16: 1114–1116.
- Slatkin M, Maddison WP, 1989. A cladistic measure of gene flow inferred from phylogenies of alleles. *Genetics* 123: 603–613.
- Summers K, Symula R, Clough M, Cronin TW, 1999. Visual mate choice in poison frogs. *Proc. Roy. Soc. Lond. B* 266: 1–5.
- Swofford DL, 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sunderland, Massachusetts: Sinauer Associates..
- Symula R, Schulte R, Summers K, 2001. Molecular phylogenetic evidence for a mimetic radiation in Peruvian poison frogs supports a Mullerian mimicry hypothesis. *Proc. Roy. Soc. Lond. B.* 268: 2415–2421.
- Symula R, Schulte R, Summers K, 2003. Molecular systematics and phylogeography of Amazonian poison frogs of the genus *Dendrobates*. *Mol. Phyl. Evol.* 26: 452–475.
- Tajima F, 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Templeton AR, 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37: 221–244.
- Yeager J, 2009. Quantification of resemblance in a mimetic radiation. Unpublished Masters Thesis, East Carolina University.
- Zwickl DJ, 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

Table 1 Tests of Discriminant Dimensions

Dimension	Canonical Correlation	Chi-square	<i>df</i>	<i>P</i>
1	0.953	236.741	9	< 0.001
2	0.938	139.680	4	< 0.001
3	0.858	53.888	1	< 0.001

Table 2 Standardized discriminant coefficients dimensions.

Dimension	1	2	3
Area	0.989	0.903	-0.504
Length	0.110	-0.973	0.868
Width	-0.603	0.258	0.960

Table 3 Summary of the results of the coalescent simulations

Comparison	Drift <i>s</i>	Min Val <i>s</i>	Hap Br Ln	Nuc Sim Min	<i>P</i> value
All morphs	3	15	1600	12	<i>P</i> < 0.001
Spotted	1	7	2500	5	<i>P</i> < 0.001
Striped	1	11	1250	9	<i>P</i> < 0.001
Banded	1	8	2100	8	<i>P</i> < 0.001
Varadero	1	3	9000	1	<i>P</i> < 0.01

All morphs is a four-way comparison between all different morphs (spotted, striped, banded and Varadero). Drift *s* refers to the value of *s* expected for complete divergence under neutral drift (with each different morph representing the fixation of a different allele). Min Val *s* refers to the minimum value of *s* calculated across the 1096 equally parsimonious trees from the parsimony analysis. Hap Br Ln refers to the branch lengths (in terms of number of generations) required to produce the minimum value of *s* in 5% of the haploid coalescent simulations. Nuc Sim Min *s* refers to the minimum value of *s* from the nuclear coalescent simulations. See text for further explanation.

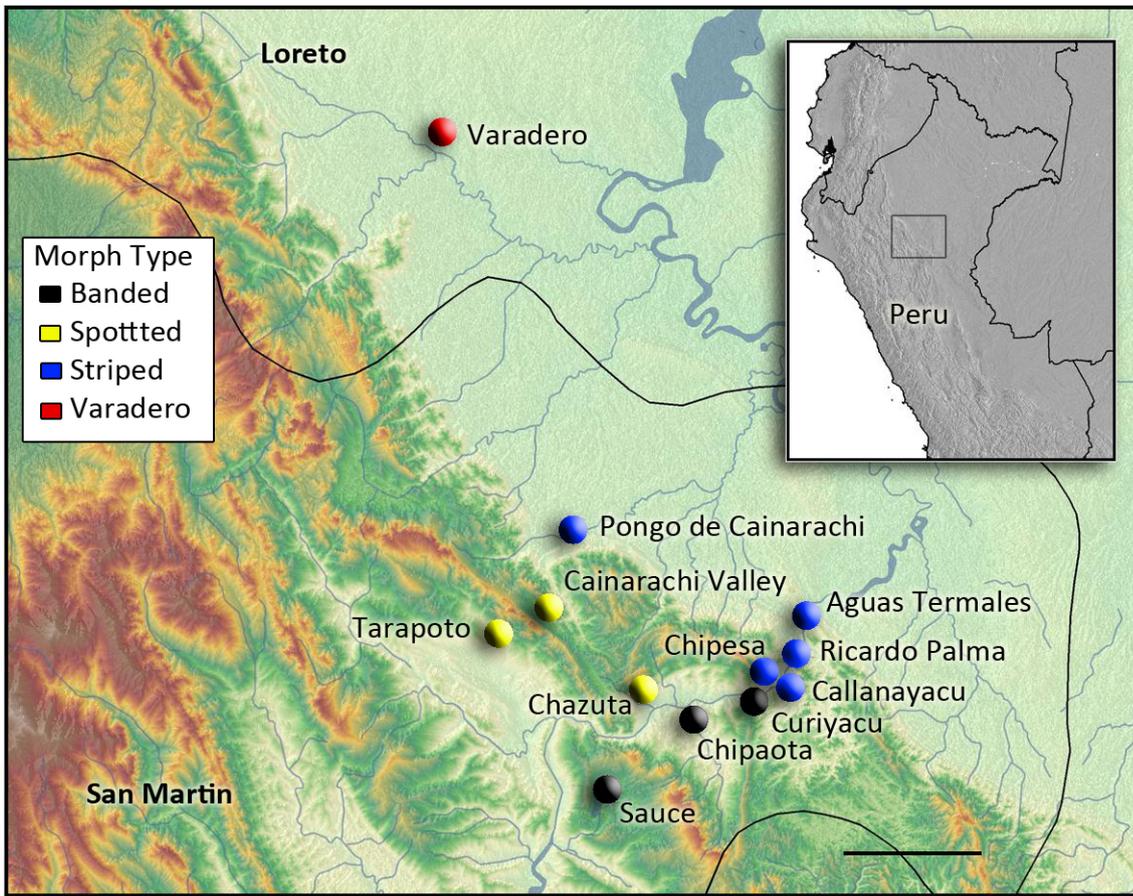


Fig. 1 Map of the main localities where samples for this study were collected

The length of the black bar is 20 km.



Fig. 2 The four mimetic morphs of *R. imitator* across the top, and the four corresponding model species across the bottom of the figure (*Ranitomeya ventrimaculata*, *R. variabilis*, *R. summersi*, and *R. fantastica*)

A. Striped; B. Spotted; C. Banded; D. Varadero. Photos by Evan Twomey.

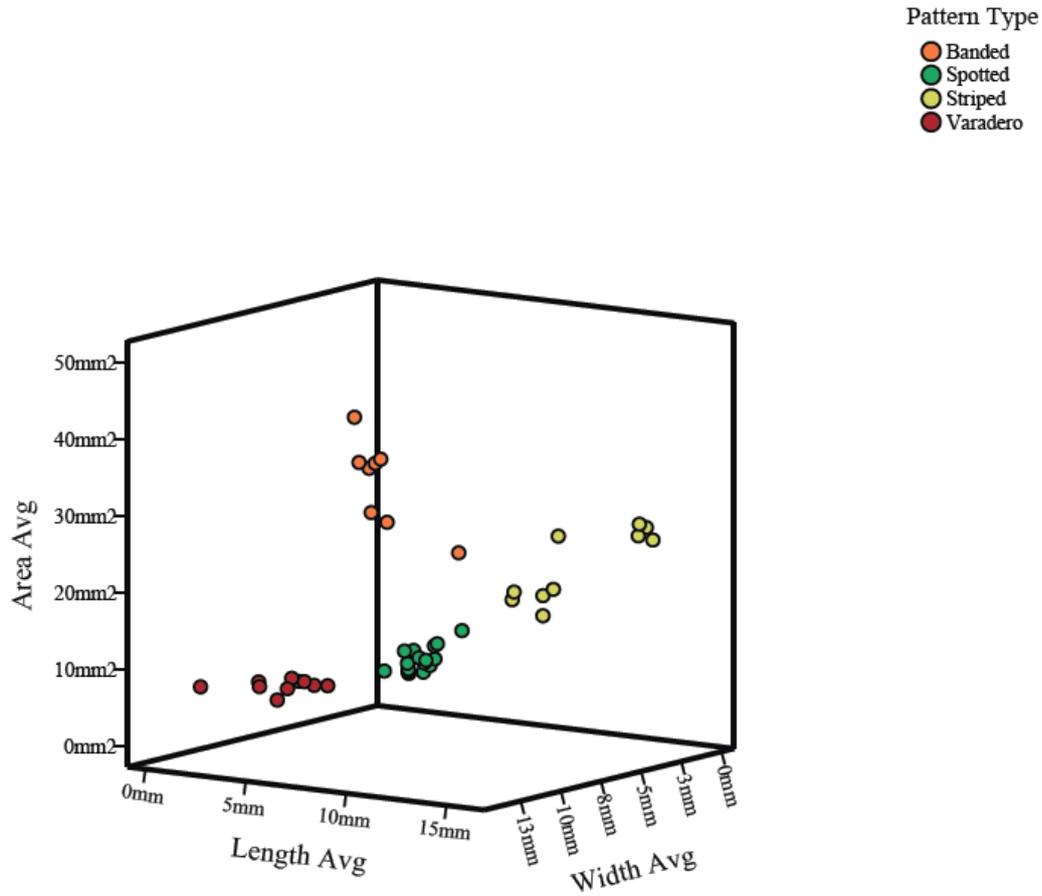


Fig. 3 Visual representation of divergence in pattern elements (length, with and area) between mimetic morphs

Points represent the average of all pattern measurements for the (black) pattern element an individual (e.g. average length of all dorsal spots).

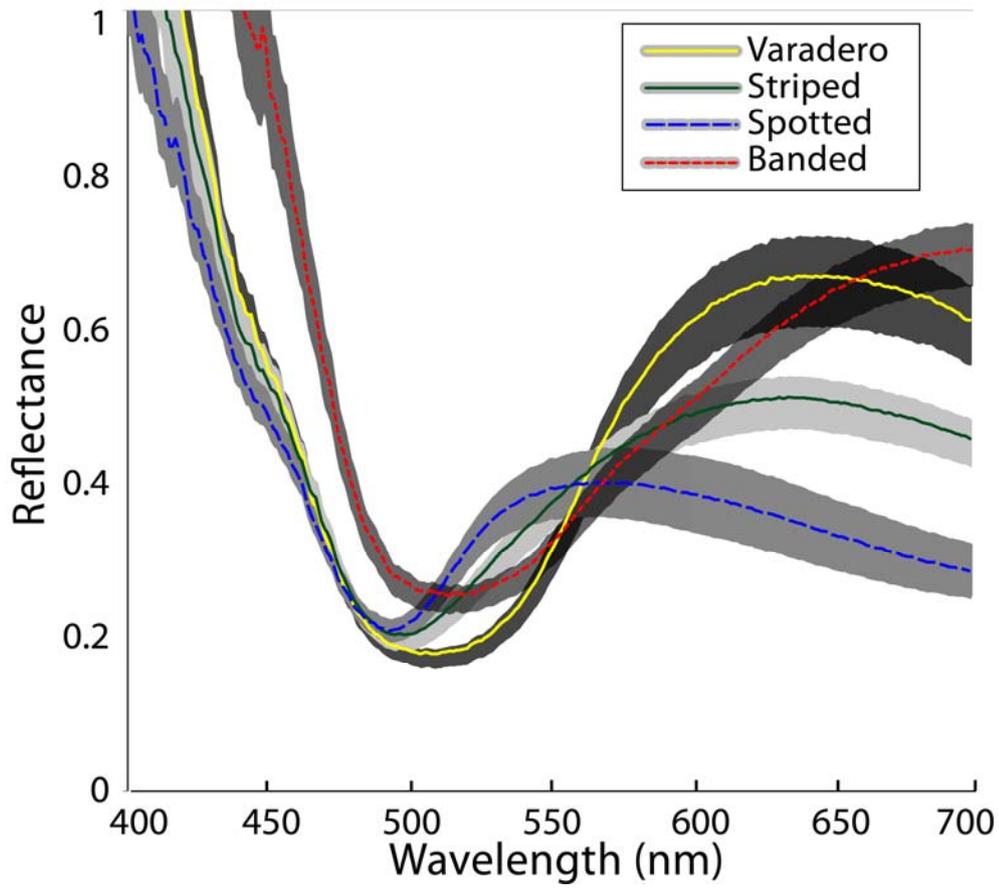


Fig. 4 Morph Dorsum Spectral Reflectances (Mean + 1 standard error bar)

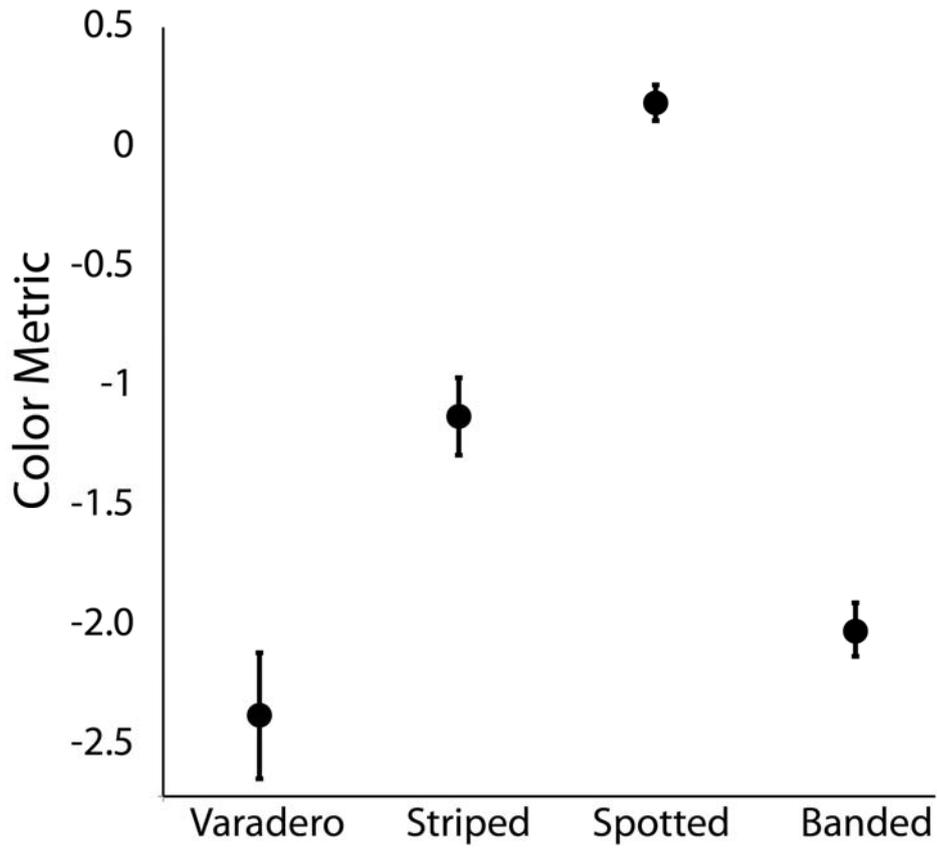


Fig. 5 Mean Color Metric +/- 95 % confidence intervals for each morph

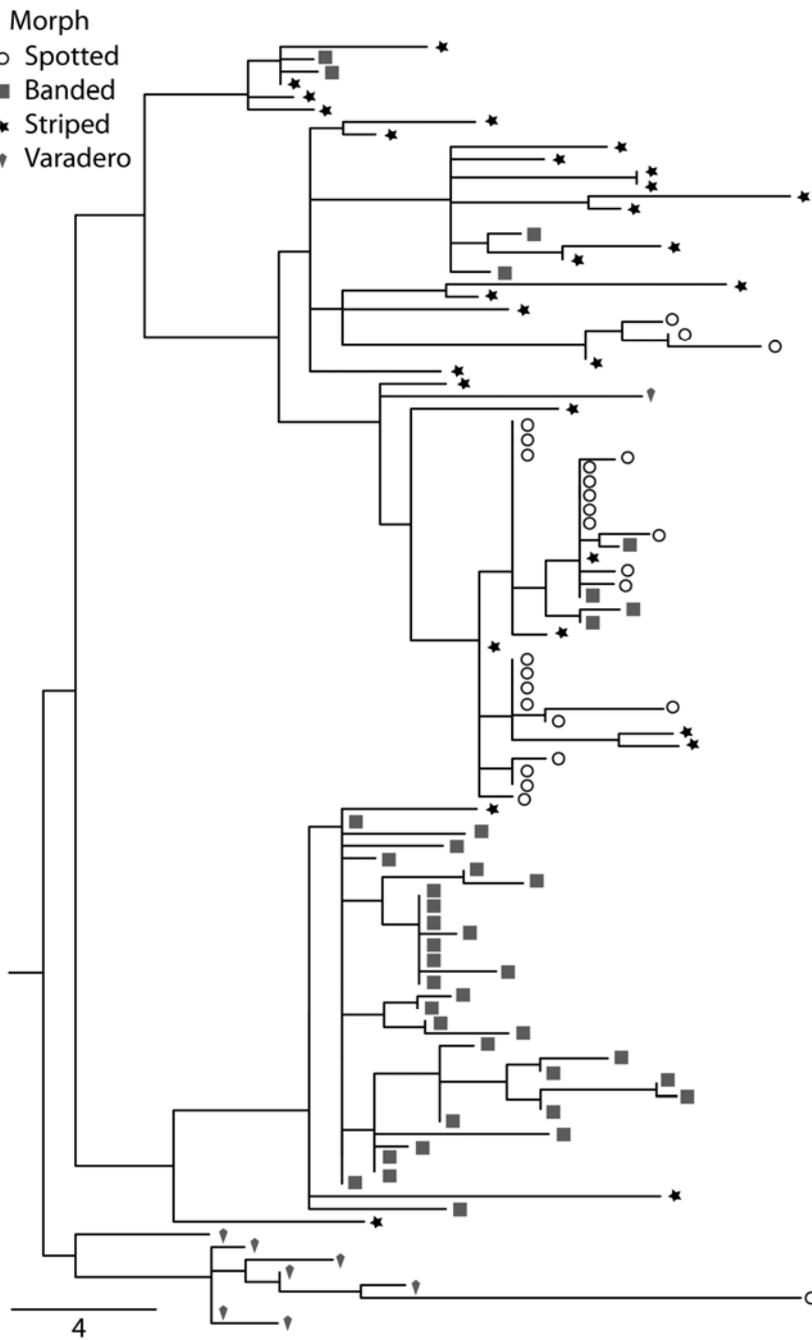


Fig. 6 maximum likelihood topology showing the phylogenetic sorting of the haplotypes of individuals with the four distinct color pattern morphs: spotted (A), banded (B), striped (C) and Varadero (D)

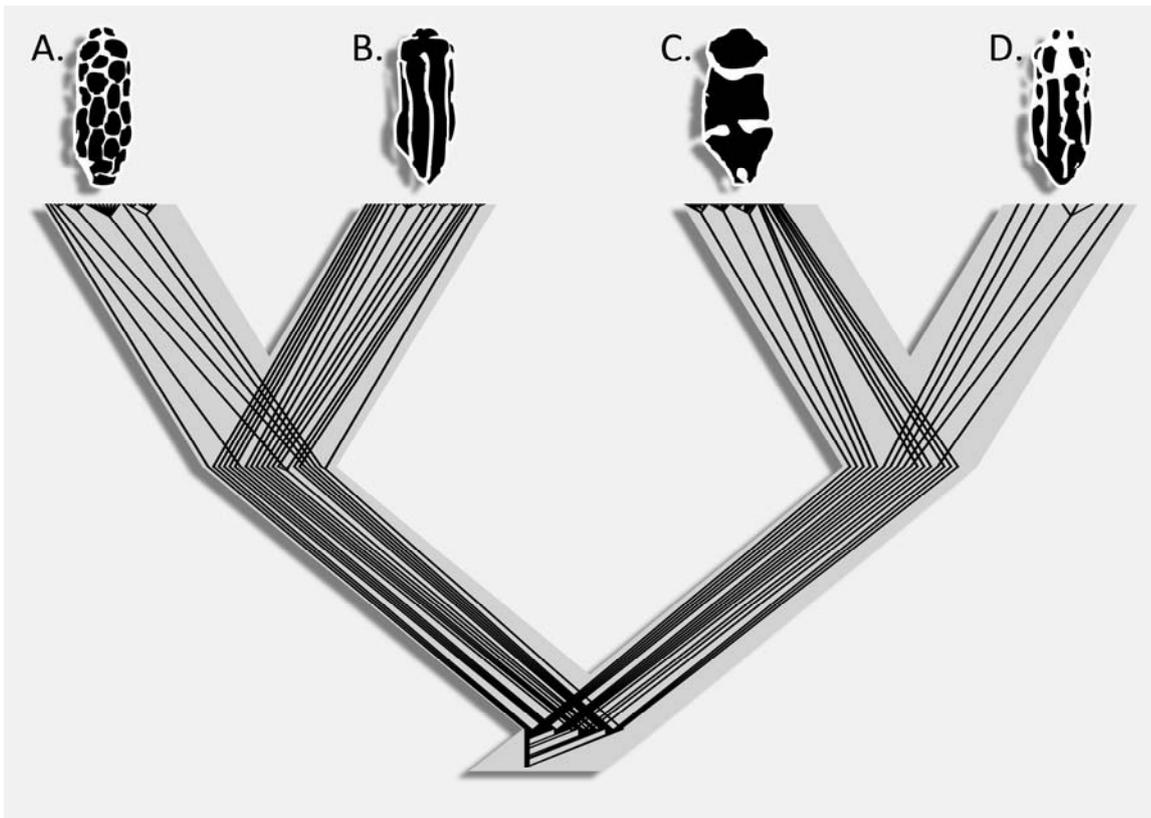


Fig. 7 Contained tree reconciling one of the most parsimonious gene trees from the cladistic analysis with a bifurcating tree representing the four color pattern morphs analyzed: A) spotted, B) striped, C) banded and D) the Varadero morph

A bifurcating tree is shown to enhance the ability to perceive the degree of reticulation of mtDNA haplotypes (or the lack of reciprocal monophyly), but any topology would show the same lack of reconciliation.