Unlinked Mendelian inheritance of red and black pigmentation in snakes: implications for Batesian mimicry

Alison R. Davis Rabosky[1,2]*, Christian L. Cox[3], and Daniel L. Rabosky[1]

[1]Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, 1109 Geddes Ave., Ann Arbor, MI 48109, USA

[2]Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, California 94720

[3]Department of Biology, Georgia Southern University, PO Box 8042, Statesboro, GA, 30460, USA

*corresponding author: ardr@umich.edu

A Brief Communication for *Evolution*

Running head: Genetics of mimetic coloration in snakes

Keywords: coral snake mimicry, coloration genetics, color polymorphism, supergene, linkage disequilibrium, *Sonora semiannulata*

Data archival: Supplementary data files are included with this manuscript and archived at Dryad (upon acceptance; *doi placeholder*)

Abstract

Identifying the genetic basis of mimetic signals is critical to understanding both the origin and dynamics of mimicry over time. For species not amenable to large laboratory breeding studies, widespread color polymorphism across natural populations offers a powerful way to assess the relative likelihood of different genetic systems given observed phenotypic frequencies. We classified color phenotype for 2,175 ground snakes (*Sonora semiannulata*) across the continental United States to analyze morph ratios and test among competing hypotheses about the genetic architecture underlying red and black coloration in coral snake mimics. We found strong support for a two-locus model under simple Mendelian inheritance, with red and black pigmentation being controlled by separate loci. We found no evidence of either linkage disequilibrium between loci or sex linkage. In contrast to Batesian mimicry systems like butterflies in which all color signal components are linked into a single "supergene," our results suggest that the mimetic signal in colubrid snakes can be disrupted through simple recombination and that color evolution is likely to involve discrete gains and losses of each signal component. Both outcomes are likely to contribute to the exponential increase in rates of color evolution seen in snake mimicry systems over insect systems.

Introduction

Mimicry, in which two unrelated species converge upon a novel phenotype for the purpose of signaling to predators, is a classic system for understanding the mechanisms that drive transitions to strikingly new and distinct character states (Mallet and Joron 1999; Reed et al. 2011). "Warning signals" in mimicry systems are usually complex traits composed of multiple distinct elements that are all required for the signal to be effective, especially for coloration (Mappes and Alatalo 1997; Kronforst and Papa 2015). Understanding the genetic determinants of these warning signals is important because theoretical models predict that the genomic architecture of signal components should vary across different kinds of mimicry systems (Charlesworth and Charlesworth 1975; Turner 1987; Charlesworth 2016). In Müllerian mimicry systems in which all members are toxic (or otherwise defended), no genetic linkage is necessary between signal components to promote the evolution of the trait. However, strong linkage into polymorphic "supergenes" is predicted for Batesian systems in which some members are completely undefended (Charlesworth and Charlesworth 1975). This difference in predictions is due to whether or not recombinant types that express just one of the traits (*e.g.*, conspicuous, but not mimetic) suffer reduced fitness in the early evolution of a multi-component mimetic phenotype, which theoretically should affect novel Batesian, but not Müllerian, mimics. In the face of strong selection against undefended conspicuous recombinants, linkage between the signal component loci should evolve rapidly (Charlesworth and Charlesworth 2011).

For the mimicry systems in which we best understand the genetic architecture underlying signal components (primarily coloration in butterflies), these predictions are fairly well supported. There is strong evidence that mimetic loci are organized into polymorphic supergenes in the Batesian *Papilio* (Kunte et al. 2014) and some *Heliconius* butterflies (Joron et al. 2006; Joron et al. 2011) achieved through chromosomal rearrangement and inversions,

while the majority Müllerian *Heliconius* species retain critical coloration components on separate chromosomes (Kronforst and Papa 2015). In both kinds of mimicry systems, the predominant pattern is that there are a few loci of major effect with a number of smaller modifier loci (Nijhout 2003; Franks and Sherratt 2007; Gamberale-Stille et al. 2012; Leimar et al. 2012), and convergence among species on the same color phenotype is at least in some cases driven by parallel evolution in the same genes (Reed et al. 2011; Martin et al. 2012).

Coral snake mimicry (the imitation of highly venomous Elapid coral snakes by harmless Colubrid snakes across North and South America) offers an opportunity to independently test the genetic predictions of Batesian mimicry systems. Although the majority of coral snake mimics are not considered toxic to predators, the cost of misidentifying an atypically colored coral snake would be very high (*i.e.,* death) relative to the cost of misidentifying a noxious insect (Pough 1988). This critical difference in cost could create altered fitness landscapes for recombinant types in the mimetic species. If selection against misidentification is strong enough, then the "conspicuous recombinant" may still be avoided by predators and linkage among signal components would not be generated. As in butterfly mimics (Joron and Mallet 1998), many snake species involved in mimicry systems have striking color polymorphism in the warning signal components (Savage and Slowinski 1992; Davis Rabosky et al. in review). As polymorphism can arise through either supergene or unlinked models of mimicry genes (Charlesworth 2016) but produces differing frequencies of "recombinants," this phenotypic variation can be leveraged to test among these competing hypotheses about the genetic architecture underlying color expression (Wright 1943).

In coral snake mimicry, the warning coloration involves variations on highly contrasting alternating bands, especially in black and red (Savage and Slowinski 1992; Brodie and Brodie 2004). Color polymorphism within coral snake mimicry occurs through

variation in either of these two signal components, often by presence/absence to produce all phenotypic combinations: 1) both red and black, 2) only red, 3) only black, and 4) neither (Davis Rabosky et al. in review). Although comparatively little is known in snakes about the genetic pathways driving synthesis and distribution of either pigment, they are thought to be endogenously produced pteridines (likely drosopterins) and melanins, respectively (Bechtel 1978; Bagnara 1983; Kikuchi and Pfennig 2012; Kikuchi et al. 2014). The differing structural morphology, biochemistry, and dermal layer distribution of the chromatophores that produce each pigment (Kuriyama et al. 2013) in vertebrates suggest that red and black pigmentation are likely controlled by separate genetic loci, which is bolstered by a limited set of common garden breeding experiments using artificially selected color mutations in corn snakes (Bechtel and Bechtel 1962, 1978). Although some quantitative genetic studies have found correlation among pigment blotch sizes in garter snakes (Westphal and Morgan 2010; Westphal et al. 2011), neither the genetic control nor linkage of red and black coloration has ever been investigated within coral snake mimicry.

For organisms with long generation times and low reproductive output that are not amenable to breeding studies within the laboratory, large-scale population sampling of color polymorphism, in combination with the predictions of Hardy-Weinberg equilibrium, can provide a powerful way to test among genetic models governing coloration in mimetic species. In this study, we leveraged such population-level data in the polymorphic Western Ground Snake (*Sonora semiannulata*) to test among a) one- vs. two-locus models, b) linkage disequilibrium among loci, and c) sex-linked vs. autosomal models and assess the relationship between observed and predicted genetic architecture of the mimetic signal in a non-insect Batesian mimic.

Materials and Methods

Study system

The Western ground snake (*Sonora semiannulata*) is a small insectivorous snake native to the western United States and northern Mexico. This species is best known for its striking color polymorphism, with up to four discrete color morphs occurring sympatrically within single populations (Cox and Davis Rabosky 2013). All four color morphs are found in both sexes and all age classes, with no ontogenetic color change (Cox and Davis Rabosky 2013). These color morphs include all combinations of red and black pigmentation, including individuals with a) both color components, b) only black, c) only red, and d) neither (Fig. 1). The black and red morph of *S. semiannulata* is commonly interpreted as a mimic of venomous coral snakes (Savage and Slowinski 1992; Brodie and Brodie 2004; Campbell and Lamar 2004; Cox et al. 2012) due both to its striking similarity to coral snake color pattern and to the evolutionary history of coloration within the larger tribe Sonorini, which also suggests the origin and maintenance of the mimetic form in this clade for the last 25 million years (Davis Rabosky et al. in review). Despite its importance in other squamates (Rosenblum 2006; Rosenblum et al. 2010), we have found no link between the melanogenesis gene *Mc1R* and black banding in Sonorine snakes (Cox et al. 2013).

Population sampling and test for sex linkage

To assess morph ratios within populations, we examined 2,175 fluid-preserved specimens from the continental United States housed in 12 institutional collections (see Acknowledgements). In addition to collecting specimen size (snout-vent length and tail length) and sex data, we scored dorsal coloration on each specimen as mimetic red and black (M), black-banded only (B), red-striped only (S), or a uniform brown or tan with neither type of marking (U). Although some pigmentation fades in the preservation process, both stripes and bands are readily visible against the remaining background body pigment even in 100-year old museum specimens (Cox and Davis Rabosky 2013). Each snake was independently scored by at least two researchers to ensure repeatability of phenotype scores.

To create a subset of individuals for population analysis of morph frequencies, we filtered the overall dataset to populations with at least 12 individuals collected within a 20-year window (N = 40 populations; see *Power analysis and simulations* below). For populations with robust collection histories over long time periods, we used the single 20-year window with the greatest number of specimens. We defined populations by county of collection in order to maximize use of specimens without georeferenced latitude-longitude coordinates. For each county, we recorded the number of specimens of each color morph, the first year of the 20 year window used for that population, and the latitude and longitude for the county midpoint using metrics from the U.S. Census Bureau (2011).

To assess whether coloration is sex-linked or autosomal, we conducted a chi-square test against the expected number of males and females of each morph under a null (autosomal) model to identify deviations suggesting sex linkage. First, we analyzed all individuals for which we had reliable sex data (N = 253 females and 482 males), irrespective of population. A greater proportion of males were sexable with high confidence because they are sometimes preserved with the hemipenes everted, accounting for much of the difference between the sexes in sample size. We also conducted within-population tests for every population with a) at least 20 individuals overall, and b) at least 5 individuals of each sex, but these analyses were limited by suitably-sized populations (N = 7).

Tests for spatial and temporal autocorrelation

To assess statistical independence of populations prior to analysis, we used Mantel tests (N = 999 permutations) in the R package 'vegan' to test for both spatial and temporal autocorrelation in morph frequencies. Because previous research found evidence that negative frequency-dependent selection is operating to maintain polymorphism in at least some populations (Cox and Davis Rabosky 2013), a temporal component was important to include in our analyses. We calculated Euclidean distance matrices among populations based

on 1) both morph frequencies and morph presence/absence (two types of phenotypic distance) and 2) time window start years (temporal distance), and by geodesic distance using the Haversine formula after converting latitude/longitude coordinates to radians (geographic distance). We then compared each phenotypic distance matrix with a) the geographic and temporal distance matrices individually, as well as b) in combination within a partial Mantel test framework. To mitigate the possibility of finding a spurious negative correlation across both long distances and a known genetic clade break through central New Mexico (Davis Rabosky et al. in review), we also repeated these analyses after dividing the overall dataset by clade affiliation into "Western" and "Great Plains" clades (as in Table S1). Significance was assessed at $P < 0.05$.

Likelihood analysis of color morph genetics

We defined three simple models for the inheritance of color phenotypes in ground snakes: *one-locus*, t*wo-locus*, and *two-locus with linkage*. Under the one-locus model, we assumed that color phenotype is a function of three alleles: red (*r*), black (*b*), and null (*n*). In this model, the black and red alleles are co-dominant in the mimetic phenotype, and both are dominant to the recessive null allele; this is the simplest genetic system that can lead to the four observed morphs. For a given population, the likelihood of the data is simply the probability of the observed phenotypic data given a vector of allele frequencies under the assumption of Hardy-Weinberg equilibrium. For example, *2rb* is the probability that a single randomly sampled individual will show the mimetic phenotype, and $r^2 + 2rn$ is the probability that an individual will exhibit the striped phenotype. The two-locus model assumes that red and black pigmentation are coded by dominant alleles at separate, independently-assorting loci. Each locus is assumed to have both a pigment-producing (*r* or *b*) and null allele (n_r or n_b), and a snake that is homozygous for the null allele at both loci would have uniform brown patterning. Note that for both models, the number of free

parameters per population is the same (np = 2), as allele frequencies for each locus must sum to 1.

To test for linkage disequilibrium among loci, we added an additional parameter D to the two-locus model that specified the extent of gametic disequilibrium for red and black alleles. D is a purely statistical measure of the gametic association of alleles from different loci. Letting n_r and n_b denote null alleles at the red and black loci, the gametic frequencies are:

$$
\begin{aligned}
f(r,b) &= rb + D \\
f(r,n_b) &= rn_b - D \\
f(n_r,b) &= n_r b - D \\
f(n_r,n_b) &= n_r n_b + D
\end{aligned}
$$

With D = 0, all gametic frequencies follow their expectation under independent assortment. D > 0 implies an excess of *rb* gametes. D takes a maximum value that depends on the allele frequencies. With D > 0, $D_{max} = min(rn_b , n_r b)$: the maximum is necessary because, were D to exceed this value, some gametic frequencies would be negative. The full probability model for the phenotypic data is derived from the above frequencies. For example, the probability of observing a mimetic phenotype is

$$f(r,b)^2 + 2f(r,b)f(r,n_b) + 2f(r,b)f(n_r,b) + 2f(r,b)f(n_r,n_b) + 2f(n_r,b)f(r,n_b)$$

which is simply the sum of all genotypic frequencies that could yield a mimetic phenotype under the model. We defined two versions of the two-locus model with linkage. In the first, we treated D as a free parameter to be estimated separately for each population. For the second model, we assessed the likelihood of the phenotypic data under *strong linkage* between red and black alleles, as predicted by genetic models of mimicry. For this latter model, we fixed the relative value of D ($D_{rel} = D / D_{max}$) to 0.99. This model assumes that the red and black alleles will show gametic associations close to their theoretical maximum value. Importantly, with a fixed relative D, the number of parameters per population is

identical to the two-locus model without linkage (D = 0).

The probability the data (X) is the simply the product of the probabilities of observing individual phenotypes across all K populations under the model (M) and parameters, or

$$\Pr(X \mid M) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} \Pr(x_{i,k} \mid M, q_k)$$

where n_k is the number of individuals in the k'th population, $x_{i,k}$ is the observed phenotype for the i'th individuals from population k, and θ_k is the parameter vector for the k'th population. We implemented all inheritance models in the R programming language (Dryad doi: *placeholder*), finding maximum likelihood estimates of parameters using a Nelder-Mead simplex algorithm on a logit-transformed parameter space. We fitted all models to the observed phenotype counts for each population, estimating allele frequencies and linkage parameters (if relevant).

We computed ΔAIC scores for each population separately with the polarity $AIC_x - AIC_2$, where AIC_x is the AIC score for model x and AIC_2 is the AIC score for the two-locus, unlinked model. With this polarity, ΔAIC > 0 indicates better fit of the unlinked two-locus model relative to the alternative model. To estimate overall model fit, we summed ΔAIC scores across populations. This operation is mathematically identical to computing ΔAIC from the overall log-probability of the complete data vector. However, by decomposing the overall ΔAIC into population-specific scores, we are able to visualize the extent to which support for different models varies among populations in addition to presenting estimates of "global" model fit.

Power analysis and simulations

We conducted parametric simulations under the fitted one- and two-locus models to test (i) whether our approach has sufficient power to distinguish between the candidate models given the sample sizes and phenotype distributions for each population, and (ii) the

overall goodness of fit of the simple models. Using the maximum likelihood estimates of allele frequencies for each population under the one and two-locus unlinked models, we simulated pseudo-populations of phenotypes assuming Hardy-Weinberg equilibrium. Each simulated population was constrained to have the same number of individuals as the true study population. For the *i'*th population, a single instance of the simulation under the one-locus model is as follows. Given maximum likelihood allele frequencies r_i, b_i, and n_i, we first computed the equilibrium distribution of phenotypes under Hardy-Weinberg equilibrium (*e.g.*, expected "red striped" frequency = $r_i^2 + 2r_in_i$). We then sampled N_i phenotypes from this distribution, where N_i is the number of individuals in the observed data for the *i'*th population.

We fitted one- and two-locus models to each such simulated dataset, to assess our power to infer each scenario given a particular (known) model of inheritance. We graphically compared distributions of phenotypes from populations simulated under each model to the observed distribution as an initial test for model adequacy. We then computed the difference in AIC scores (ΔAIC) for each population under the two models, which we always computed with the polarity described above ($AIC_{one\text{-}locus}$ - $AIC_{two\text{-}locus}$). Thus, under a true (generating) one-locus model, we expect ΔAIC to be negative, as the one-locus model should have a lower (better) AIC score. Conversely, under true (generating) two-locus model, we expect ΔAIC to be positive. We performed 1000 simulations under both one- and two-locus models per population. Finally, we assessed overall goodness-of-fit by comparing the observed ΔAIC value (summed across all populations) to the simulated distributions of ΔAIC under perfect one- or two-locus inheritance.

Results

Sex linkage

We found no evidence for significant sex linkage of ground snake coloration either in the overall dataset (χ^2 = 6.26, P = 0.10) or within any of the seven populations with large enough sample sizes to test (χ^2 = 0-3.35, P = 0.34-1.0), suggesting that coloration is an autosomal trait.

Spatial and temporal autocorrelation

While we did find evidence of spatial autocorrelation among populations when analyzing the frequency of morphs across all populations together (Mantel statistic r = 0.472, P = 0.001), this effect was generally driven by low-sample size populations in close proximity (< 100km) across the Great Plains and an interaction among populations across a known taxonomic break (Western vs. Great Plains; Fig. 2; Table S1). We found that spatial autocorrelation was weaker or absent when morph presence, which is perhaps a more robust metric of polymorphism, was used instead of morph frequency (r = 0.062-0.168, P = 0.017-0.346; Fig. 2). Across the Western clade populations, we found no evidence of spatial autocorrelation in any phenotypic metric or distance class (r = 0.117-0.190, P = 0.128-0.346). Overall, spatial autocorrelation was also reduced or eliminated when populations were analyzed under the more stringent sampling criterion of 20 individuals per population (compare Fig. 2a to 2b), although this reduced the dataset from 40 to 21 total populations. We found no effect of temporal autocorrelation, either when analyzed separately (r = -0.074-0.243, P = 0.089-0.928) or in a partial Mantel test simultaneously with the geographic distance matrix (no change to significance class when compared to the pure spatial model in any case).

Model comparison

We found strong support across populations for a multi-locus model over a single-locus model, with an overall ΔAIC = 154.38 in support of the two-locus model (Fig. 3; Table S2). Twelve of these populations had AIC differences greater than +2 units, while no

281 population had significant support for the one-locus model (≤ -2 units; Fig. 3a). We found
282 that these differences in model support were not biased by levels of polymorphism ($F_{1,38}$ =
283 1.53, P = 0.223; Fig. 3b) or sample size ($F_{1,38}$ = 1.69, P = 0.202; Fig. 3c). Although we
284 found many populations with ΔAIC scores at or near 0 (support for neither model, N = 20),
285 the majority of these were populations from the Great Plains clade with a high frequency of
286 the uniform morph, usually in close geographic proximity to each other (*e.g.*, those
287 populations with statistically detectable spatial autocorrelation; Fig. 3a; Table S1). Four of
288 the 5 populations with the greatest ΔAIC score, which accounted for two thirds of the total
289 ΔAIC sum, were from the Western clade with no spatial dependency of phenotype.
290 We found no statistical evidence for strong linkage disequilibrium between the red
291 and black loci, with the unlinked two-locus model supported over both a) the linkage model
292 with D as a free parameter (ΔAIC = 29.52 favoring the unlinked model), and b) the strong
293 linkage model with D fixed at the theoretical D_{max} for each population (ΔAIC = 230.71
294 favoring the unlinked model). These results for the strong linkage model were quite striking,
295 as the fit of this model was even worse than the fit of the one-locus model. Because the
296 number of parameters in the two-locus unlinked model and the strong-linkage model are
297 identical, this is result is not merely driven by a difference in the number of parameters.
298 Additionally, these results do not depend on the precise relative D value used to define strong
299 linkage: even when we reduce relative D to 0.8, the unlinked model fits much better (ΔAIC =
300 33.9). Although two populations had significant support for the linkage models when
301 analyzed independently (Carter Co., OK and Cowley Co., KS; Table S2), the low number of
302 mimetic individuals (3 and 2, respectively) driving this result suggests that this outcome more
303 likely results from violations of Hardy-Weinberg assumptions or population/morph
304 classification error than from true linkage.
305 Our power simulations (Fig. 4) indicate that we had high discrimination power among

306 the one- and two-locus models, even with low sample sizes. In populations with few uniform
307 morphs, we found that even low sample sizes (*e.g.*, $N = 10$) yielded substantially different
308 distributions of predicted numbers of mimetic individuals under the two models; these
309 distributions became even more distinct with greater sample size (Fig. 4a). In populations
310 with mostly uniform morphs, we found that increasing sample size had no effect on
311 distinguishing among models due to low information content (Fig. 4a). For the true sample
312 sizes in the observed data, we found that this effect of information content held across all
313 morphs, with observed numbers of the three non-uniform morphs supporting the two-locus
314 model for populations with high information content, while the models are indistinguishable
315 for the populations with low informativeness (Fig. 4b). When all populations are simulated
316 under one- vs. two-locus models, we found that support for the two models are easily
317 distinguishable (note non-overlapping distributions in Fig. 4c). Thus, our results cannot be
318 explained by an asymmetry in power to distinguish between one- and two-locus models.
319 Finally, the observed ΔAIC value across all populations is close to the distribution of
320 simulated values under a two-locus model and very different from the ΔAIC distribution
321 simulated under a one-locus model. The correspondence between the two-locus and observed
322 values is especially striking given the numerous ways that real populations can deviate from
323 the strict assumptions of Hardy-Weinberg equilibrium present in the simulation model.
324
325 **Discussion**
326 Here we have provided evidence that pigmentation genes for the mimicry of coral
327 snake coloration are controlled by separate loci and assort in an unlinked fashion in natural
328 populations of mimics. The support for a multi-locus system is unsurprising given what was
329 already known or suspected about pigmentation in snakes, but the evidence against strong
330 linkage between loci in a Batesian mimic is an exciting and unexpected result given the

supergene linkage repeatedly found in butterfly mimicry (Joron et al. 2006; Joron et al. 2011; Kunte et al. 2014).

An important question that follows from our study is how generalized a lack of linkage between red and black coloration may be across coral snake mimics, especially in species that are not polymorphic. It is possible that *Sonora* represents an atypical loss of linkage, perhaps as part of a northward expansion beyond the range of model species after the last glacial maximum (Westphal et al. 2011), while linkage has been maintained across most other mimetic species. However, many populations of *S. semiannulata* remain sympatric with coral snakes (Cox and Davis Rabosky 2013), including as far south as the Mexican state of Jalisco (*e.g.*, MZFC-17246). Additionally, the frequency and quality of color polymorphism across both coral snakes and their mimics suggests that polymorphism is so common (Davis Rabosky et al. in review) and repeatable (always the same general set of color morphs) that *Sonora* likely represents a common condition in the evolutionary dynamics of coral snake mimicry.

There are two competing explanations for this lack of genetic linkage that require further testing. The first is that it is possible that the increased toxicity to the predator of lethal coral snakes relative to merely noxious butterflies may lead to wholly different dynamics and fitness costs in snakes during the evolution of mimicry from a cryptic ancestor (Pough 1988). In this case, it may be useful to theoretically assess toxicity thresholds in model species that are sufficient to prevent locus linkage in the mimics, perhaps akin to the approach taken by Turner (Turner et al. 1984) or palatability, abundance, and fitness. However, a second explanation suggests that coral snake mimicry may be more Müllerian than commonly described (Greene and McDiarmid 1981). Coral snake mimicry has traditionally been considered Batesian because there are no New World Colubrid snakes with the advanced venom delivery system found in Elapids (front fangs, muscular control of

venom glands, complex venom cocktails with high LD_{50} values, etc.) and no human or avian deaths have ever been reported from a New World Colubrid bite (Weinstein et al. 2011). However, many coral snake mimics, including *Sonora semiannulata,* have enlarged, grooved rear teeth in association with a Duvernoy's gland and some level of toxicity to their prey (Greene and McDiarmid 2005). The success of these modifications for predator defense rather than prey acquisition, and therefore the nature and consequences of the boundary between Batesian and Müllerian mimicry in snakes, is presumed to be inconsequential but remains untested.

In either case, the lack of strong linkage between the two color components in coral snake mimicry has major implications for the evolution of the mimetic signal over time. First, because simple recombination is all that is needed for snake coloration to drastically change, both the buildup and breakdown of the mimetic signal is likely to occur a) in discrete steps, and b) quite quickly. Two unexplained observations about the relative rates of color evolution among mimicry systems that may be impacted by genetic linkage is that the rates recovered across phylogenies of snakes (Davis Rabosky et al. in review) are several orders of magnitude faster than rates recovered in invertebrates (Kunte 2009; Oliver and Prudic 2010; Penney et al. 2012) and that reversals back to the cryptic state are much more common in snakes. Intriguingly, rates of color evolution across the genus *Heliconius*, which contains mostly Müllerian mimics with unlinked coloration components (Kronforst and Papa 2015), have been qualitatively inferred to be quite high (Kozak et al. 2015) and much more similar to those recovered for snakes. Thus, coral snake mimicry may emerge as a surprising demonstration of the greater importance of the genetic architecture behind the mimetic signal than the classification of a system as Müllerian or Batesian in predicting the macroevolutionary dynamics of mimetic signals.

Acknowledgements

We thank the following museums and their curatorial staff for access to specimens: Arizona State University, California Academy of Sciences, Museum of Southwestern Biology at the University of New Mexico, Museum of Vertebrate Zoology at the University of California, New Mexico State University, Sam Noble Museum at the University of Oklahoma, San Diego Natural History Museum, Sternberg Museum of Natural History at Fort Hayes State University, University of Arizona, University of Kansas, University of Texas, University of Texas at Arlington Amphibian and Reptile Diversity Research Center, and University of Texas at El Paso. We thank two anonymous reviewers and M. Alfaro for helpful comments on this manuscript, and Yann Surget-Groba for help sampling museum specimens. This work was supported by a National Science Foundation Postdoctoral Fellowship in Biology (DBI-0906046) to ARDR and by the University of Michigan. The authors declare no conflicts of interest.

Literature Cited

Bagnara, J. T. 1983. Developmental aspects of vertebrate chromatophores. American Zoologist 23:465-478.

Bechtel, H. B. 1978. Color and pattern in snakes (Reptilia, Serpentes). Journal of Herpetology 12:521-532.

Bechtel, H. B. and E. Bechtel. 1962. Heredity of albinism in the corn snake, *Elaphe guttata guttata*, demonstrated in captive breedings. Copeia 2:436-437.

Bechtel, H. B. and E. Bechtel. 1978. Heredity of pattern mutation in the corn snake, *Elaphe g. guttata*, demonstrated in captive breedings. Copeia 4:719-721.

Brodie, E. D., III and E. D. Brodie, Jr. 2004. Venomous snake mimicry. Pp. 617-633 *in* J. A. Campbell, and W. W. Lamar, eds. Venomous reptiles of the Western Hemisphere. Cornell University Press.

Campbell, J. A. and W. W. Lamar. 2004. The Venomous Reptiles of the Western Hemisphere. Cornell University Press, Ithaca, NY.

Charlesworth, D. 2016. The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. Evolutionary Applications 9:74-90.

Charlesworth, D. and B. Charlesworth. 1975. Theoretical genetics of Batesian mimicry II. Evolution of supergenes. Journal of Theoretical Biology 55:305-324.

Charlesworth, D. and B. Charlesworth. 2011. Mimicry: The hunting of the supergene. Current Biology 21:R846-R848.

Cox, C. L. and A. R. Davis Rabosky. 2013. Spatial and temporal drivers of phenotypic diversity in polymorphic snakes. The American Naturalist 182:E40-E57.

Cox, C. L., A. R. Davis Rabosky, and P. T. Chippindale. 2013. Sequence variation in the *Mc1r* gene for a group of polymorphic snakes. Gene 513:282-286.

Cox, C. L., A. R. Davis Rabosky, J. Reyes-Velasco, P. Ponce-Campos, E. N. Smith, O. Flores-Vilella, and J. A. Campbell. 2012. Molecular systematics of the genus *Sonora* (Squamata: Colubridae) in central and western Mexico. Systematics and Biodiversity.

Davis Rabosky, A. R., C. L. Cox, D. L. Rabosky, P. Title, I. A. Holmes, A. Feldman, and J. A. McGuire. in review. Coral snakes predict the evolution of mimicry across New World snakes. Nature Communications.

Franks, D. W. and T. N. Sherratt. 2007. The evolution of multicomponent mimicry. J Theor Biol 244:631-639.

Gamberale- Stille, G., A. C. Balogh, B. S. Tullberg, and O. Leimar. 2012. Feature saltation and the evolution of mimicry. Evolution 66:807-817.

Greene, H. and R. McDiarmid. 2005. Wallace and Savage: Heroes, theories, and venomous snake mimicry. Pp. 190-208. Ecology and Evolution in the Tropics: A Herpetological Perspective. University of Chicago Press.

Greene, H. W. and R. W. McDiarmid. 1981. Coral snake mimicry: Does it occur? Science 213:1207-1212.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, P. A. Wilkinson, C. Salazar, C. Davidson, R. Clark, M. A. Quail, H. Beasley, R. Glithero, C. Lloyd, S. Sims, M. C. Jones, J. Rogers, C. D. Jiggins, and R. H. ffrench-Constant. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477:203-206.

Joron, M. and J. L. Mallet. 1998. Diversity in mimicry: paradox or paradigm? Trends in Ecology & Evolution 13:461-466.

Joron, M., R. Papa, M. Beltrán, N. Chamberlain, J. Mavárez, S. Baxter, M. Abanto, E. Bermingham, S. J. Humphray, and J. Rogers. 2006. A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. PLoS Biology 4:e303.

Kikuchi, D. W. and D. W. Pfennig. 2012. A Batesian mimic and its model share color production mechanisms. Current Zoology 58.

Kikuchi, D. W., B. M. Seymoure, and D. W. Pfennig. 2014. Mimicry's palette: widespread use of conserved pigments in the aposematic signals of snakes. Evolution & Development 16:61-67.

Kozak, K. M., N. Wahlberg, A. F. Neild, K. K. Dasmahapatra, J. Mallet, and C. D. Jiggins. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. Systematic Biology 64:505-524.

Kronforst, M. R. and R. Papa. 2015. The functional basis of wing patterning in *Heliconius* butterflies: The molecules behind mimicry. Genetics 200:1-19.

Kunte, K., W. Zhang, A. Tenger-Trolander, D. H. Palmer, A. Martin, R. D. Reed, S. P. Mullen, and M. R. Kronforst. 2014. *doublesex* is a mimicry supergene. Nature 507:229-232.

Kuriyama, T., H. Misawa, K. Miyaji, M. Sugimoto, and M. Hasegawa. 2013. Pigment cell mechanisms underlying dorsal color-pattern polymorphism in the Japanese four-lined snake. Journal of Morphology 274:1353-1364.

Leimar, O., B. S. Tullberg, and J. Mallet. 2012. Mimicry, saltational evolution, and the crossing of fitness valleys. The Adaptive Landscape in Evolutionary Biology. Oxford University Press, Oxford, UK:259-270.

Mallet, J. and M. Joron. 1999. Evolution of diversity in warning color and mimicry: polymorphisms, shifting balance, and speciation. Annual Review of Ecology and Systematics 30:201-233.

Mappes, J. and R. V. Alatalo. 1997. Batesian mimicry and signal accuracy. Evolution 51:2048-2051.

Martin, A., R. Papa, N. J. Nadeau, R. I. Hill, B. A. Counterman, G. Halder, C. D. Jiggins, M. R. Kronforst, A. D. Long, W. O. McMillan, and R. D. Reed. 2012. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a *Wnt* ligand. Proceedings of the National Academy of Sciences 109:12632-12637.

Nijhout, H. F. 2003. Polymorphic mimicry in *Papilio dardanus*: mosaic dominance, big effects, and origins. Evolution and Development 5:579-592.

Pough, F. H. 1988. Mimicry of vertebrates: Are the rules different? The American Naturalist 131:S67-S102.

Reed, R. D., R. Papa, A. Martin, H. M. Hines, B. A. Counterman, C. Pardo-Diaz, C. D. Jiggins, N. L. Chamberlain, M. R. Kronforst, R. Chen, G. Halder, H. F. Nijhout, and W. O. McMillan. 2011. *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333:1137-1141.

Rosenblum, E. B. 2006. Convergent evolution and divergent selection: Lizards at the White Sands ecotone. The American Naturalist 167:1-15.

Rosenblum, E. B., H. Rompler, T. Schoneberg, and H. E. Hoekstra. 2010. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proceedings of the National Academy of Sciences of the United States of America 107:2113-2117.

Savage, J. S. and J. B. Slowinski. 1992. The colouration of the venomous coral snakes (family Elapidae) and their mimics (families Aniliidae and Colubridae). Biological Journal of the Linnean Society 45:235-254.

Turner, J. R. G. 1987. The evolutionary dynamics of Batesian and Muellerian mimicry: similarities and differences. Ecological Entomology 12:81-95.

Turner, J. R. G., E. P. Kearney, and L. S. Exton. 1984. Mimicry and the Monte Carlo predator: the palatability spectrum and the origins of mimicry. Biological Journal of the Linnean Society 23:247-268.

U.S. Census Bureau. 2011. USA counties data file downloads. Spreadsheet LND01.xls.

Weinstein, S. A., D. A. Warrell, J. White, and D. E. Keyler. 2011. "Venomous" Bites from Non-Venomous Snakes: A Critical Analysis of Risk and Management of "Colubrid" Snake Bites. Elsevier.

Westphal, M. F., J. L. Massie, J. M. Bronkema, B. E. Smith, and T. J. Morgan. 2011. Heritable variation in garter snake color patterns in postglacial populations. PLoS ONE 6:e24199-e24199.

Westphal, M. F. and T. J. Morgan. 2010. Quantitative genetics of pigmentation development in 2 populations of the Common Garter Snake, *Thamnophis sirtalis*. Journal of Heredity 101:573-580.

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

Figure Legends

Figure 1. Three dimensional bar plots of counts for four sympatric color morphs by population across the continental United States (a) mimetic, (b) banded, (c) striped, (d) uniform. All bars are plotted on the same scale, and the highest bar (Maricopa Co., AZ, striped morph) represents 150 individuals.

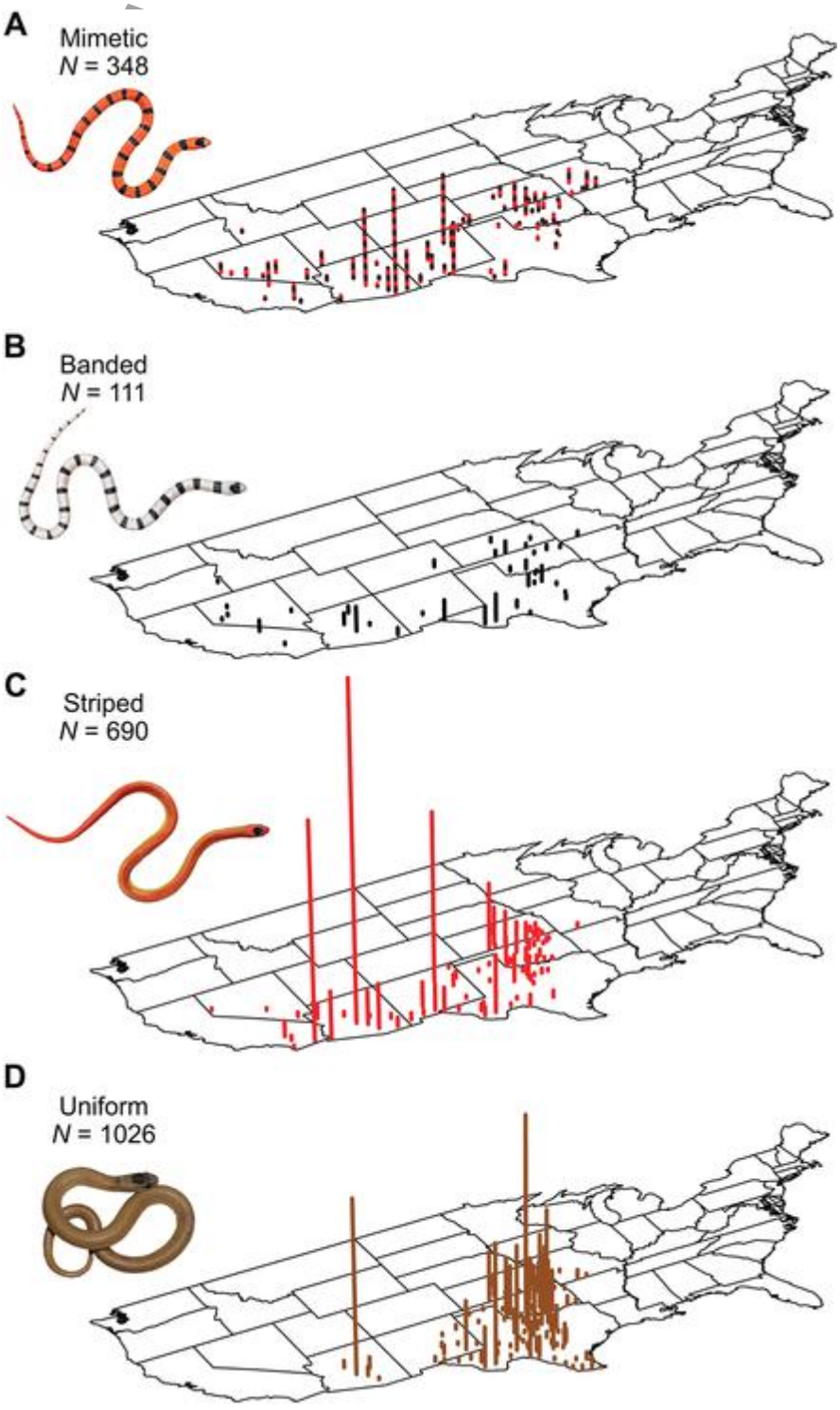

Figure 2. Mantel correlograms between phenotypic and geographic distance by class using populations with (a) at least N =12 and (b) at least N = 20 individuals sampled within a 20-year time window. Filled symbols indicate distance classes with significant correlation. Y-axis ranges are standardized to aid comparison among plots.

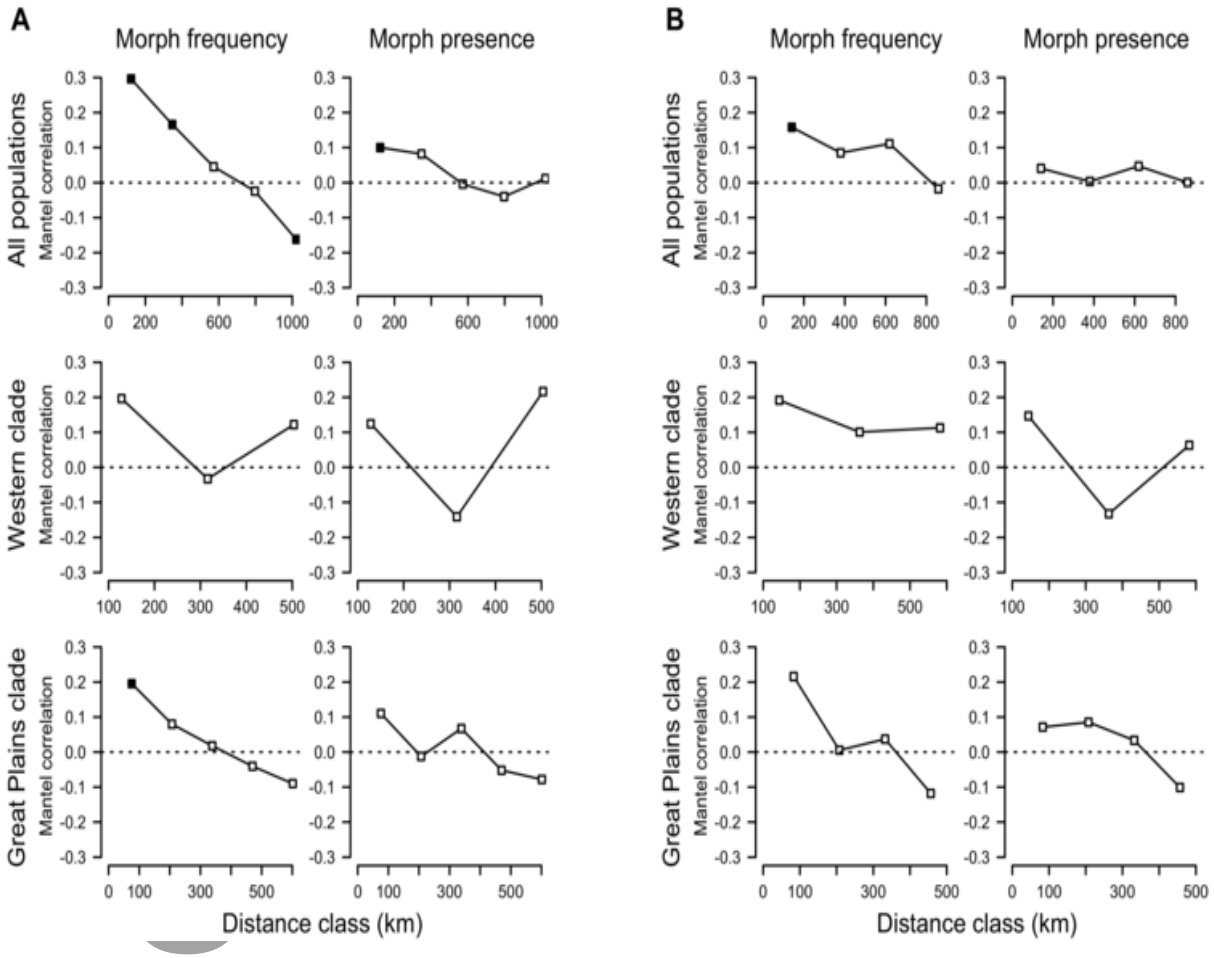

Figure 3. (a) Comparisons of model fit (N = 40 populations) show strong overall support for the two-locus model (sum of ΔAIC=154.36; note that 4 of the 5 highest scoring populations are from the Western clade). No populations show significant support for the one-locus model. Dashed reference line for equivocal model support is given at 0. (b) Model fit shows no relationship with level of polymorphism ($F_{1,38}$ = 1.53, P = 0.223). (c) There is no correlation between between the number of snakes collected within a population and the number of morphs detected ($F_{1,38}$ = 1.69, P = 0.202). Asterisked P-values are from linear models without the respective outlier values. Note the broken axis in each panel to allow the display of outlying values.

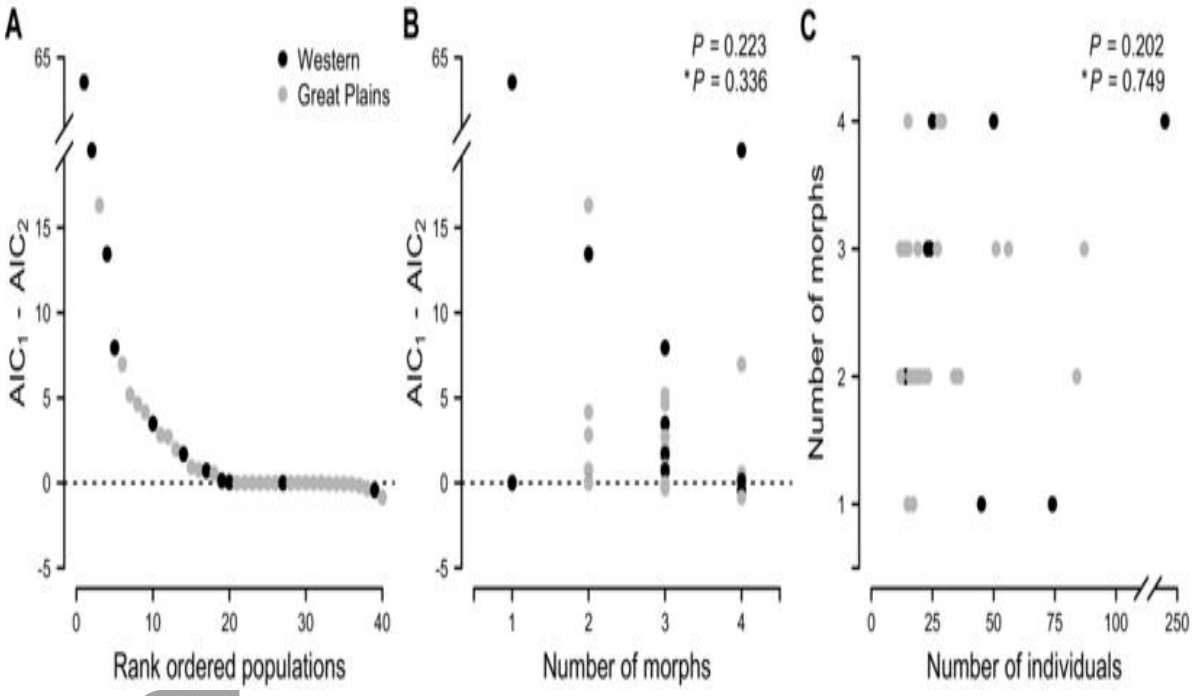

Figure 4. Simulated power analyses, with one-locus model predictions in light blue and two-locus predictions in dark blue. (a) Kernel density plots of populations with mostly non-uniform morphs (e.g., Cochise Co., AZ) show substantially different distributions of predicted numbers of mimetic individuals under the two models at all sample sizes, while increasing sample size has no effect on distinguishing among models in populations with high frequencies of the uniform morph (e.g., Stephens Co., TX) due to low information content. (b) For the true sample sizes in the observed data for the same two populations (N = 14 for Cochise Co. and N = 19 for Stephens Co.), the observed numbers of each morph (dashed lines) support the two-locus model for the population with high information content, while the models are indistinguishable for the population with low informativeness. (c) When all populations are simulated under one- vs. two-locus models (N = 1000 iterations), support for the two models are easily distinguishable (note non-overlapping distributions) and our observed ΔAIC value (dashed line) fits a two-locus model much better than a one-locus model.