Abstract
REPRODUCTIVE ISOLATION AND MIMETIC DIVERGENCE

IN THE POISON FROG *RANITOMEYA IMITATOR*

By

Evan Twomey

November 27, 2014

Director of Dissertation: Kyle Summers, Department: Biology
Interdisciplinary Doctoral Program in Biological Sciences

Understanding the process of speciation requires examination of various stages of its progress. This work focuses on the early stages of population divergence, where populations of a single species may show varying levels of incipient reproductive isolation. This work focuses on a species of poison frog, *Ranitomeya imitator*, that has undergone population divergence in color pattern to mimic different model species. This is an example of a Müllerian mimetic radiation, which has led to the establishment of four distinct morphs of *R. imitator* occurring in different geographic areas. There are two main goals of this research: (1) to examine to what extent (if any), different mimetic morphs are reproductive isolated, and (2) to examine factors influencing variation in the strength of between-morph reproductive isolation. To do this, the work focuses on sampling across three mimetic transition zones. In Chapter 1, by using a combination of color pattern quantification, landscape genetics, and mate choice experiments, results show that a mimetic shift in *R. imitator* is likely driving incipient speciation among two of these mimetic morphs. In Chapter 2, two additional mimetic transition zones are studied. Results suggest that these mimetic morphs are strongly differentiated in aspects of the mimetic phenotype, but show little neutral genetic divergence, a``nd random mating with respect to mimetic morph at the

transition zones, suggesting that mimetic divergence has failed to generate reproductive isolation in the two transition zones studied in this chapter. It is suggested that multifarious selection on both mimetic color-pattern and body size may be responsible for explaining variation in progress toward speciation. In Chapter 3, variation in mating calls among mimetic morphs is studied. While there are shifts in certain aspects of the mating call across two mimetic transition zones, phonotaxis trials indicate that call differences among morphs may not be responsible for generating reproductive isolation. Overall, the this dissertation work supports the idea that mimicry may be driving speciation in *Ranitomeya imitator*, and provides a basis for future work examining ecological speciation in this system.

REPRODUCTIVE ISOLATION AND MIMETIC DIVERGENCE IN THE POISON FROG *RANITOMEYA IMITATOR*

A Dissertation
Presented to

The Faculty of the Interdisciplinary Doctoral Program in Biological Sciences

The Brody School of Medicine, East Carolina University

In Association with the Department of Biology, Thomas Harriot College of Arts and Sciences

Submitted In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

Interdisciplinary Doctoral Program in Biological Sciences

By

Evan Twomey

November 27, 2014

REPRODUCTIVE ISOLATION AND MIMETIC DIVERGENCE IN THE POISON FROG *RANITOMEYA IMITATOR*

by

Evan Twomey

APPROVED BY:

DIRECTOR OF DISSERTATION:_____
KYLE SUMMERS, PhD

COMMITTEE MEMBER:_____
JEFF MCKINNON, PhD

COMMITTEE MEMBER:_____
TRIP LAMB, PhD

COMMITTEE MEMBER:_____
JASON BOND, PhD

COMMITTEE MEMBER:_____
ANDRADA IVANESCU, PhD

DIRECTOR, INTERDISCIPLINARY DOCTORAL PROGRAM IN BIOLOGICAL SCIENCES:

_____
TERRY L. WEST, PhD

DEAN OF THE GRADUATE SCHOOL:

_____
PAUL J. GEMPERLINE, PhD

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Reproductive isolation related to mimetic divergence in the poison frog *Ranitomeya imitator*

Abstract

In a mimetic radiation—when a single species evolves to resemble different model species—mimicry can drive within-species morphological diversification, and, potentially, speciation. While mimetic radiations have occurred in a variety of taxa, their role in speciation remains poorly understood. We study the Peruvian poison frog *Ranitomeya imitator,* a species that has undergone a mimetic radiation into four distinct morphs. Using a combination of colour pattern analysis, landscape genetics, and mate choice experiments, we show that a mimetic shift in *R. imitator* is associated with a narrow phenotypic transition zone, neutral genetic divergence, and assortative mating, suggesting that divergent selection to resemble different model species has led to a breakdown in gene flow between these two populations. These results extend the effects of mimicry on speciation into a vertebrate system and characterize an early stage of speciation where reproductive isolation between mimetic morphs is incomplete but evident.

Introduction

Elucidating the factors that promote population divergence and initiate speciation is key to understanding the evolution of biodiversity. Several studies have identified cases where divergent selection on ecologically relevant traits leads to partial or complete reproductive isolation or speciation (Chamberlain *et al.* 2009; Hatfield & Schluter 1999; McKinnon *et al.* 2004; Nosil 2012). Speciation is frequently studied by examining pairs of "good" species and identifying current reproductive barriers (Via 2009). However, these reproductive barriers may

have arisen after speciation was complete, whereas other, currently incomplete barriers may have arisen earlier and been important during initial population divergence (Coyne & Orr 2004). With the goal of investigating initial divergence, one can focus on the early stages of speciation, for example, populations of a single species showing incipient reproductive isolation.

Mimicry can drive phenotypic convergence between distantly related species, but can also drive within-species diversification. This has led to impressive morphological radiations in diverse taxonomic groups such as catfish (Alexandrou *et al.* 2011), millipedes (Marek & Bond 2009), snakes (Greene & McDiarmid 1981), bees (Plowright & Owen 1980), frogs (Symula *et al.* 2001), moths (Turner 1970), and, most famously, *Heliconius* butterflies (Bates 1862). In *Heliconius*, selection for Müllerian mimicry (mimicry between unpalatable species) has led to intraspecific divergence in wing patterns as different populations radiate into distinct mimicry rings (Bates 1862). These wing patterns are also used in mate choice, and morph-based assortative mating can arise as a byproduct of selection for wing mimicry (Kronforst *et al.* 2006) if accompanied by evolution of preferences. Studies of mimetic hybrid zones in *Heliconius* have yielded a range of examples highlighting the continuous nature of speciation. On one end of the continuum, hybrid zones can be narrow and characterized by strong assortative mating, neutral genetic divergence, and infrequent hybridization (Arias *et al.* 2008; Jiggins *et al.* 1997). On the other end of the continuum, hybrid zones can be wide, with little or no assortative mating, and with genetic divergence generally restricted to genomic regions controlling colour pattern differences (Mallet *et al.* 1990). There are, however, few examples of 'intermediate' hybrid zones, where distinct mimetic morphs show intermediate levels of genetic divergence and/or premating isolation (but see Arias *et al.* 2012). By identifying cases where speciation appears to have started, but is not yet complete, we can better understand how freely interbreeding

populations transition to reproductively isolated species.

Neotropical poison frogs (Dendrobatidae) are diurnal, toxic frogs known for their striking warning colours. A number of species display remarkable intraspecific diversity in colour pattern (Brown *et al.* 2011; Myers & Daly 1976; Silverstone 1975; Summers *et al.* 2003), although in most cases, the source of divergent selection among populations is unclear (Comeault & Noonan 2011; Hegna *et al.* 2012; Maan & Cummings 2009; Richards-Zawacki *et al.* 2013; Tazzyman & Iwasa 2010). In *Ranitomeya imitator*, intraspecific divergence in colour pattern is associated with selection for Müllerian mimicry (Yeager *et al.* 2012), which led to the establishment of four distinct mimetic morphs of this species in central Peru (Twomey *et al.* 2013). These morphs resemble three different model species (one of the model species, *R. variabilis*, has two morphs itself, both mimicked by *R. imitator*), and occur in different geographic regions, forming a "mosaic" of mimetic morphs. Where different morphs come into contact, narrow hybrid (or "transition") zones are formed (Twomey *et al.* 2013), similar to what has been observed in *Heliconius* butterflies. We have identified three such transition zones, making this study system useful for comparative analyses.

Here, we show that a mimetic shift in *R. imitator* is likely driving early-stage reproductive isolation among two of these mimetic morphs. We focus on the narrowest transition zone, which is found in the lowlands of north-central Peru and is formed between the "varadero" morph, which mimics *R. fantastica,* and the "striped" morph, which mimics the lowland morph of *R. variabilis* (Yeager 2009) (Fig. 1-1 and Fig. 1-S1). Our sampling along a transect crossing this transition zone reveals that there is a shift in several aspects of colour pattern in *R. imitator*, including dorsal colour (yellow to orange), arm colour (pale greenish-blue to orange), leg colour

(pale greenish-blue to navy blue), and dorsal pattern (uniform longitudinal stripes to colouration concentrated around the head). These shifts correspond to the colour pattern of each model species (Fig. 1-1 and Fig. 1-S2), and are therefore likely involved in mimicry. Analyses of colour pattern clines show that the transition zone is approximately 1–2 km wide and composed of phenotypic intermediates. Landscape genetic analyses indicate that neutral genetic divergence between morphs is primarily associated with divergence in mimetic colour pattern, rather than geographic distance, suggesting that mimetic divergence has reduced gene flow between morphs. Using mate choice experiments, we find evidence for assortative mating in one of the mimetic morphs, however, this mating preference is only present near the transition zone, consistent with reproductive character displacement. Taken together, these results suggest that mimetic divergence in *Ranitomeya imitator* has led to a breakdown in gene flow between these two populations, potentially facilitated by assortative mating.

Results

Colour pattern clines

Selection for different mimetic morphs across geographical areas should cause differentiation in mimetic traits. At the interface between distinct mimetic morphs, traits subject to divergent selection are expected show a sigmoidal pattern of variation across this zone of mixing (Barton & Hewitt 1985). To quantify colour pattern variation along the mimicry transect, we used a pattern variables in *R. imitator*. Transect variation in three of these colour pattern variables (head colour, body colour, and leg pattern) was best described by a linear model (Fig. 1-2), suggesting gradual spatial change. However, due to our sampling pattern, we cannot rule out the possibility

of a sigmoidal cline with a displaced centre for these colour pattern variables. The remaining three colour pattern variables (arm colour, leg colour, and body pattern) were best described by a sigmoidal model (Fig. 1-2), suggesting that these aspects of the colour pattern are under divergent selection. If multiple aspects of the colour pattern are involved in mimetic resemblance, then shifts in traits should coincide geographically. We tested for cline coincidence among arm colour, leg colour, and body pattern by comparing Akaike weights (w_i) between two models: one where cline centre is constrained to a single parameter shared across all three datasets, and one where centre is unshared. A common centre was found for all three colour pattern variables without a significant reduction in model fit (w_i shared centre model = 0.841; w_i unshared centre model = 0.159), indicating coincidence among the three colour pattern clines. The point estimate for the shared centre parameter was 0.54 km (i.e., 0.54 km N from the *a priori* estimated centre), corresponding to 1.25 km N from the village of Varadero (Fig. 1-S1). An alternative explanation for coincident clines is recent secondary contact between divergent populations (see below for discussion of primary vs. secondary contact).

In a tension zone model (Barton & Hewitt 1985), where divergent selection is opposed by dispersal, the width of the cline reflects a balance of selection (which narrows a cline) and dispersal (which widens a cline). Differences in cline widths among different traits may be due to differences in the strength of selection on loci underlying those traits. If different traits are clines than those under weaker selection. We tested for a common cline width (concordance) among all six colour pattern variables. A common width could not be found among all six variables without a reduction in model fit (w_i shared width model = 0.273; w_i unshared width model = 0.727), indicating that some colour pattern variables show non-concordant widths. This was due to the inclusion of the three non-sigmoidal variables, as it was possible to fit a common

width of 2.27 km among the three variables showing a sigmoidal pattern of variation (w_i shared width model = 0.747; w_i unshared width model = 0.253). Cline width should be primarily a function of selection strength (assuming constant dispersal), so the evidence that these three clines can be constrained to a common width suggests equivalent strength of selection on arm colour, leg colour, and body pattern. This could also suggest a common genetic basis or linkage among all three traits, although colour and pattern elements in dendrobatids are likely controlled by different genes (Summers *et al.* 2004).

Landscape genetics

Reduced gene flow between adaptively diverged populations (isolation-by-adaptation; IBA) is a key prediction of ecological speciation (Nosil 2012). This results in a positive correlation between adaptive ecological divergence and genetic differentiation among populations after controlling for the effect of isolation-by-distance (IBD) (Wang & Summers 2010). Results from the Structure analysis (Fig. 1-3a) indicate the presence of three genetic groups within the study area. One of these groups (Fig. 1-3b) is associated with an allopatric population, while two of the groups (Fig. 1-3c, d) form a sharp break at the mimetic transition zone. These latter two groups still show some evidence of genetic exchange, as there were a few individuals with a striped colour pattern but a varadero genotype, and vice-versa (Fig. 1-3). The narrow genetic cline is also characterized by a peak in linkage disequilibrium (Supplementary Fig. 1-3), further suggesting a barrier to gene flow among the two mimetic morphs. The coincidence of genetic clines and colour pattern clines was supported by a factorial correspondence analysis, where the cline centre on the first major axis (0.31 km) is almost identical to the shared colour pattern cline

centre (0.54 km), supporting the hypothesis that a shift in mimicry has led to a breakdown in gene flow among mimetic morphs. Using a causal modelling framework, the best-supported hypothesis was one where colour pattern distance (IBA), but not geographic distance (IBD), was correlated with genetic distance among populations. Multiple matrix regression (Wang 2013) yielded similar results, except that both colour pattern distance ($r^2 = 0.427$, $P = 0.006$) and geographic distance ($r^2 = 0.230$, $P = 0.001$) were accounted as significant predictors of genetic distance. However, the correlation coefficient for colour pattern distance is nearly twice that of geographic distance, indicating that IBA is a stronger determinant of among-population genetic divergence than is IBD. An alternative interpretation for these results is mimetic divergence in allopatry followed by secondary contact. This could explain the neutral genetic divergence among these two populations, however, one would expect the microsatellite cline to be wider than the observed 0.54 km unless contact happened very recently (see below).

Mate choice experiments

One potential mechanism for a breakdown of gene flow between adaptively diverged populations is morph-based assortative mating (Gavrilets 2004). To address the role of assortative mating, we conducted triad mate choice experiments in which we introduced two females (one of each morph) into the terrarium of a given male, and measured the amount of courtship time between the male and female. This is equivalent to a mutual choice test, which is appropriate here as *R. imitator* is monogamous (Brown *et al.* 2010), and therefore both sexes should be choosy. We tested preferences in three populations: striped-allopatric, striped-transition, and varadero, allowing us to address two questions (1) whether courtship preferences differ among between the

striped-transition and varadero populations, and (2) whether courtship preferences differ among the two populations of the striped morph. Using generalized linear mixed models, we found an overall significant effect of male origin ($\chi^2_{1,55} = 16.518$, $P = 0.00026$), indicating that mate preferences were significantly different across populations. A post-hoc test revealed that the preferences in the striped-allopatric and varadero populations were not significantly different ($\chi^2_{1,22} = 3.096$, FDR-adjusted $P = 0.078$), with neither population showing a significant preference (Fig. 1-4). However, preferences between the striped-transition and varadero populations were significantly different ($\chi^2_{1,33} = 11.986$, FDR-adjusted $P = 0.00161$), mainly due to the striped-transition population's preference towards its own morph (Fig. 1-4), which indicates that mating preferences have diverged between these two populations across the transition zone. Finally, preferences between the striped-allopatric and striped-transition populations were significantly different ($\chi^2_{1,29} = 9.748$, FDR-adjusted $P = 0.00269$), suggesting that mating preferences in the striped-transition population are stronger at the mimetic transition zone.

Bioacoustics

During our sampling, there were apparent differences in the advertisement calls of the striped and varadero morphs, which could represent a potential premating isolating mechanism between the two morphs. To determine whether the pattern of call variation coincided with the mimetic transition zone, we recorded the calls of *R. imitator* across the sampling transect. The call of *R. imitator* is a short, musical trill of 0.44–1.07 seconds, with trills (or "notes") repeated roughly every 4–20 seconds, and a dominant frequency of 4710–5660 Hz. Each note is composed of 16–32 pulses, with an average pulse rate of 24–30 pulses per second (Brown *et al.* 2011). Note length was negatively correlated with temperature ($r^2 = 0.115$, $P = 0.006$), and pulse rate was positively correlated with temperature ($r^2 = 0.259$, $P < 0.001$). To account for this, we standardized each of the three bioacoustic variables by calculating regression residuals against temperature. After temperature standardization, two bioacoustic variables showed a sigmoidal rather than linear pattern of variation across the transect (note length: Akaike weight (w_i) linear model = 0.032, w_i sigmoidal model = 0.967; dominant frequency: w_i linear model = 0.006; w_i sigmoidal model = 0.992). The point estimates of cline centre were similar (note length centre = -0.14 km; dominant frequency centre = -0.41 km) indicating that the shift in these two call parameters occurs in roughly the same geographic location. Furthermore, the estimated cline centres both occur very close to the estimated colour pattern and microsatellite cline centres (within < 1 km), indicating that the shift in call characteristics occurs in the same place as the shift in colour pattern and microsatellites. For pulse rate, the linear model was favoured (w_i linear model = 0.830, w_i sigmoidal model = 0.138), indicating a smooth, rather than abrupt, transition across the putative transition zone. To derive a single metric describing call variation, we used a

linear discriminant analysis to derive a discriminant score where the two groups for classification were defined as the populations on the endpoints of the transect (i.e., populations 1 and 10 in Table 1-S1). Both note length and dominant frequency contributed substantially to the discriminant function, whereas pulse rate did not (standardized canonical discriminant function coefficients: dominant frequency = 1.343; note length = -1.189; pulse rate = 0.162). This metric showed a sigmoidal pattern of variation with similar cline centre and width as observed in the colour pattern metrics (Fig. 1-2, Table 1-S2).

Discussion

Our mate choice trials found that the preferences in the two striped populations we studied were stronger in the striped-transition zone population relative to the striped-allopatric population, a pattern consistent with reproductive character displacement (RCD). However, our experimental design is limited in terms of inferring RCD given that we only tested three populations, and therefore, assuming that mating preferences vary among populations, there is a one in three chance that the strongest preference will be in striped-transition population. A much more robust test of RCD would involve testing multiple populations to determine whether contact among morphs explains variation in mating preferences. Patterns of enhanced mating preferences in areas of contact have, however, been observed in mimetic *Heliconius* butterflies. For example, *H. melpomene* populations that are sympatric with *H. cydno* display stronger mating preferences relative to allopatric populations (Jiggins *et al.* 2001). In another example, mating preferences in both *H. cydno* and *H. pachinus* are much stronger in sympatry than allopatry (Kronforst *et al.* 2007). One explanation for this pattern is reinforcement, where mate preferences are

strengthened in zones of sympatry to avoid producing unfit hybrids. However, several other processes can result in a pattern of enhanced mate preferences in zones of sympatry (e.g., differential fusion hypothesis, "noisy neighbour" hypothesis; see ref. (Coyne & Orr 2004) for review). In this case, as non-mimetic hybrids may suffer fitness costs if they experience higher predation rates (McMillan *et al.* 1997), adaptations to avoid cross-morph matings are expected to be favoured by selection.

In addition to a shift in colour pattern and microsatellites at the transition zone, we found a shift in body mass and certain aspects of the advertisement call. Striped frogs south of the transition zone tend to have a smaller body size and a shorter, more highly pitched call compared to the varadero morph north of the transition zone. For both body size and advertisement calls, variation along the transect is best described by a sigmoidal cline with centres coinciding with the colour pattern and genetic clines (Table 1-S4), further supporting the existence of a transition zone. This also supports the possibility for secondary contact, where clines are expected to be congruent for multiple traits. One possible explanation for the shift in body size is that *R. variabilis*, the model species of the smaller, striped morph of *R. imitator*, is smaller than *R. fantastica*, the model species of the varadero morph (*R. variabilis* mass: = 0.52 g, $n = 3$; *R. fantastica* mass: = 0.68 g, $n = 4$). Thus, size could represent a mimetic adaptation. As our experiments did not address the specific cue used in mate choice, the roles of colour pattern, body size, and advertisement calls in mediating mate choice in this system should be investigated further.

As we have mentioned above, secondary contact among differentially adapted populations could give rise to many of the observed cline patterns. One plausible scenario here would be

mimetic divergence in allopatry, followed by secondary contact. Determining whether hybrid zones are the result of primary or secondary contact without historical evidence is difficult (Barton & Hewitt 1985). However, secondary contact with neutral diffusion is unlikely given our dispersal estimate in *R. imitator* of 0.095 km per generation (see Supplementary Methods for details on dispersal calculations). The cline created by secondary contact with subsequent neutral diffusion would exceed the observed overall cline width (0.97 km, see Table 1-S4, model D) in only 17 generations, or approximately 11 years. Considering secondary contact, a more likely scenario is that the cline is maintained by some isolating barrier. In either case (primary or secondary contact), the cline is associated with a shift in mimicry, and may be maintained, at least in part, by assortative mating. Overall, the existence of a narrow cline, as well as moderate genetic divergence between morphs (F_{ST} between mimetic morphs is 0.065–0.077), suggests that mimetic divergence may be playing a key role driving early-stage speciation in a vertebrate system.

Methods

Data availability

Colour pattern data, advertisement call data, mate choice data, and the full microsatellite dataset is available at Dryad (doi:10.5061/dryad.rd586).

Sample collection and transect description

For colour pattern analyses, we sampled a total of 127 *Ranitomeya imitator* from 15 localities in the department of Loreto, Peru. Ten of these localities (localities 1–10 in Fig. 1-S1) lie on a

rough north-south transect 40 km in length, running from the village of Micaela Bastidas in the south to 7 km N from the village of San Gabriel de Varadero in the north. We sampled an additional five localities off the transect but still relevant for inferring the spatial arrangement of the two focal morphs of *R. imitator*. For genetic analyses, we sampled 136 *R. imitator* from 10 localities. Tissue samples for genetic analysis (toe-clips) were taken with sterile surgical scissors and preserved in 96% ethanol prior to extraction. In most cases, both tissue samples and colour pattern measurements were taken from each frog, although there were some localities where only genetic data were collected or only colour pattern data were collected (see Table 1-S1 for details). Additionally, we took colour pattern measurements from the two putative model species: 7 *R. variabilis* from Pongo de Cainarachi (representative of the typical lowland *R. variabilis* morph) and 7 *R. fantastica* collected from San Gabriel de Varadero.

Because the transect is not perfectly linear, we calculated transect position as straight-line distance from the putative transition zone centre, with localities south of this point given a negative sign and localities north of this point given a positive sign. The initial centre-point (lat/long: -5.70653° -76.41427°) used in these calculations was estimated from field observations where an apparent shift in colour pattern occurred. Therefore, instances where the estimated cline centre from nonlinear regression was close to zero indicate a close fit to our field observations. Cline centre estimates with a negative sign indicate the inferred cline centre to be south of the initial centre-point, whereas positive values indicate the cline centre to be north of the initial centre-point.

Colour and pattern quantification

To quantify frog colour, we measured the spectral reflectance at specific points on the dorsal surfaces of the mimic species (*R. imitator*) and both model species (*R. variabilis* and *R. fantastica*). Two measurements were taken on the head (right and left sides), four on the body (right and left sides of mid-body and rump), and two on the legs (dorsal surface of right and left thighs). Reflectance measurements were taken with an Ocean Optics USB4000 spectrometer with an LS-1 tungsten-halogen light source and Ocean Optics SpectraSuite software. A black plastic tip was used on the end of the probe so that measurements were always taken at a distance of 3 mm from the skin and at a 45° angle. White standards were measured for every other frog using an Ocean Optics WS-1-SL white reflectance standard to account for lamp drift. Spectral data were then processed in Avicol version 6 software (Gomez 2006) using Endler's segment model (Endler 1990) calculated between 450–700 nm. This model calculates brightness (Qt), chroma (C), hue (H), and two Euclidean coordinates representing position in a 2D colour space: blue-yellow axis position (MS), and red-green axis position (LM). Measurements within body regions (head, body, and legs) were averaged. In addition to the spectrometer measurements, we measured upper-arm colouration using the colour-picker tool (set to a 5x5 pixel average) in Adobe Photoshop CS4 on dorsal photos of each frog, recording the average intensities of red, green, and blue channels on two points on each upper arm. Photos were taken on a white background with a Canon Rebel XS DSLR with a Canon EF 100 mm macro lens and the camera flash.

We quantified frog pattern by a collection of local image descriptors. The descriptors were automatically extracted from images of every individual and collected in a feature matrix. Three

types of descriptors were extracted: a colour/non-colour ratio, gradient orientation histograms and shape index histograms (Dalal & Triggs 2005; Koenderink & van Doorn 1992; Larsen *et al.* 2014). Collectively, these capture zeroth, first and second order image structure. A spatial pooling scheme was used to separately collect information at four interest points: left leg, right leg, lower and upper back. At each of these interest points, pattern variation occurs on a distinct scale, wherefore the descriptors were extracted according to a scale space formulation (Lindeberg 1996). Colour/non-colour ratios were extracted for every interest point on a single scale; gradient orientation histograms for every interest point on two different scales and two orientation bins (horizontal and vertical), and shape index histograms were only extracted for the legs on two scales in five bins equidistantly spaced between $-\pi/2$ and $\pi/2$. This summed up to a total of $4 \cdot (1 + 2 \cdot 2 + 2 \cdot 5) = 60$ features per individual.

To reduce the multivariate colour and pattern data to a single descriptive metric per body region, we used kernel discriminant analysis (Mika *et al.* 1999), where the two model species (*R. variabilis* and *R. fantastica*) represented the training groups used for classification. This procedure assigns a discriminant score to each *R. imitator* individual on the basis of their similarity to either model species, and thus can be thought of as a "mimicry score". The analysis can be constrained to include only subsets of the variables to derive a metric for different body regions, e.g., leg colour variation in *R. imitator*. Kernel based analysis is implicitly capable of estimating non-linear effects, making it more suitable for non-normally distributed features, such as the colour metrics output from Avicol. Using this procedure, we derived colour metrics for four body regions (head, body, legs, and arm), and pattern metrics for two body regions (dorsum and legs). For additional details on kernel discriminant analysis, see Supplementary Methods.

Cline analysis

To describe clinal variation in colour pattern elements (as well as average male mass, advertisement call, and microsatellites; see Supplementary Methods), and in particular to estimate cline width, we performed nonlinear regression using a 4-parameter sigmoid tanh function

where *c* is the centre of the cline, *w* is the cline width, and *ymax* and *ymin* are the maximum and minimum trait values (i.e., the trait values at the tails of the cline). This uses the cline model of Szymura & Barton (1986) except that the minimum and maximum trait values are free to take on any value. Parameter searches were done using the solver function in Excel using a least-squares optimality criterion. Solver was run using the GRG nonlinear algorithm with the following settings: convergence = 0.0001; central derivatives; multistart on; population size = 100.

To evaluate whether the data were adequately described by a "flat" model (constant trait value across the transect), or a linear model (smooth transition), we fit these models, in addition to the sigmoid model, as candidate models. A flat model consists of a single parameter (population mean) defined as the grand mean of all individuals and is invariant across the transect. A linear model has two parameters, slope and y-intercept, and was fit with linear regression. To evaluate which of the three models (1-parameter flat, 2-parameter linear, or 4-parameter sigmoid) was a better fit to the data, we calculated ΔAICc and Akaike weights (w_i) for

each model (methods following Burnham & Anderson 2002) using the residual sum of squares divided by the sample size as the likelihood criterion.

Confidence intervals on parameter estimates were calculated using a Monte Carlo resampling method using the software GraphPad Prism. Briefly, this procedure involves the following steps: First, data are simulated for each observed x-value using best-fit parameters of the observed cline, with scatter added by drawing data points randomly from a hypothetical normally distributed population with a standard deviation equal to the observed $S_{y.x}$ (standard deviation of the residuals). A cline is then fit to the simulated data and best-fit values of each parameter are recorded. This process is then repeated for a number of iterations, each time generating a new simulated dataset, fitting a cline to that dataset, and recording parameter estimates. By using observed x-values (i.e., actual sampling locations) and observed residual variation, we are essentially simulating the distribution of cline parameter estimates we might observe if we resampled the entire transect multiple times. Simulations were run for 10,000 generations and 95% confidence intervals were calculated on the simulation parameter estimates.

We tested for common centre (coincidence) and common width (concordance) among clines using global nonlinear regression. This method compares model fit when certain parameters are shared vs. unshared among different variables. Under a scenario where all variables shift at the same position on the transect, a common centre parameter can be fit across all measured variables without a substantial reduction in model fit. This is expected, for example, in a scenario where there is a shift in the selective regime for one mimetic colour pattern in one area vs. another area along the transect. If the width of a cline on a phenotypic trait is a function of the strength of selection on that trait, a global width parameter may be expected when selection acts

at similar strength on all traits; however, if selection is strong on some traits and weak on others, this will cause different cline widths and thus a common width parameter will not adequately fit all the data. A common width may also be expected when linkage disequilibrium is high in the centre of the hybrid zone, as is observed here (Fig. 1-S3). We evaluated four models representing different combinations of shared and unshared parameters (Table 1-S4): (a) no constraint (each variable with unique centre and width), (b) centres constrained to be equal, width unconstrained, (c) width constrained to be equal, centres unconstrained, (d) centre constrained to be equal and width constrained to be equal. Best-fit shared parameter searches were done by fitting shared parameters to all datasets simultaneously, while unshared parameters were free to take on unique values for each dataset. Goodness-of-fit was assessed by calculating ΔAICc for each model.

Mate choice experimental design and analysis

To test for morph-based mating preferences, we conducted triad mate choice experiments in which we introduced two females (one of each morph) into the terrarium of a given male for one hour, and measured the amount of courtship time between the male and the varadero female vs. the male and the striped female. For details on the populations we sampled, as well as details on husbandry and experimental protocols, see Supplementary Methods. In *Ranitomeya imitator*, courtship is usually initiated when a calling male approaches a female. The female may then either reciprocate by following the male to a suitable oviposition site while the male continues calling, or show no interest (Brown *et al.* 2008a). The conditions of our mate choice experiments allowed these behaviours to take place in that a male was free to initiate courtship with either female, and the female was free to reciprocate interest or not. Male initiation of courtship is

readily observable in captivity as males (a) initiate a courtship call (shorter and more rapid than an advertisement call), and/or (b) begin to move in a staccato-like walk, often moving their rear legs erratically. Thus, when a male engaged in either of these behaviours in the vicinity of a female, this marked the initiation of courtship. Courtship was deemed to have ended under the following conditions: (a) the male, having initiated courtship, moves away and the female does not pursue, or (b) the female moves away and the male does not pursue. A trial was excluded when one or both females remained hidden in the gravel during the trial, thus precluding any possibility for choice. Using these criteria, we measured in each trial the total amount courtship between the male and the varadero female vs. the male and the striped female.

Typically, in this kind of experimental setup, the two females to be introduced to the male would be matched for mass to control for any confounding effects of mass on preference. However, in this case, matching for mass was not feasible because the varadero population is larger than either striped population (striped allopatric females = 0.56 g, s.d. = 0.05 g, $n = 43$; striped-transition females = 0.49 g, s.d. = 0.04 g, $n = 18$; varadero females = 0.61 g, s.d. = 0.05 g, $n = 30$), severely limiting the number of potential female combinations (for example, only the four heaviest striped-transition females would have qualified to be matched with the six smallest varadero females). To control for differences between females, we used a paired-samples experimental design whereby a given pair of females was presented to a male of each morph. This design therefore addresses the question of how changing male morph type alters courtship probabilities when female identity is held constant.

To analyse mate choice data, we used generalized linear mixed models (GLMM) using the glmmADMB package (Skaug *et al.* 2011) in R version 3.0.2 (R Development Core Team 2005)

with an underlying beta-binomial error distribution to test if the time males spent courting each female morph was influenced by male population origin. "Pair ID" (i.e., a unique identifier assigned to each female pair) was used as a random effect to account for the paired-samples experimental design. Following a significant result of the overall GLMM, we conducted post-hoc tests to determine: (1) whether courtship preferences differ among morphs (specifically, comparing striped-allopatric to varadero, and striped-transition to varadero), and (2) whether courtship preferences differ among populations of the same morph (comparing striped-allopatric to striped-transition). Post-hoc tests were run using the same GLMM procedure described above except that we restricted the analysis to the populations of interest. To account for multiple comparisons, we adjusted p-values using a false discovery rate protocol (Benjamini & Hochberg 1995) accounting for the fact that we conducted three post-hoc tests. The protocols we used were approved by East Carolina University's Institutional Animal Care and Use Committee (AUP permit #D225a) prior to the start of this study.

Landscape genetics

We used a causal modelling framework (Cushman *et al.* 2006; Wang & Summers 2010) to test specific hypotheses of how geographic distance and colour pattern differentiation between populations are associated with genetic distance. In landscape genetics, causal modelling uses Mantel tests and partial Mantel tests to evaluate alternative models explaining genetic distance between populations. Each model carries a set of statistical predictions; the model with all its predictions upheld is the one with the strongest support. In an isolation-by-distance scenario, a significant correlation is expected between a geographic distance matrix (independent variable

set) and a genetic distance matrix (dependent variable set). By using partial Mantel tests, the correlation between two dissimilarity matrices can be quantified while controlling for the effect of a third covariant matrix. For example, a partial Mantel test between colour pattern distance and genetic distance with geographic distance as a covariant matrix tests for the correlation between colour pattern distance and genetic distance after the effects of geographic distance are removed. We used one measure of geographic distance (Euclidean distance), one measure of genetic distance (Nei's *D*), and one measure of colour pattern distance (difference in discriminant score, see below) to test three models of genetic isolation. Details on each model and their associated predictions are given in Table 1-S3.

Causal modelling is often used to test how various landscape factors influence genetic isolation among populations (Cushman *et al.* 2006; Richards-Zawacki 2009). This can be useful for species occupying heterogeneous habitats, where straight-line distance between populations may not be the most likely corridor of gene flow. However, in our case, all populations of *R. imitator* are from a contiguous lowland rainforest habitat without any geographic barriers separating populations. The only two substantial barriers in this area, the Huallaga River and the Cordillera Escalera Mountains, are located to the east and south, respectively, of all sampling sites. Therefore, for the geographic distance matrix, we simply calculated pairwise straight-line distance between populations. For the genetic distance matrix, we calculated Nei's genetic distance (*D′*) between all pairs of populations in GenAlEx version 6.5 (Peakall & Smouse 2006). To generate a colour pattern distance matrix, we calculated pairwise differences in discriminant score from the kernel discriminant function analysis. Thus, because this analysis takes into account features of the model species, it can be thought of as a composite difference in mimetic

colour pattern. In addition to causal modelling, we used a multiple matrix regression method (Wang 2013) to quantify the relative effects of geographic distance and colour pattern distance on genetic distance. This method is similar to Mantel and partial Mantel tests but incorporates multiple regressions, such that the relative effects of two or more predictor variables on genetic distance can be quantified, as can the overall fit of the model. Multiple matrix regression was run with 10,000 permutations using the R script provided in Wang (2013).

For details on microsatellite genotyping protocols, Structure analyses, and factorial correspondence analyses, see Supplementary Methods.

Bioacoustics

We recorded the advertisement calls of 58 *R. imitator* from eight localities. These localities are all located on the colour pattern/microsatellite transect spanning the transition zone and thus can be used for cline analysis. Calls were recorded on a Marantz PMD660 solid state recorder with a Sennheiser ME 66-K6 microphone and analysed in Raven Pro version 1.3 (Charif *et al.* 2008). We quantified advertisement calls by measuring the following parameters: note length (measured from the start of the first pulse to the end of the last pulse), pulse rate (defined as pulse count divided by note time), and dominant frequency (the frequency at which peak amplitude is registered). For each male, a recording generally consisted of several notes. Measurements were always taken on at least 3 notes and then averaged for each male. As temperature is known to have a strong influence on certain aspects of amphibian calls (Myers & Daly 1976), we took temperature measurements alongside each call recording in the same microhabitat as the calling male, which we then used to standardize call parameters by calculating regression residuals

against temperature. We fit clines to each call parameter separately, plus a "composite" call score which we calculated using a linear discriminant analysis.

Supplementary Methods

Kernel discriminant analysis supplement

We regularized the kernel discriminant analysis solution with λ times the identity matrix and used a Gaussian kernel with width γ (Table 1-S5). These parameters were selected to minimize intra-location variance of the *R. imitator* discriminant scores while keeping them within the span of the two model species' discriminant scores. The formulation of kernel discriminant analysis unfortunately makes it impossible to inspect, e.g., the loadings as one would do in linear discriminant analysis to determine variable importance. However, the mutual information (MI) between each of the original variables $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ and the discriminant scores (metric) $\mathbf{z}$ can be inspected to elucidate which variables are relevant for discrimination between the model species. Mutual information is defined as

where *H(x)* and *H(x,y)* are marginal and joint entropies respectively. Parzen window estimates of MI between the variables used to define each of the six metrics, and the resulting metric can be seen in Figures 1-S4 and 1-S5. The values $MI(\mathbf{x}_i, \mathbf{z})$ are normalized such that $MI(\mathbf{x}_i, \mathbf{x}_i) = MI(\mathbf{z}, \mathbf{z}) = 1$. For colour data (Fig. 1-S4), we see that for head colour, red-green axis (LM), chroma (C), and hue (H) have relatively high MI, whereas for body colour blue-yellow axis (MS) and chroma have high MI. For leg colour, brightness (Qt) had the highest MI. For arm colour, blue channel intensity (B) had the highest MI. For pattern data (Fig. 1-S5), we see that for dorsal pattern, most

variables associated with pattern variation on the lower dorsum (prefix c3) show high MI, whereas variables associated with pattern variation on the head (prefix c4) show lower MI. For leg pattern, both shape index histograms (sih) and gradient orientation histograms (goh) showed high MI on each leg (c1 and c2), although this depended mainly on the associated extraction parameters.

Mate choice: Animal collection, husbandry, and protocols

We used *Ranitomeya imitator* from three populations for mate choice experiments (GPS points given in Table 1-S1):

(1) *Striped-allopatric (site 1; Fig. 1-S1)* – Frogs from this population are mimics of *R. variabilis*, with yellow pinstripes along the dorsum and a pale greenish or bluish reticulum on the legs and venter. These frogs were collected from near Micaela Bastidas, a village 33 km to the southeast of the transition zone and represent the 'striped-allopatric' population in our analyses.

(2) *Striped-transition (sites 3, 4, and 5; Fig. 1-S1)* – These frogs are also *R. variabilis* mimics, with yellowish-orange stripes on the dorsum and pale greenish legs. Due to the difficulty of collecting striped frogs in this area because of deforestation, we collected frogs from three different sites. Two of these sites (Varadero Stream and Varadero Bridge on Table 1-S1) are on the north side of the Paranapura river and one site (Varadero South Bank) is on the south side. We treated these three collecting sites as a single population for mate choice trials because the Paranapura river is small (50-80 m across in most places) and appears to present no noticeable barrier to gene flow. For example, the genetic distance (Nei's D') between Varadero South Bank and Varadero Stream (same morph, opposite sides of river) is 0.226 over 1 km, whereas the

genetic distance between Varadero Stream and Varadero Forest 1 (different morphs, same side of river) is 0.572 over 2.9 km. For comparison, this is roughly equivalent to the genetic distance between Varadero Stream and Micaela Bastidas ($D' = 0.580$), two striped populations separated by 33.4 km airline distance.

(3) *Varadero (site 8; Fig. 1-S1)* – This is the *R. fantastica* mimic morph, with orange dorsal colouration, orange upper arms, and navy blue reticulation on the legs and lower body. Frogs were collected from a single site 3.5 km north of the village of San Gabriel de Varadero. For consistency with previous studies (Twomey *et al.* 2013; Yeager *et al.* 2012) we refer to this morph as the varadero morph, despite the fact that striped frogs can also be found near this village.

Male and female *Ranitomeya imitator* were collected in the field and kept in captivity in Tarapoto, Peru. Sexual dimorphism in this species is subtle and mainly related to size (Brown *et al.* 2008a), so when possible frogs were sexed based on behavioural observations made while collecting (e.g., calling, territorial fighting, tadpole transport, courtship behaviour). Frogs were weighed to the nearest 0.01 g, which was also useful for sex identification as females are generally heavier than males (across all known *R. imitator* populations, females: = 0.60 g, s.d. = 0.07 g, $n = 130$; males: = 0.48 g, s.d. = 0.06 g, $n = 176$). Frogs were housed individually in glass terraria (dimensions in cm 50 x 30 x 30). Terraria had roughly two inches of washed gravel as a substrate (primarily for temperature stability throughout the day), leaf litter, and were planted with two bromeliads (pineapple tops). Water and food (wild fruit flies) were both constantly available. For each of the three populations, we targeted a sample size of 20 responses to analyse mating preferences, as specified by the animal use protocol permit (AUP #225a).

Mate choice trials were initiated by releasing two females simultaneously into the terrarium of a male. Trials were filmed for one hour. After the trial, the same females were then released into a terrarium of a male of the opposite morph of the first male, and again filmed for one hour. To account for any order effects, the morph of the male tested first was determined at random. In cases where the male or one or more females were unresponsive (typically by hiding in the gravel), trials were re-run at a later date. Terraria were illuminated with a full-spectrum ZooMed AvianSun 5.0 UVB 26 watt compact fluorescent bulb. To allow the full spectrum of light to pass into the terrarium, we constructed a special terrarium cover out of UV-transparent acrylic that we used during trials.

Population genetics

Genetic divergence between populations was assessed by genotyping 136 *R. imitator* individuals at 11 microsatellite loci. We amplified the following loci: RimiA06, RimiA07, RimiB01, RimiB02a, RimiB07a, RimiB11, RimiC05a, RvarD01, RimiD04, RimiE02a, and RimiF06 (see Table 1-S6 for primer sequences) following extraction and amplification protocols described in Brown *et al.* (2009a), with the exception that 56°C was used as the annealing temperature for B07, C05a, and E02a and 54°C for D01. Forward primers were labelled with a fluorescent tag for visualisation (6-FAM, NED, PET, or VIC). Loci were amplified individually and multiplexed for sequencing. Sequencing was done on an ABI 3130 sequencer and fragment sizes were analysed using GeneMapper software (Applied Biosystems). We used Micro-Checker software version 2.2.3 (Van Oosterhout *et al.* 2004) to check for presence of null alleles. Three out of the

original eleven loci (A07, B07, E02) showed evidence for high null allele frequencies (mean across populations > 0.09), thus these loci were omitted from further analyses.

We used the program Structure version 2.3.4 (Pritchard *et al.* 2000) to investigate population genetic structure from the microsatellite data. This program employs a Bayesian clustering algorithm to assign individuals probabilistically to each of *K* populations, where *K,* the number of populations, is unknown. The program was run with a burn-in of 50,000 generations and 500,000 subsequent generations, from one to five genetic clusters ($K = 1–5$), with five replicates at each value of *K.* The program was run using the admixture model with allele frequencies correlated. No prior information on sampling location was used in the model. To determine the number of clusters that best describe the data, we used the method described in Evanno *et al.* (2005), which is based on the second-order rate of change of the log-likelihood. This method was implemented using Structure Harvester (Earl & vonHoldt 2012).

To estimate cline shape for the microsatellite data, we used the first major axis from a factorial correspondence analysis (FCA), calculated using the software Genetix version 4.5 (Belkhir *et al.* 1996). This method is conceptually similar to principal components analysis, except it takes into account features of genetic data such as heterozygosity and homozygosity. This analysis was run using the eight microsatellite loci and the nine localities on the sampling transect for which we had genetic data. The program was run without any population information on the samples.

Dispersal estimates

As individuals disperse into the centre of a hybrid zone, they carry with them the gene combinations characteristic of their parental populations (Szymura & Barton 1986). This influx of parental genotypes into the hybrid zone creates linkage disequilibria among unlinked genetic loci, which are broken down through recombination in hybrids. Thus, strong linkage disequilibrium in a hybrid zone is evidence of reduced hybridization and increased reproductive isolation among parental types. Similarly, the influx of parental phenotypes into a hybrid zone will create covariance among independent phenotypic traits (Barton & Gale 1993). As with linkage disequilibrium, this phenotypic covariance should peak in the centre of the hybrid zone, and will be maximized when reproductive isolation is complete (Barton & Gale 1993).

We used estimates of cline width, linkage disequilibrium, and phenotypic covariance to calculate the scale of dispersal (σ) in two different ways. First, we used the relationship (from Barton & Gale 1993)

where r is the recombination rate among loci (assumed to be 0.5 for unlinked loci), w is the the hybrid zone. For cline width, we used a value of 0.97 km, which corresponds to the point estimate for cline width for the entire dataset (see Table 1-S4, model D). To calculate linkage disequilibrium, we used the software Multilocus version 1.3 (Agapow & Burt 2001) to calculate the multilocus linkage disequilibrium estimator $_d$ which has a form similar to a correlation coefficient and can have a maximum value of 1. We found that, consistent with the predictions of a hybrid zone, linkage disequilibrium among the eight microsatellite loci peaked ($_d$ = 0.066)

near the hybrid zone centre (Fig. 1-S3). Second, we used phenotypic covariance to estimate dispersal using the following relationship (Barton & Gale 1993; Gay *et al.* 2008)

recombination rate (again assumed to be 0.5 for unlinked traits). We chose two phenotypic traits showing clear sigmoidal variation across the transect: arm colour and leg colour. These two traits have a maximum slope of 0.96 and 1.68, respectively, in the centre of the hybrid zone, and have a maximum covariance of 0.15, which occurs near the hybrid zone centre (Fig. 1-S3). Using linkage disequilibrium, we estimated the dispersal rate (σ) to be 0.095 km per generation. Using phenotypic covariance, we estimated $\sigma = 0.248$ km per generation. Although we have no direct survey data in *R. imitator* with which to compare these dispersal estimates, given the available field observations, these estimates seem reasonable. While adult *R. imitator* are highly territorial and occupy small home ranges (approximately 5–14 m^2, Brown *et al.* 2009b), these home ranges are centred around reproductive resources and tightly packed in space such that most reproductive resources in a given site will be monopolized by a breeding pair. Thus, a juvenile or an individual in search of a territory may be forced to disperse a large distance in search of suitable breeding habitat. For example, two transient adult *R. imitator* moved distances of 19 and 23 m during a field study (Brown *et al.* 2009b). Additionally, one juvenile moved across an entire study plot (a distance of approximately 40 m) over the course of about one week (J. Brown, pers. comm.). Finally, during the course of a separate study (Tumulty *et al.* 2014), a male

who was removed from his territory and released 160 m away was able to return to his territory in only five days (J. Tumulty, pers. comm.).

If the current hybrid zone reflects secondary contact with neutral diffusion, the width of the cline (w) can be predicted using the number of generations since contact (T) and the dispersal rate (σ), with the following equation (Endler 1977)

Using our lower estimate of dispersal (0.095 km per generation), a cline formed by secondary contact with neutral diffusion should exceed the overall observed cline width (0.97 km; see Table 1-S4 model D) in only 17 generations. Using the upper estimate of dispersal (0.248 km per generation), a neutral cline would only take three generations to exceed the observed cline width. *Ranitomeya imitator* has a generation time of roughly eight months, which means secondary contact (assuming neutral diffusion) would have occurred not more than 11.3 years ago. Therefore, if the current hybrid zone is due to secondary contact, it is likely maintained by some barrier to gene flow (possibly assortative mating, divergent selection, or selection against hybrids), otherwise the cline is too narrow to be reasonably described by a neutral diffusion model.

Acknowledgements

We thank Jesse Delia and Jeff McKinnon for discussions and comments on the manuscript, Anders B. L. Larsen for discussion on image texture quantification, Morgan Kain for help with multiple-matrix regression, and Justin Touchon for help with the generalized linear mixed model analysis. We thank Jason Brown, Santiago Cisneros, César Lopez, Manuel Guerra-Panaifo,

Michael Mayer, Mark Pepper, Neil Rosser, Manuel Sanchez-Rodriguez, Lisa Schulte, Adam Stuckert, James Tumulty, and Justin Yeager for help in the field. For help with research permits and museum specimens, we thank Pablo Venegas. This research was funded by a NSF DDIG (1210313) grant awarded to ET and KS, a National Geographic Society grant awarded to KS (8751-10), and the NCCB scholarship (2012) at East Carolina University awarded to ET. Research permits were obtained from the Ministry of Natural Resources (DGFFS) in Lima, Peru (Authorizations No. 050-2006-INRENA-IFFS-DCB, No. 067-2007-INRENA-IFFS-DCB, No. 005-2008-INRENA-IFFS-DCB). Tissue exports were authorized under Contrato de Acceso Marco a Recursos Geneticos No. 0009-2013-MINAGRI-DGFFS/DGEFFS, with CITES permit number 003302. All research was conducted following an animal use protocol approved by the Animal Care and Use Committee of East Carolina University.

Figures and tables

Figure 1-1 – Mimetic divergence in *Ranitomeya imitator.* In central Peru, the mimic poison frog *R. imitator* **(a)** exhibits two mimetic morphs corresponding to two different model species **(b)**. These morphs occupy distinct geographic areas **(c)**, and form a narrow transition zone (grey box, **c**) characterized by phenotypic intermediates **(d)**. Scale bar in panel **(c)** equals 3 km. **(e)** Map of Peru showing study area (scale bar equals 500 km).

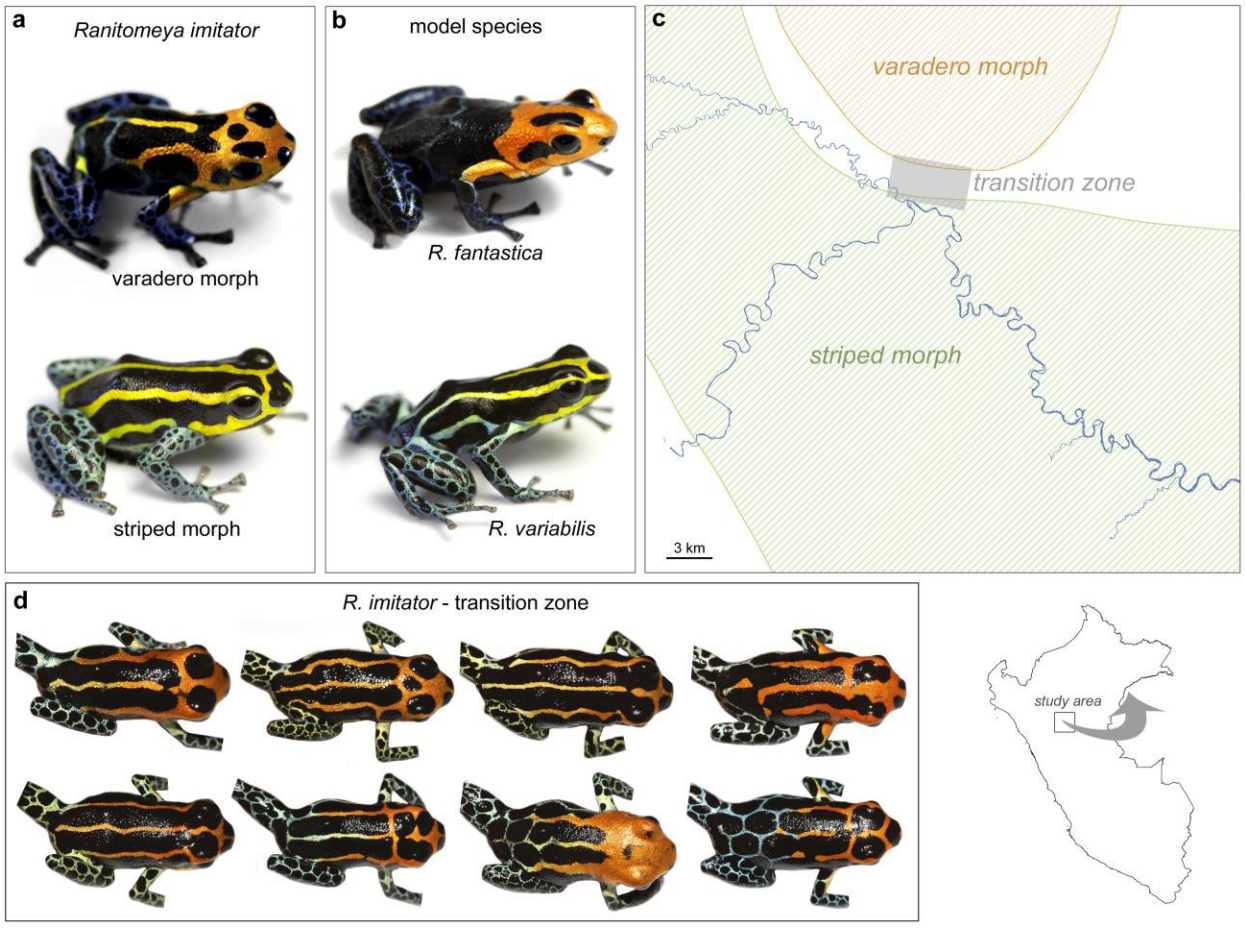


Figure 1-2 – Clines in colour pattern, microsatellites, male mass, and advertisement calls. In all panels, trait values for individual *R. imitator* (represented by dots) are plotted along the geographic transect (x-axis). (**a–f**) colour pattern variation (y-axis: kernel discriminant score; values closer to +1 indicate closer similarity to *R. variabilis* and closer to -1 *R. fantastica*); **(g)** microsatellite variation (y-axis: first major axis from FCA); **(h)** male mass (y-axis: grams); **(i)** advertisement call variation (y-axis: linear discriminant score). The fit line for each variable represents the best supported model describing transect variation and parameter point estimates.

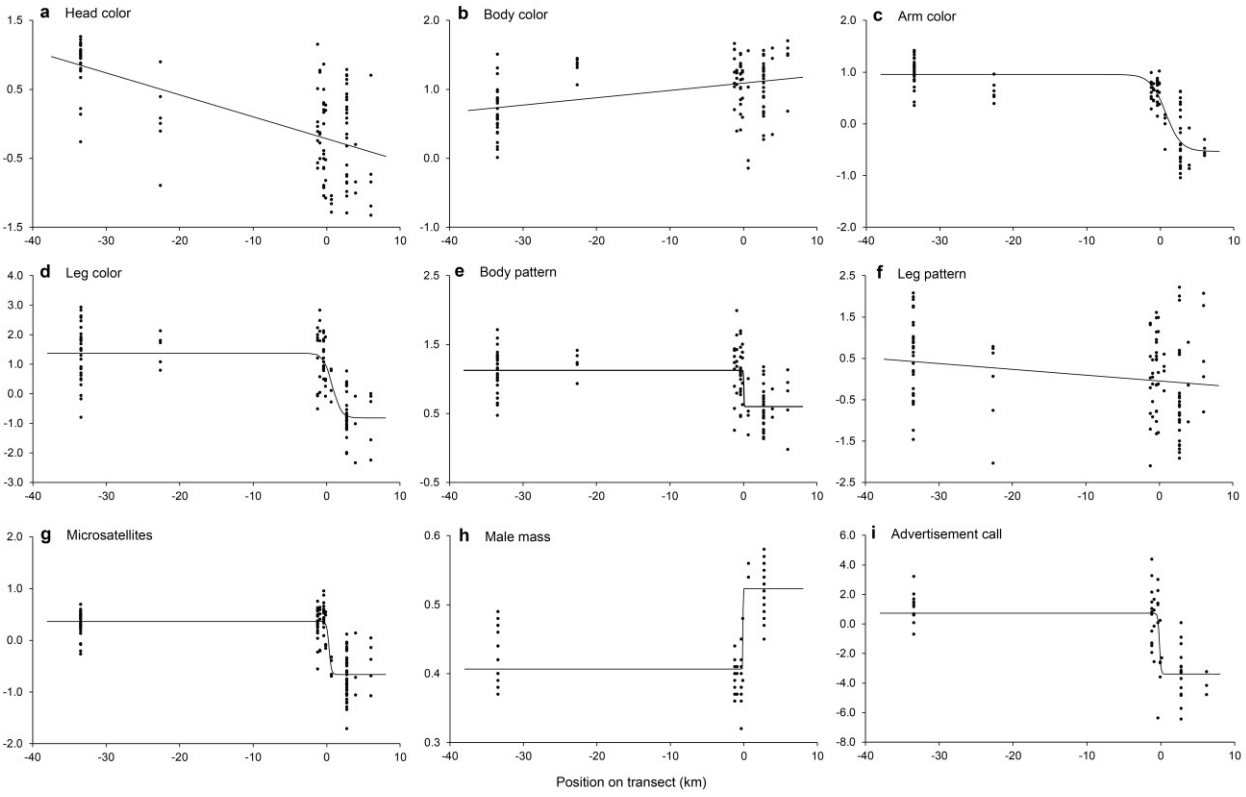


Figure 1-3 – Genetic structure between mimetic morphs of *R. imitator*. **(a)** We used the software Structure 2.3.4 to analyse the multilocus microsatellite dataset and assign individuals of *R. imitator* to each of *K* populations. The optimal number of inferred populations was *K*=3 (shown). Horizontal grey bars represent the morph (upper bar) and sampling localities (lower bar). **(b–d)** Spatial genetic structure of each of the three genetic groups as inferred by Structure. We projected the Structure output to a map by interpolating the average probability assignment score of each population to each inferred group using inverse-distance weighted interpolation in ArcGIS. **(b)** Probability assignment to group 1, **(c)** probability assignment to group 2, **(d)** probability assignment to group 3.

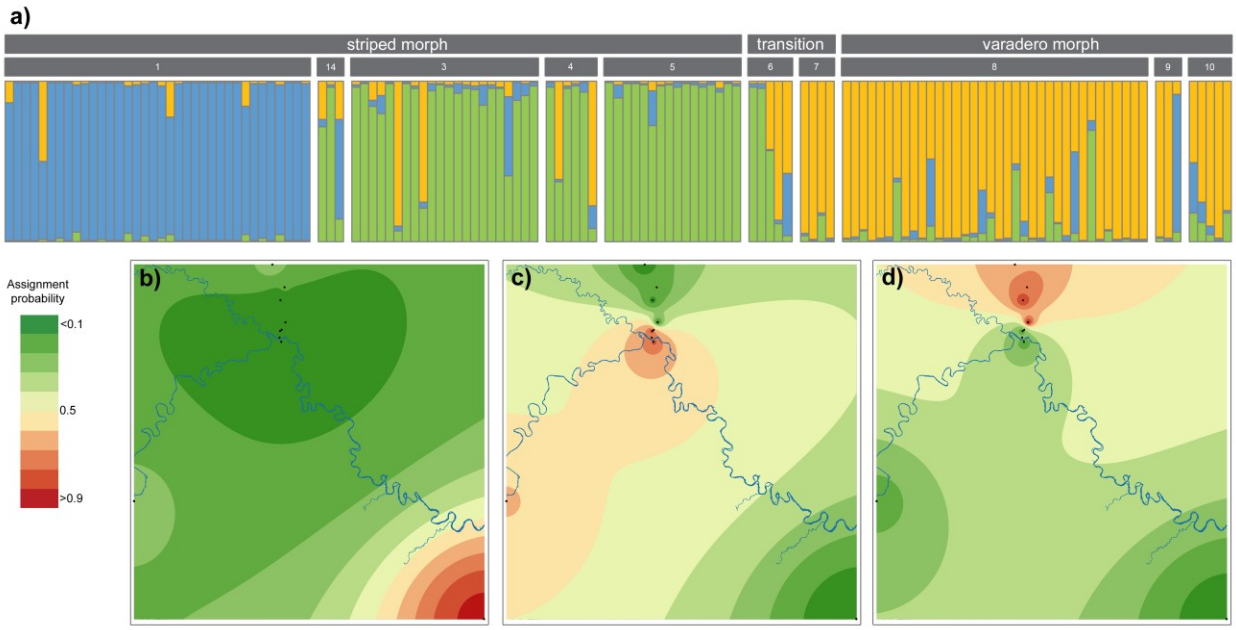


Figure 1-4 – Courtship preferences in *R. imitator*. For display on the figure, raw courtship times for each trial were converted to a "preference index", which was calculated by dividing the time a male spent courting the varadero female by the time spent courting the varadero female + time spent courting the striped female (i.e., dividing by total courtship time). This index therefore ranges from 0 (all courtship with striped female) to 1 (all courtship with varadero female), with a value of 0.5 (indicated by the dotted line) indicating no preference. Open circles show the mean preference index for each population; error bars represent 95% confidence intervals. Icons next to the error bars represent the morph of the male used in the experiment. Asterisks indicate significant differences from the GLMM ($P < 0.05$) from the post-hoc tests, following false-discovery rate adjustment for multiple comparisons. Sample sizes are as follows: striped-allopatric, N=10; striped-transition, N=19; varadero, N=26.

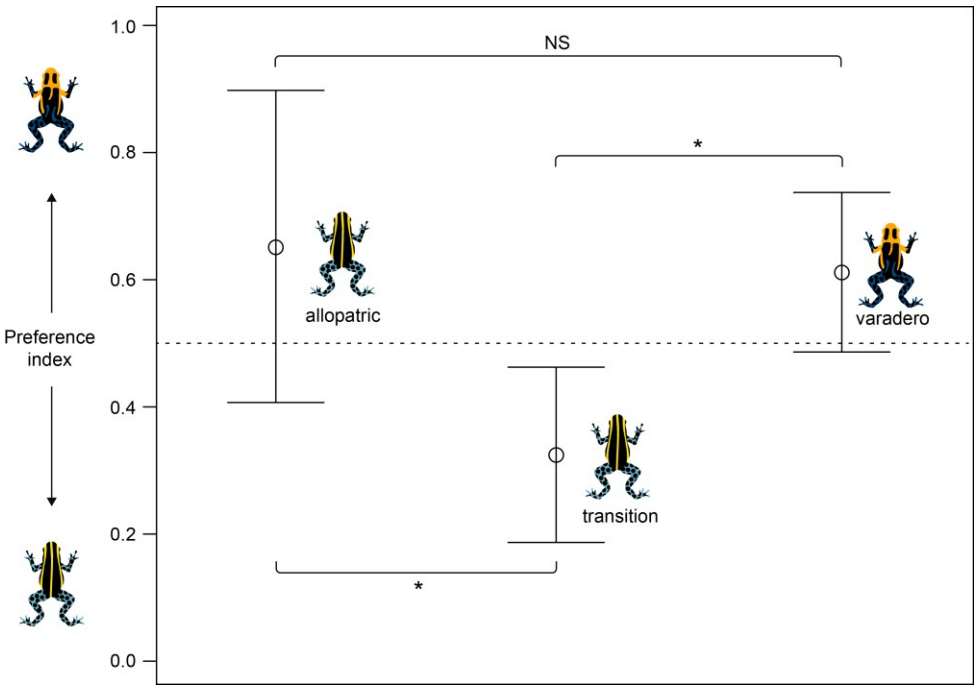


Figure 1-S1 – Sampling localities and variation in *Ranitomeya imitator*. Numbered boxes show *R. imitator* variation at a given locality (numbers correspond to localities in Table 1-S1). The colour of the dot indicates the putative morph: green dots indicate the striped morph (*R. variabilis* mimic), orange dots indicate the varadero morph (*R. fantastica* mimic) and grey dots indicate the transitional form between the two morphs. Model species are shown in middle-left box. Composite colour pattern score was calculated from the colour pattern data using a kernel discriminant function analysis, with the model species representing the training groups. Mean discriminant scores at each sampling locality were then interpolated for visualisation on the map using IDW interpolation in ArcGIS. Scale bar equals 10 km.

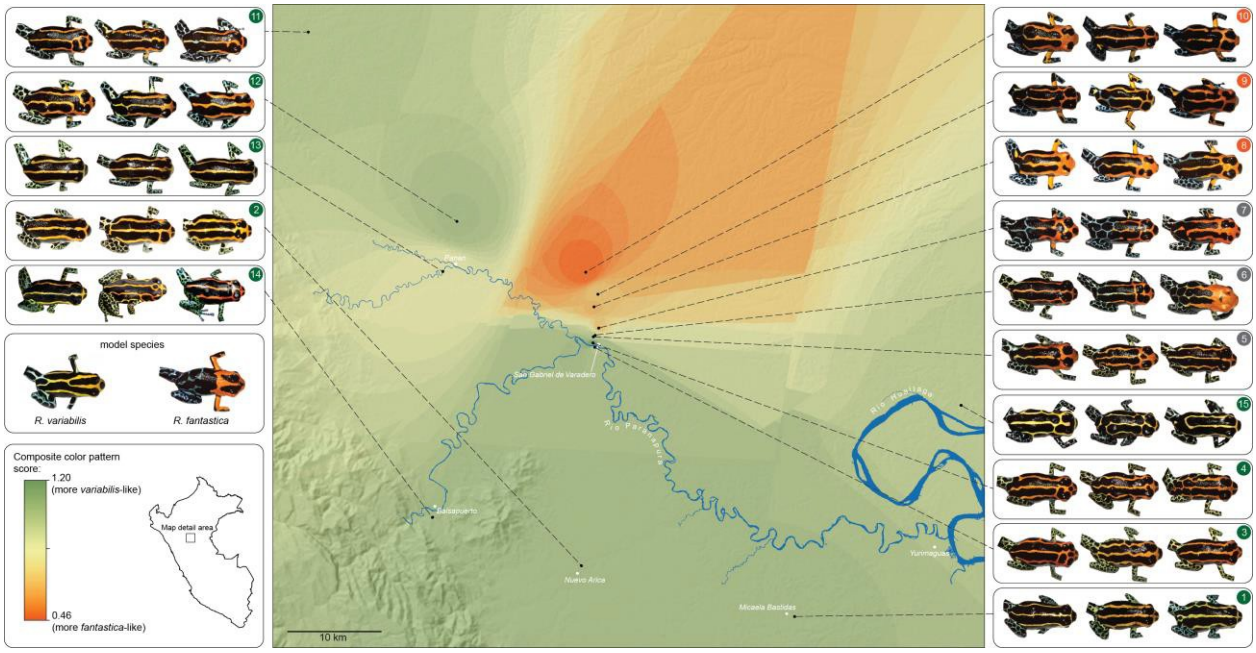


Figure 1-S2 – Mimicry comparison between model species and *R. imitator*. In each panel the average discriminant scores (±1 s.e.m.) from the kernel discriminant analysis are compared among model species (*R. variabilis*, denoted *R. v.*; *R. fantastica,* denoted *R. f.*) and two populations of *R. imitator* (*R. i. striped* and *R. i. varadero).* The two *R. imitator* populations plotted here are Micaela Bastidas and Varadero Forest 1 (see Table 1-S1), which are representative of "pure" striped and varadero morphs, respectively.

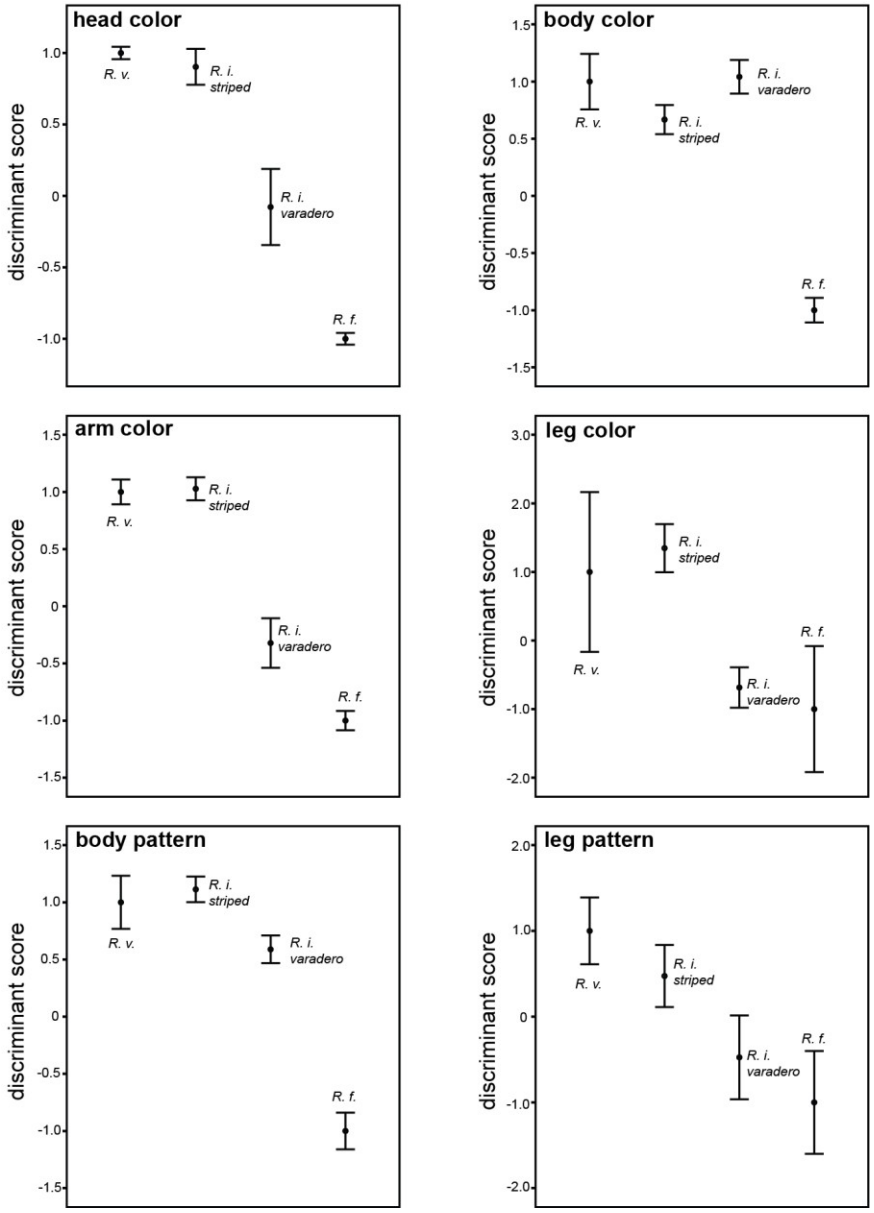


Figure 1-S3 – Linkage disequilibrium and phenotypic covariance across the sampling transect. (**a**) Multilocus linkage disequilibrium ($_d$) and (**b**) covariance between arm colour and leg colour in each population was calculated plotted along the transect (x-axis). The dashed grey lines show the position of the colour pattern cline centre. For both linkage disequilibrium and phenotypic covariance, the peaks occur near the centre of the colour pattern cline, consistent with the predictions of a hybrid zone.

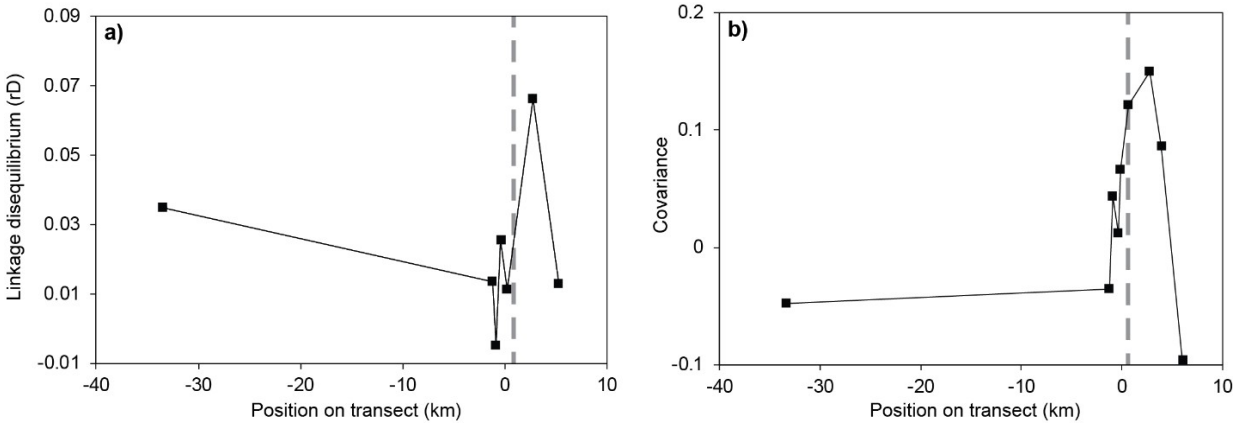


Figure 1-S4 – Mutual information plot for colour variables. See methods for details on how mutual information is calculated. In general, larger bars indicate greater variable contribution to the discriminant function. For panels (**a–c**), variables were derived from the Avicol analyses on the spectrometer data. Variable prefixes are as follows: Qt = brightness, MS = blue/yellow axis position, LM = red/green axis position, C = chrominance, H = hue. For panel (**d**), variables were derived from measuring the intensities of the red (R), green (G), and blue (B) channels in Photoshop.

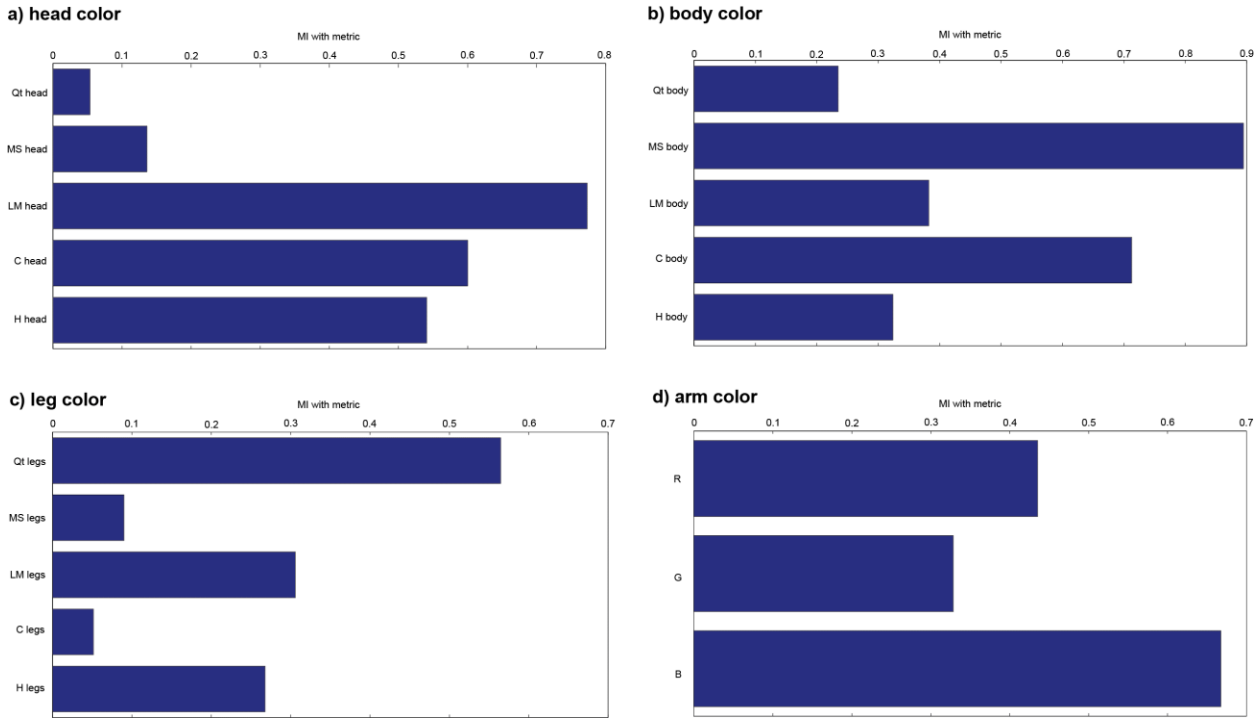


Figure 1-S5 – Mutual information plot for pattern variables. Mutual information calculations are detailed in the methods. As in Supplementary Figure 4, larger bars indicate greater variable contribution to the discriminant function. Panels are split by body region: (**a**) body pattern and (**b**) leg pattern. Variables use the following naming convention: body region, metric, and associated parameters. For example, c3goh, sigma=2, b=0 indicates that the body region of interest was c3, the metric was goh (gradient orientation histogram), and the associated parameters for its extraction were sigma=2 and b=0. Body regions are as follows: c3=lower dorsum, c4=head, c1=right leg, c2=left leg. Extracted metrics are as follows: goh=gradient orientation histogram, sih=shape index histogram, bwratio=colour/non-colour ratio.

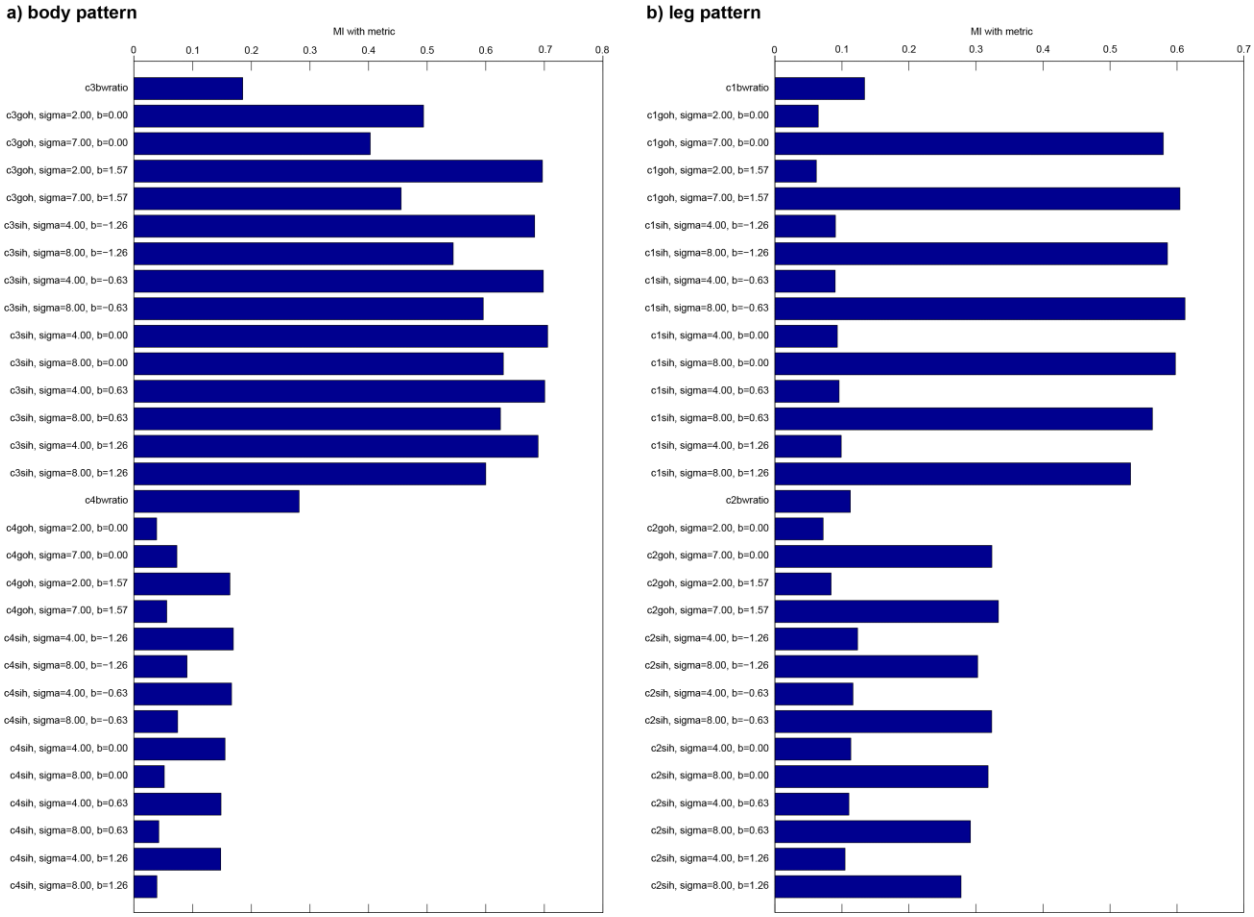


Table 1-S1 – Sampling localities and sample sizes for each kind of data collected for the study. Numbers in the first column correspond to localities shown on Fig. 1-S1.

locality	number on Suppl. Fig. 1	position on transect (km from centre)	latitude	longitude	colour pattern sample size	microsatellites sample size	male mass sample size	advertisement call sample size
Micaela Bastidas	1	-33.46	-5.9554	-76.2424	31	36	23	11
Nuevo Arica	2	-22.63	-5.9107	-76.4280	6	—	—	—
Varadero - South Bank	3	-1.25	-5.7177	-76.4163	8	22	13	13
Varadero - Bridge	4	-0.92	-5.7142	-76.4178	6	6	4	4
Varadero - Stream	5	-0.39	-5.7084	-76.4174	16	16	11	6
Varadero - Transition 1	6	-0.16	-5.7073	-76.4161	5	5	3	4
Varadero - Transition 2	7	0.64	-5.7009	-76.4128	4	4	2	—
Varadero - Forest 1	8	2.72	-5.6821	-76.4171	25	36	22	17
Varadero - Forest 2	9	3.92	-5.6710	-76.4137	3	3	—	—
Varadero - Forest 3	10	6.03	-5.6515	-76.4241	5	5	—	3
Monte Cristo	11	not included in transect	-5.4395	-76.6655	6	—	—	—
Panan North	12	not included in transect	-5.6067	-76.5361	3	—	—	—
Panan South	13	not included in transect	-5.6510	-76.5484	4	—	—	—
Balsapuerto	14	not included in transect	-5.8542	-76.5431	—	3	—	—
Bajo Huallaga	15	not included in transect	-5.7618	-76.0780	5	—	—	—

Table 1-S2 – Model fit results for each of three candidate models describing transect variation in all six variables measured along the transect. In all cases, the best supported model (as indicated by AICc) is shown in bold. For cases where the sigmoid model was best supported, point estimates and 95% Monte Carlo confidence intervals on centre and width parameters are given.

variable	model	AICc	ΔAICc	Akaike weight	centre	centre 95% CI	width	width 95% CI
arm colour	flat	-89.5	134.0	0.000	—	—	—	—
	linear	-150.2	73.3	0.000	—	—	—	—
	sigmoid	**-223.4**	**0.0**	**1.000**	**0.72**	**0.02 – 1.65**	**4.15**	**1.79 – 6.43**
body colour	flat	-181.8	15.9	0.000	—	—	—	—
	linear	**-197.7**	**0.0**	**0.969**	—	—	—	—
	sigmoid	-190.8	6.9	0.031	—	—	—	—
head colour	flat	-55.8	58.6	0.000	—	—	—	—
	linear	**-114.4**	**0.0**	**0.894**	—	—	—	—
	sigmoid	-110.1	4.3	0.106	—	—	—	—
leg colour	flat	46.9	85.2	0.000	—	—	—	—
	linear	20.5	58.8	0.000	—	—	—	—
	sigmoid	**-38.3**	**0.0**	**1.000**	**0.77**	**0.04 – 1.93**	**2.38**	**0.03 – 4.86**
body pattern	flat	-191.6	44.0	0.000	—	—	—	—
	linear	-203.3	32.4	0.000	—	—	—	—
	sigmoid	**-235.6**	**0.0**	**1.000**	**0.03**	**-0.17 – 1.82**	**0.10**	**0.00 – 4.75**
leg pattern	flat	13.0	2.9	0.180	—	—	—	—
	linear	**10.1**	**0.0**	**0.772**	—	—	—	—
	sigmoid	15.6	5.6	0.048	—	—	—	—
microsatellites FCA axis 1	flat	-138.8	145.6	0.000	—	—	—	—
	linear	-164.8	119.6	0.000	—	—	—	—
	sigmoid	**-284.4**	**0.0**	**1.000**	**0.31**	**-0.15 – 0.63**	**0.64**	**0.01 – 1.54**
male mass	flat	-423.3	82.0	0.000	—	—	—	—
	linear	-424.5	80.9	0.000	—	—	—	—
	sigmoid	**-505.3**	**0.0**	**1.000**	**-0.13**	**-0.15 – 0.63**	**0.07**	**0.00 – 1.50**
advertisement call	flat	114.8	39.2	0.000	—	—	—	—
	linear	102.5	26.9	0.000	—	—	—	—
	sigmoid	**75.6**	**0.0**	**1.000**	**-0.19**	**-0.5 – 1.91**	**0.40**	**0.00 – 5.41**

Table 1-S3 – Causal modelling results of factors potentially influencing genetic distance and associated statistical predictions under each hypothesis. Results from the partial Mantel tests are given as a p-value and a yes/no indication of whether the prediction was supported. For each statistical prediction, a × separates the two dependent matrices, with the covariate matrix separated by a period. For example, Dist × Gen . Cp tests for the correlation between Dist (geographic distance) and Gen (genetic distance) controlling for the effect of Cp (colour pattern distance).

Factor(s) influencing genetic structure	Statistical predictions	Result (p-value)	Prediction supported?
Geographic distance	Dist × Gen . Cp = sig.	0.434	no
	Cp × Gen . Dist = n.s.	0.011	no
Colour pattern	Cp × Gen . Dist = sig.	0.011	yes
	Dist × Gen . Cp = n.s.	0.111	yes
Geographic distance and colour pattern	Cp × Gen . Dist = sig.	0.010	yes
	Dist × Gen . Cp = sig.	0.115	no

Table 1-S4 – Global regression results for common centre and common width parameters for the six variables showing sigmoidal variation across the transect. Models were compared globally using AICc. The best supported model (shared centre) is shown in bold.

		a) No constraints			**b) Shared centre**			c) Shared width			d) Shared centre and width		
	n	centre	width	RSS	**centre**	**width**	**RSS**	centre	width	RSS	centre	width	RSS
arm colour	108	0.72	-4.14	29.18	**0.52**	**-3.68**	**29.24**	0.00	-0.35	32.55	0.22	-0.97	31.09
leg colour	108	0.77	-2.38	45.85	**0.52**	**-1.86**	**45.94**	0.66	-0.35	46.84	0.22	-0.97	47.26
body pattern	108	0.03	-0.10	67.10	**0.52**	**-0.10**	**67.11**	0.21	-0.35	67.12	0.22	-0.97	67.56
male mass	78	-0.12	-0.07	24.74	**0.52**	**-0.10**	**24.89**	0.07	-0.35	24.83	0.22	-0.97	26.46
call	58	-0.22	-0.31	26.20	**0.52**	**-4.15**	**27.28**	-0.22	-0.35	26.21	0.22	-0.97	28.60
microsatellites	133	0.31	-0.64	14.72	**0.52**	**-0.27**	**14.76**	0.48	-0.35	14.76	0.22	-0.97	14.94
total RSS				207.79			**209.23**			212.31			215.92
parameters				24			**19**			19			14
AICc				-571.74			**-578.44**			-569.78			-570.38
ΔAICc				6.70			**0.00**			8.66			8.06

Table 1-S5 – Parameters used in the kernel discriminant analysis.

	Regularization λ	*Kernel width* γ
head color	0.002	2.14
body color	0.001	9.00
leg color	0.310	6.30
arm color	0.000	4.78
leg pattern	0.010	10.00
body pattern	0.001	22.90

Table 1-S6 – Primer sequences for microsatellites.

locus	Forward primer sequence	Reverse primer sequence
RimiA06	CTTAATTGAGTAATTGTCAAG	GCTTTTGGATAATCAGTATCG
RimiA07	TTCTTAATTGAGTAATTGTC	TCCTTAATATACCAGTTAAGC
RimiB01	TAATTGTATTTGTCACTGAC	ATTTTTGCGGGCATATTCGG
RimiB02a	TCGAGATTTTAGCAGTGTTTTATCC	CATGAAAACCATATTTCGGACA
RimiB07a	CACCGTGCACTGGTTATCTATC	GTTTCGCTCAACCCTAGTGC
RimiB11	GTAAGTCCGTATATGTCGATG	CCTGAGAGTGTAATGGATAGAC
RimiC05a	CGTTTCGCTCAACCCTAGTC	ATGGAGGCAATCCACAAATC
RvarD01	GAAAAAGCATTACAGCTCATCAA	GCCGAAACATTGCCATAAAT
RimiD04	CTCCAAAACACACCCCAAAC	AGAGGTGCTGCCCTTTTGTA
RimiE02a	GCAGAGGGGATTAGGGACTC	TGGGTAGCTGTGTTCCATGA
RimiF06	TTGATATTCTGAGGTATG	GTAGCTTATGGCAGCTACG

Chapter 2: Mimetic divergence and the speciation continuum in the mimic poison frog *Ranitomeya imitator*

Abstract

While divergent ecological adaptation can drive speciation, understanding the factors that facilitate or constrain this process remain a major goal in speciation research. By studying cases of replicated ecological divergence, we can understand properties of both species and environments that influence progress toward speciation. We study the poison frog *Ranitomeya imitator*, a species that has undergone a Müllerian mimetic radiation to establish four morphs in Peru. Ecological divergence is replicated as there are three mimetic transition zones where a putatively ancestral morph has "shifted" to an alternate morph. We use a combination of color pattern quantification, cline analysis, landscape genetics, and mate choice experiments to study two transition zones, comparing them to a third transition zone described previously. We find that mimetic morphs are strongly differentiated in aspects of the mimetic phenotype, producing geographic clines with varying widths. However, distinct mimetic morphs show little neutral genetic divergence, and landscape genetic analyses implicate isolation-by-distance as the primary determinant of among-population genetic differentiation. Mate choice experiments suggest random mating with respect to mimetic morph at the transition zones, although certain allopatric populations do appear to show a preference for their own morph, and we present evidence that this preference may be mediated by color-pattern specifically. Overall, our results suggest that mimetic divergence has failed to generate reproductive isolation in the two transition zones studied here. These results contrast with an earlier study of the third transition zone, in

which a shift in mimicry was associated with reproductive isolation. We discuss possible explanations for this variable progress toward speciation, and suggest that multifarious selection on both mimetic color-pattern and body size may be responsible for generating reproductive isolation.

Introduction

Divergent selection among populations inhabiting different ecological conditions has been shown to generate reproductive isolation (Chamberlain *et al.* 2009; Hatfield & Schluter 1999; Jiggins *et al.* 2001; McKinnon *et al.* 2004; Nosil *et al.* 2003), a process known as ecological speciation (Nosil 2012; Schluter 1996). Despite the wide appreciation for the role of divergent ecological adaptation in speciation, our understanding of the factors that promote or inhibit incipient population divergence remains relatively limited. Studies have found that, while divergent selection can often initiate speciation, there is substantial variability among natural systems in how far speciation has progressed (Funk *et al.* 2006; Hendry *et al.* 2009; Mallet *et al.* 2007; Merrill *et al.* 2011; Nosil *et al.* 2009; Rosenblum 2006; Rosenblum & Harmon 2011; Seehausen 2009). This variability in stage of speciation is often referred to as the "speciation continuum", and a major goal in speciation research is to understand the conditions that influence progress along this continuum.

Comparative studies based on natural replicates of population divergence offer a powerful approach for understanding the conditions that facilitate or constrain ecological speciation (Berner *et al.* 2009; Jiggins *et al.* 2004; Rosenblum 2006; Rosenblum & Harmon 2011). One approach involves studying multiple species that have diverged across a common ecological

setting, revealing inherent properties of species that facilitate or constrain ecological speciation. For example, in a study of replicated ecological divergence of three lizard species across a common ecological gradient (a white sands/dark soil ecotone), (Rosenblum & Harmon 2011) found that population divergence was strongest in a habitat-specialist species with low dispersal capabilities, and weakest in a generalist species with high dispersal rates, suggesting that variation in dispersal rates among species may explain differential progress toward speciation. Alternatively, by studying a single species (or species pair) that has undergone replicated divergence across a range of ecological settings, we can understand how ecological differences can facilitate or constrain speciation (Berner *et al.* 2009; Jiggins *et al.* 2004; Seehausen 2009). A good example of this approach is the cichlid species-pair *Pundamilia pundamilia* and *P. nyererei*, in which reproductive isolation between the two species increases with increasing water clarity (Seehausen 2009; Seehausen *et al.* 2008). By identifying additional examples of replicated ecological divergence, we can better understand the properties of species and environments that are conducive to ecological speciation.

Hybrid zones (also referred to as contact zones, transition zones, or clines depending on the context) have been important for speciation research in that they often reflect a balance between the diversifying effects of local adaptation and the homogenizing effects of gene flow and migration (Barton & Hewitt 1985). As hybrid zones occur between species or populations that can still exchange genes, they can provide insight into reproductive barriers that arise relatively early in the process of speciation. Three general mechanisms are thought to produce stable clines: (1) environmental ecotones, (2) heterozygote disadvantage, and (3) positive frequency-dependence (Endler 1977; Haldane 1948; Mallet & Barton 1989; Szymura & Barton 1986). A

cline formed from an ecotone reflects local adaptation of a single species to distinct, abutting environments (e.g., the White Sands lizards mentioned above). In a cline formed by heterozygote disadvantage, selection against recombinants in the zone center acts as a hybrid "sink" and slows gene flow between parental forms (Barton & Hewitt 1985). Finally, positive frequency-dependent selection can lead to stable transition zones between distinct forms. As with clines formed by heterozygote disadvantage, clines based on positive frequency-dependence are not regulated by external environmental factors, and thus may move over time (Barton & Hewitt 1985).

Aposematic signals represent a clear example of a trait that is under frequency-dependent selection (Endler & Greenwood 1988; Greenwood *et al.* 1989). Signals under positive frequency-dependence are predicted to be uniform (Joron & Mallet 1998), bringing into question the initial source of signal divergence among populations. In certain cases, such as Müllerian mimicry, signal divergence is predicted if different populations are subject to selection to resemble different model species. Thus, Müllerian mimicry is an intriguing mechanism driving population divergence in that we frequently observe (a) signal uniformity within populations, and (b) signal divergence between populations. Divergence in mimicry appears to be responsible for speciation in the *Heliconius* butterflies in many cases (Jiggins 2008), and may also have driven speciation in certain groups of reef fish (Puebla *et al.* 2007). The dendrobatid poison frog *Ranitomeya imitator* is a Müllerian mimic (Stuckert *et al.* 2014a; b; Symula *et al.* 2001) that has undergone rapid diversification in color-pattern to establish four distinct morphs in north-central Peru (Twomey *et al.* 2013). Based on previous research (Twomey *et al.* 2013; Yeager 2009; Yeager *et al.* 2012) and ongoing field work, a relatively clear geographic picture of the *R. imitator* mimicry

system is emerging. There are four mimetic morphs of *R. imitator*, each of which resembles a distinct model species (Fig. 2-1a). The geographic distributions of each morph appear to form a mosaic (Fig. 2-1b), with narrow transition zones forming where two mimetic morphs come into contact. Thus far we have found three such transition zones (Fig. 2-1b), all of which involve the striped morph transitioning into some other morph. The mimetic radiation of *R. imitator* represents an excellent system in which to study the role of mimicry in speciation for four reasons. First, the initial source of mimetic divergence is relatively clear in that various morphs of a single species have diverged to resemble several different model species. Second, all the mimetic morphs belong to a single species, thus we can study reproductive barriers that arise in the earliest stages of population divergence. Third, there is replicated ecological divergence in *R. imitator* in that there are three known instances of mimetic divergence between the striped morph and another morph. Fourth, different mimetic morphs appear to have undergone varying levels of morphological divergence, which allows us to address whether the strength or diversity of color-pattern divergence is positively correlated to the strength of reproductive isolation.

The main purpose of this paper is to study two mimetic transition zones in *Ranitomeya imitator* in order to address whether mimetic divergence is driving reproductive isolation between different mimetic morphs. In a previous study (Twomey *et al.* 2014), we used a combination of color-pattern quantification, cline analysis, mate choice experiments, and landscape genetics to study reproductive isolation across the striped-varadero transition zone. In that study, we found that mimetic divergence appears to have led to a breakdown in gene flow between morphs, possibly facilitated by assortative mating, which is consistent with the idea that mimetic divergence is driving ecological speciation in this system. In this study, we use the same

methods to study two additional transition zones (banded-striped and spotted-striped, Fig. 2-1b) to determine to whether mimetic divergence has led to reproductive isolation more generally. More importantly, as we now have data on three independent cases of mimetic divergence within a single species, we can compare all three transition zones, which allows us to investigate the factors explaining differential progress toward speciation.

Methods

The methods of this paper largely follow the methods of (Twomey *et al.* 2014). That study focused on a single mimetic transition zone in *R. imitator*. Here, we focus on two additional transition zones. Therefore, for consistency, and to facilitate direct comparisons among all three transition zones, the description of the methods herein closely follows those in (Twomey *et al.* 2014).

Sample collection and description of transects

For the color-pattern analysis across the banded-striped transect we sampled a total of 125 *Ranitomeya imitator* from 11 localities in the departments of San Martin and Loreto, Peru (Table 2-S1). These localities form a roughly southwest-northeast linear transect along the Huallaga River (Fig. 2-S1), approximately from the town of Sauce in San Martin department, to the lowlands near the city of Yurimaguas. For the genetic analyses, we sampled 158 *R. imitator* from 11 localities. We included an additional 91 samples from 11 localities from (Twomey *et al.* 2013), and 36 samples from one locality from (Twomey *et al.* 2014), for a combined dataset of 285 individuals from 16 localities (see Table 2-S1). We calculated transect position for each

sampling locality as the straight-line distance from the estimated center of the transition zone. Localities southwest of this point were given a negative sign and localities northeast of this point were given a positive sign. The initial center point (lat/long: -6.5584° -75.9517°) used in this calculation was based on field observations where an apparent shift in color-pattern occurred. Therefore, in our cline analyses, center estimates close to zero indicate that the inferred cline center was consistent with our field observations.

For the color-pattern analysis of the spotted-striped transect we sampled a total of 140 *R. imitator* from 8 localities in San Martin and Loreto provinces in Peru (Table 2-S2). With the exception of the Pinto Recodo locality, these localities form a roughly linear transect running south-to-north, extending approximately from the city of Tarapoto to the lowlands near the city of Yurimaguas. For the genetic analysis of this transect, we sampled 82 *R. imitator* from 7 localities. In addition, we included 33 samples from 3 localities from (Twomey *et al.* 2013) and 36 samples from 1 locality from (Twomey *et al.* 2014), for a combined dataset of 151 individuals from 8 localities (see Table 2-S2). As with the previous transect, we calculated transect position for sampling localities as a straight-line distance from an estimated transition zone center (for the spotted-striped transect, this center point was -6.3629° -76.2939°).

We also took color-pattern measurements from the model species and populations: 7 lowland *R. variabilis* from Pongo de Cainarachi (the model for the striped morph of *R. imitator*), 7 highland *R. variabilis* from San Jose (the model for the spotted morph of *R. imitator*), and 6 *R. summersi* from Sauce (the model for the banded morph of *R. imitator*).

Color and pattern data

Frog color (hue and brightness) was quantified by measuring the spectral reflectance at specific points on the dorsum of each frog. One measurement was taken on each side (i.e., right and left) of the head, each side of the mid-body, each side of the rear-body, and dorsal surfaces of right and left thighs, for a total of eight spectral reflectance measurements per frog. Spectral reflectance measurements were taken with an Ocean Optics USB4000 spectrometer with a LS-1 tungsten-halogen light source and Ocean Optics SPECTRASUITE software. We used a black plastic tip on the end of the probe to ensure that measurements were taken at a consistent distance (3 mm) and angle (45°) relative to the skin. To account for lamp drift we measured a white reflectance standard (Ocean Optics WS-1-SL) after every other frog. Raw spectra were processed in AVICOL version 6 software (Gomez 2006) using Endler's segment model (Endler 1990) calculated between 450–700 nm. This model calculates brightness (Qt), chroma (C), hue (H), brightness in the yellow-blue range (MS), and brightness in the red-green range (LM). After processing in Avicol, readings were averaged within body regions (head, body, and legs).

To quantify pattern, we used dorsal photographs and an automated feature extraction method to extract a suite of pattern descriptor variables. All photos were taken on a white background with a Canon Rebel XS DSLR with a Canon EF 100 mm macro lens and the camera flash. Image descriptors were automatically extracted from images of every individual and collected in a feature matrix. We extracted three types of descriptors: color/non-color ratio, gradient orientation histograms, and shape index histograms (Dalal & Triggs 2005; Koenderink & van Doorn 1992; Larsen *et al.* 2014). These descriptors capture zeroth, first, and second order image structure. We used a spatial pooling scheme to collect information separately at four interest points: left leg,

right leg, posterior dorsum, and anterior dorsum. At each of these interest points, pattern variation occurs on a distinct scale, so the descriptors were extracted according to a scale space formulation (Lindeberg 1996). Color/non-color ratios were extracted for every interest point on a single scale; gradient orientation histograms for every interest point on two different scales and two orientation bins (horizontal and vertical), and shape index histograms were extracted for the legs on two scales in five bins equidistantly spaced between $-\pi/2$ and $\pi/2$. This sums to a total of $4 \cdot (1 + 2 \cdot 2 + 2 \cdot 5) = 60$ features per individual.

To reduce the multivariate color and pattern data to a single descriptive metric for each body region, we used kernel discriminant analysis (Mika *et al.* 1999) with individuals of the model species representing the training groups used for classification. In both sampling transects, the phenotypic transition in *R. imitator* involves a shift from the striped morph (which mimics lowland *R. variabilis*) to either the spotted morph (which mimics highland *R. variabilis*), or the banded morph (which mimics *R. summersi*). Therefore, for the spotted-striped transect, the populations of the model species used for classification were the highland *R. variabilis* and lowland *R. variabilis*; for the banded-striped transect, the model species used were *R. summersi* and lowland *R. variabilis.* This procedure assigns a discriminant score to each *R. imitator* individual on the basis of its similarity to either model species, and thus can be thought of as a mimicry score. The analysis can be constrained to include only subsets of the variables to derive separate color and pattern metrics for different parts of the body. We derived color and pattern metrics for two body regions: dorsum and legs. We regularized the solution with λ times the identity matrix and used a Gaussian kernel with width γ (Table 2-S3). These parameters were selected to minimize the intra-location variance of the *R. imitator* discriminant scores while

keeping them within the range of the model species' discriminant scores. The formulation of kernel discriminant analysis makes it impossible to inspect the loadings as one would do in linear discriminant analysis to determine variable importance. To determine which variables were relevant for the discriminant analysis, we calculated the mutual information between each of the original variables $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ and the discriminant scores (metric) $\mathbf{z}$. Mutual information is defined as

$$MI(x,y) = \frac{H(x) + H(y) - H(x,y)}{H(x) + H(y)}$$

where *H(x)* and *H(x,y)* are marginal and joint entropies respectively. Parzen window estimates of MI between the variables were used to define each of the six metrics, and the resulting metric can be seen in Figures 2-S2 and 2-S3. The values MI($\mathbf{x}_i$, $\mathbf{z}$) are normalized such that MI($\mathbf{x}_i$, $\mathbf{x}_i$)= MI($\mathbf{z}$, $\mathbf{z}$)=1.

Cline fitting

To describe clinal variation in color-pattern elements (and microsatellites, see below), we used nonlinear regression using a 4-parameter sigmoid tanh function

$$y = \frac{1 + \tanh\left(\frac{2x - c}{w}\right)}{2\left(\frac{1}{y_{\max} - y_{\min}}\right)} + y_{\min}$$

where *c* is the center of the cline, *w* is the cline width, and *ymax* and *ymin* are the maximum and minimum trait values (i.e., the trait values at the tails of the cline). This uses the cline model of (Szymura & Barton 1986) except that the minimum and maximum trait values are free to take on any value. Parameter searches were done using the solver function in EXCEL using a least-

squares optimality criterion. Solver was run using the GRG nonlinear algorithm with the following settings: convergence = 0.0001; central derivatives; multistart on; population size = 100.

To evaluate whether the data were adequately described by a "flat" model (in which the trait value is constant across the transect), or a linear model (with a smooth transition across the transect), we fit these models, in addition to the sigmoid model, as candidates. A flat model consists of a single parameter (population mean) defined as the grand mean of all individuals and is invariant across the transect. A linear model has two parameters, slope and y-intercept, and was fit with linear regression. To evaluate which of the three models (1-parameter flat, 2-parameter linear, or 4-parameter sigmoid) was a better fit to the data, we calculated ΔAIC_c and Akaike weights (w_i) for each model (Burnham & Anderson 2002) using the residual sum of squares divided by the sample size as the likelihood criterion.

Confidence intervals on parameter estimates were calculated using a Monte Carlo resampling method implemented in the software GRAPHPAD PRISM (methods following (Twomey *et al.* 2014). We tested for a common center (coincidence) and width (concordance) among clines using global nonlinear regression, which compares model fit when certain parameters are shared or not different variables. For example, under a scenario where all variables shift at the same point on the transect, a common center parameter can be fit across all measured variables without a substantial reduction in model fit. For each transect, we evaluated seven models representing different combinations of shared and unshared parameters (Tables 2-S4 and 2-S5). Best-fit shared parameter searches were done by fitting shared parameters to all datasets simultaneously,

while unshared parameters were free to take on unique values for each dataset. Goodness-of-fit was assessed by calculating ΔAIC_c for each model.

Landscape genetics

Tissue samples for genetic analysis (toe-clips) were taken with sterile surgical scissors and preserved in 96% ethanol prior to extraction. We amplified the following loci: RimiA06, RimiA07, RimiB01, RimiB02a, RimiB07, RimiB11, RimiC05a, RvarD01, RimiD04, RimiE02a, and RimiF06 following extraction and amplification protocols described in (Brown *et al.* 2009a), with the exception that 56°C was used as the annealing temperature for B07, C05a, and E02a and 54°C for D01. Forward primers were labelled with a fluorescent tag for visualization (6-FAM, NED, PET, or VIC). Loci were amplified individually and multiplexed for sequencing. Sequencing was done on an ABI 3130 sequencer and fragment sizes were analyzed using GENEMAPPER software (Applied Biosystems). We used MICRO-CHECKER software version 2.2.3 (Van Oosterhout *et al.* 2004) to check for the presence of null alleles. One locus (RimiE02a) showed evidence for high null allele frequencies (mean across populations > 0.10), and this locus was omitted from further analyses. Another locus (RimiB01) did not amplify for most of the banded populations, and was omitted from the banded-striped transect analysis.

We used the program STRUCTURE version 2.3.4 (Pritchard *et al.* 2000) to investigate population genetic structure along each transect from the microsatellite data. This program employs a Bayesian clustering algorithm to assign individuals probabilistically to each of *K* populations, where *K,* the number of populations, is unknown. The program was run with a burn-in of 50,000 generations and 500,000 subsequent generations, from one to five genetic clusters

(K = 1–5), with five replicates at each value of K. The program was run using the admixture model with correlated allele frequencies. No prior information on sampling location was used in the model. To determine the number of clusters that best describe the data, we used the method described in (Evanno *et al.* 2005), which is based on the second-order rate of change of the log-likelihood. This method was implemented using STRUCTURE HARVESTER (Earl & vonHoldt 2012).

We conducted landscape genetic analyses to quantify the relative effects of geographic distance and color-pattern distance on genetic distance between populations. Under an isolation-by-distance scenario (IBD), geographic distance alone will account for much of the observed genetic distance between populations. If mimetic divergence is also an isolating barrier (i.e., isolation-by-adaptation, or "IBA", which would be expected under morph-based reproductive isolation), population-wise mimetic divergence should also account for a significant proportion of the observed genetic distance between populations. We used a multiple matrix regression method (Wang 2013) in order to quantify the relative effects of two distance matrices (geographic distance and color-pattern distance) on the genetic distance matrix. This method incorporates multiple regression, such that the relative effects of two or more predictor variables on genetic distance can be quantified, as can the overall fit of the model. Results consist of regression coefficients (β) and associated P-values. Regression coefficients herein are denoted β_D (for geographic distance) and β_{CP} (for color-pattern distance), and are thus measures of the relative importance of IBD and IBA, respectively. We ran the multiple matrix regression analyses with 10,000 permutations using the R script provided in (Wang 2013). Our distance matrices consisted of one measure of geographic distance (Euclidean distance), one measure of

genetic distance (Nei's *D*), and one measure of color-pattern distance (difference in discriminant score, see below). For the geographic distance matrix, we calculated pairwise straight-line distance between populations. For the genetic distance matrix, we calculated Nei's genetic distance (*D′*) between all pairs of populations in GenAlEx version 6.5 (Peakall & Smouse 2006). To generate the color-pattern distance matrix, we calculated pairwise differences in discriminant score from the kernel discriminant function analysis. Because this analysis takes into account features of the model species, it can be thought of as a composite difference in mimetic color-pattern. We conducted landscape genetic analyses separately for each transect.

To estimate cline shape for the microsatellite data on each transect, we used the first major axis from a factorial correspondence analysis (FCA), calculated using the software Genetix version 4.5 (Belkhir *et al.* 1996). This method is conceptually similar to principal components analysis, except it takes into account features of genetic data such as heterozygosity and homozygosity. The program was run without any prior population information on the samples.

Mate choice: Animal collection and husbandry

We tested mate preferences across each of the sampling transects to determine whether any of the morphs showed assortative mating preferences, and, furthermore, to determine whether distance from the transition zone influenced preference strength. For the banded-striped transect, we conducted mate choice experiments on the following four populations:

(1) Banded-allopatric (site 1, Fig. 2-S1) – Frogs from this population mimic *R. summersi*, with orange bands crossing the body and running down the legs and arms. These frogs

were collected from the village of Sauce. This population appears to be somewhat genetically distinct from other banded populations (Fig. 2-2).

(2) Banded-transition (site 3, Fig. 2-S1) – These frogs also mimic *R. summersi*, but are closer to the transition zone center, and fall in the same genetic cluster as the banded, intermediate, and striped populations that span the transition zone (Fig. 2-2). These frogs were collected near the village of Santa Rosa de Chipaota (herein shortened to "Chipaota").

(3) Striped-transition (site 13, Fig. 2-S1) – This population mimics the lowland morph of *R. variabilis*, with yellow stripes running along the dorsum and bluish-green reticulation on the limbs. Genetically, they are closely related to other populations occurring along the region of the Huallaga River spanning the transition zone (including the banded-transition population). These frogs were collected near the village of Achinamisa.

(4) Striped-allopatric (site 16, Fig. 2-S1) – Also a mimic of lowland *R. variabilis*, this population occurs in the lowlands, distant from any known transition zone. Phenotypically, this population is nearly identical to the striped morph from the transition zone. These frogs were collected near the village of Micaela Bastidas.

For the spotted-striped transect, we conducted mate choice experiments on three populations. Due to the difficulty of collecting the spotted morph (mainly because it occurs in steep, mountainous areas), we were only able to assess mating preferences in a single spotted population.

(1) Spotted (site C, Fig. 2-S1) – These frogs are the typical highland *R. variabilis* mimic, known from the Cainarachi valley north of Tarapoto. This population has green dorsal reticulation and black dorsal spots.

(2) Striped-transition (site G, Fig. 2-S1) – These frogs are fixed for the striped phenotype, corresponding to the lowland morph of *R. variabilis*, and occur within a few kilometers of the spotted-striped transition zone.

(3) Striped-allopatric (site H, Fig. 2-S1) – This population is also fixed for the striped phenotype, and occurs in the lowlands. These frogs were collected near the village of Micaela Bastidas.

Male and female *R. imitator* were collected in the field and kept in captivity in Tarapoto, Peru. This species is only weakly sexually dimorphic, with females slightly larger, on average, than males (Brown *et al.* 2008a), so frogs were sexed on the basis of behavioral observations made while collecting if possible (e.g., calling, territorial fighting, tadpole transport, courtship behavior). Frogs were weighed to the nearest 0.01 g, which was useful for sex determination as females are on average 25% heavier than males (Twomey *et al.* 2014). Frogs were housed individually in glass terraria (dimensions in cm 50 x 30 x 30). Terraria had roughly two inches of washed gravel as a substrate (primarily for temperature stability throughout the day), leaf litter, and were planted with two bromeliads (pineapple tops). Water and food (wild fruit flies) were both constantly available.

Mate choice: Free-release trial protocols and statistical analysis

To test for morph-based mating preferences, we conducted triad mate choice experiments in which we introduced two females (one of each morph) into the terrarium of a given male, and

measured the amount of courtship time between the male and the same-morph female vs. alternate-morph female. In *R. imitator*, courtship starts when a calling male approaches a female. The female can reciprocate interest by following the male to an oviposition site while the male continues calling, or otherwise show no interest (Brown *et al.* 2008a). Our mate choice experiments facilitated these behaviors in that a male was free to initiate courtship with either female, and the female was free to reciprocate interest or reject the male. Initiation of courtship is readily observable in captivity as males produce a rapid courtship call, and/or begin to walk with a staccato-like rhythm in the vicinity of the female. Therefore, when a male displayed either of these behaviors near a female, this marked the initiation of courtship. Courtship was judged to have ended under either of the following conditions: (a) the male moves away and the female does not respond, or (b) the female moves away and the male does not respond. Using these criteria, we measured in each trial the total amount of courtship time between the male and each female. These trials were initiated by placing two females into the terrarium of a male at the same time. Trials were filmed for one hour. After the trial, the same females were then released into a terrarium of a male of the other morph, and again filmed for one hour (see below for justification of paired-samples design). To account for any order effects, the morph of the male tested first was randomly selected. In cases where the male or one or more females were unresponsive (for example, by hiding in the gravel), trials were repeated at a later date. Terraria were illuminated with a full-spectrum ZooMed AvianSun 5.0 UVB 26 watt compact fluorescent bulb. To allow the full spectrum of light into the terrarium, we used a custom-built terrarium cover made of UV-transparent acrylic.

Although it would be desirable to match the two females for mass, this was not feasible due to substantial mass differences among populations. For example, the striped-allopatric population has an average female mass of 0.56 g (Twomey *et al.* 2014), whereas the banded-allopatric population has an average female mass of 0.65 g (n=14). To control for differences between females, we used a paired-samples experimental design whereby a given pair of females was presented to a male of each morph. This design addresses the question of how changing male morph type alters courtship probabilities when female identity is held constant.

To analyze the free-release mate choice data, we used generalized linear mixed models (GLMM) using the glmmADMB package (Skaug *et al.* 2011) in R version 3.0.2 (R Development Core Team 2005) with an underlying beta-binomial error distribution to test whether the time males spent courting each female morph was influenced by male population origin. We used "male origin" as a fixed effect and "pair ID" (i.e., a unique identifier assigned to each female pair) as a random effect to account for the paired-samples experimental design. The significance of male origin was determined using χ^2 tests comparing models where male origin was either included or excluded as a parameter.

For the banded-striped transition zone, we essentially ran two distinct experiments: one where frogs originated from allopatric populations, and one where frogs originated from transition zone (parapatric) populations. In the former experiment, frogs originated from Sauce (allopatric banded) or Micaela Bastidas (allopatric striped); in the latter experiment, frogs originated from Chipaota (transition zone banded) or Achinamisa (transition zone striped). Because of this design, it is not appropriate to make direct comparisons between two populations of the same morph because male origin is confounded with female origin (i.e., it is not possible

to assign variation in preference strength to male origin, as females came from different sites in the different experiments). Therefore, our design allowed for two comparisons: allopatric banded vs. allopatric striped, and transition zone banded vs. transition zone striped. Under a scenario where mate preferences have diverged in allopatry, we expect allopatric populations to show stronger assortative mating than transition populations. Alternatively, if mate preferences are enhanced at the transition zone (for example, due to reinforcement), we expect that assortative mating should be strongest between the two transition populations. Here, we ran two separate GLMMs, one for each experiment.

For the spotted-striped transition zone, because we used a single spotted population for mate choice trials, our experiment addresses whether preferences in the allopatric striped population (Micaela Bastidas) and transition zone striped population (Pongo) are different from each other. This design also allows us to address preferences in the spotted population under two conditions: when the striped female was from a transition zone population, and when the striped female was from an allopatric population. For this experiment, we ran a single overall GLMM (i.e., across all three populations).

Mate choice: Plastic model protocols and statistical analysis

Based on the results from the free-release mate choice experiments (see Results section for more details) we found that none of the populations included in the spotted-striped transect showed any significant mating preferences. However, in the banded-striped transect, we found that at least one population (Sauce) showed a preference for its own morph. As described above, the free-release mate choice trials assess total behavioral reproductive isolation (within the confines

of the experimental arena) between two mimetic morphs, using all potential cues (e.g., color-pattern, bioacoustics, behavioral signals, olfactory cues). To assess whether color-pattern alone mediates assortative mating, we ran mate choice experiments using plastic frog models painted with banded or striped color-patterns. For this experiment, we tested the same four populations as in the free-release banded-striped experiment, but we presented males with artificial model frogs rather than live females. The plastic models were printed on a 3D printer using a digital frog model (courtesy of Laura Crothers). The models, when printed, had a snout-vent length of 19 mm (corresponding to the size of a female *R. imitator*) and had the general morphology of a small dendrobatid frog (these models were originally designed to resemble *Oophaga pumilio*). We first painted the models with a black matte spray paint as a primer base, then used custom-mixed acrylic paints to paint either striped or banded color-patterns on the models. We checked the colors by measuring the spectral reflectance of the artificial frogs to ensure the paint colors accurately matched the colors of a real frog. We analyzed the data in AVICOL (as described above), and confirmed that the results from the frog models fell within the natural variation shown by the particular morph. We then glued each model to the second-hand of a clock, removed the minute and hour hands, and glued a compact disc (spray painted black) to the face of the clock (herein referred to as a platform). In this way, each model moved in a circle on its own platform, on a black disc, and with the same movement speed (i.e., one "hop" every second).

Trials were initiated by placing a striped model and a banded model in the front corners of the terrarium of a male, with side (right/left) determined at random. Housing and lighting conditions were identical to those described above. We then filmed for an hour, and scored two

types of behaviors. First, we measured the total amount of time the male spent on each platform. Second, we counted the total number of calls a male directed towards each model. For this, we counted calls that were given either from the platform itself, or within an interaction zone defined as < 10 cm from the platform. For calls, we used an interaction zone threshold because, in our initial trials, many males called from the terrarium glass just above the model platform. We also attempted to test females, but they usually tried to fight with the plastic models. This may have been because during female trials we used a small speaker in the tank to play a male advertisement call. Females seemed to recognize that the call was not emanating from the speaker; therefore, they may have perceived that the models were females trying to court a nearby (unseen) male. Therefore, this may represent female-female aggression in the context of competition for mates, something that was also observed in free-release trials. Given the high levels of male parental investment in this species, this type of aggression is expected. Because females displayed aggressive behaviors rather than courtship behaviors, we did not analyze female responses.

The two response variables (time on platform and calls) were analyzed separately in R using GLMMs with a beta-binomial error distribution (glmmADMB package). For this analysis, we included two fixed effects: male morph (banded/striped), and transect position (allopatric/transition zone), with individual male identity included as a random effect. The significance of the fixed effects was tested by removing the fixed effect of interest and comparing model fit to the full model with a χ^2 test. In addition, we tested for an interaction between male morph and transect position by comparing the full model to a model including an interaction term. An interaction is expected if transect position influences the mate preferences of

each morph differently. For example, if striped frogs have a stronger "preference for striped" near the transition zone, and banded frogs have a stronger "preference for banded" at the transition zone, a model including the interaction between morph and transect would provide a better fit than one where transect affected each morph equally.

Results

Color-pattern and microsatellite cline analysis: banded-striped transect

For the banded-striped transect, we found evidence of a sigmoidal cline in all four color-pattern metrics (i.e., dorsal color, dorsal pattern, leg color, and leg pattern; see Table 2-S6). Moving along the transect from the banded to the striped morph, we find a rapid shift in dorsal color (orange to yellow), dorsal pattern (stripes running across body to stripes running along body), leg color (orange to blue), and leg pattern (stripes running along the leg to a reticulated pattern) (Fig. 2-3). These shifts occur over a narrow geographic zone near Chipesa, Callanayacu, and Ricardo Palma. With the exception of leg color, all color-pattern metrics had a center parameter estimate close to zero (Table 2-S6), indicating that the center of the transition zone occurred close to our initial estimate of the transition zone center. The leg color cline, however, appears to be shifted 1.5–3 km into the striped morph side of the transition zone, and the 95% confidence interval for the center parameter does not include zero (Table 2-S6). There is variation in the widths of the color-pattern clines, ranging from narrow (1.16 km for dorsal color) to relatively wide (8.97 km for leg pattern).

We can inspect the mutual information plots (Fig. 2-S2) to determine the contribution of the original color-pattern variables to the respective discriminant metrics. This reveals which

specific aspects of the color-pattern shift along the transect. For dorsal color, the highest mutual information (MI) with the discriminant function metric is body hue (H body), followed by brightness in the red-green range of the head (LM head), head chroma (i.e., color purity) (C head), and head hue (H head). For leg color, all variables except leg brightness (Qt legs) showed high MI with the leg color metric. Based on these results, aspects of color (hue, in particular) contribute substantially to both the dorsal color and leg color metrics, whereas brightness contributes relatively little. For dorsal pattern, all variables except color/noncolor ratios (c3bwratio and c4bwratio) and gradient orientation histograms (goh) where the extraction parameter *b* was equal to zero showed high MI with the dorsal pattern discriminant metric. For leg pattern, several metrics showed moderately high MI, with many of the gradient orientation histograms showing higher MI, on average, than the shape index histograms (sih).

The factorial correspondence analysis on the microsatellite data generated four main axes which together accounted for 54.1% of the genetic variation. Most of this was captured by axes 1 and 2, which accounted for 20.6% and 16.8% of the genetic variation, respectively. For the cline analyses, we only analyzed the first major axis from the FCA. We did however visually inspect FCA axes 2–4 and found that they showed no clear trend across the transect. Our cline analysis of FCA axis 1 showed that the data were best explained by a sigmoidal model of variation along the transect (Table 2-S6, Fig. 2-3c). However, the point estimate for cline width was so wide (106.97 km) that the resulting model was effectively linear (Fig. 2-3e). Furthermore, the center estimate for the microsatellite cline was highly discordant (-47.68 km) relative to center estimates from the color-pattern metrics.

Based on our results from global regression, the best supported model was one where all four color-pattern metrics and microsatellite clines had unique center and width parameters. In other words, by constraining parameters to be equal across datasets, there was always a substantial reduction in model fit. However, upon comparing AICc values, we see that all models where the microsatellite cline was constrained to be coincident or concordant with the color-pattern clines resulted in substantially worse model fit than when the microsatellite cline was unconstrained (Table 2-S4). For example, in Table 2-S4, models b–d all involve constraining the microsatellite cline to share a parameter with the color-pattern clines, whereas in models e–g the microsatellite cline is free to take on its own independent parameter values. We see that by relaxing the constraints on the microsatellite cline (models e–g), we get a much better model fit (average AICc = -1406.12) than when microsatellites and color-pattern metrics are constrained simultaneously (models b–d, average AICc = -1337.30), corresponding to a ΔAICc of 68.82. Based on this, and the point estimates for cline parameters, we can conclude that the microsatellite cline has a very different shape and position from the color-pattern clines, and is effectively linear across the transect. Finally, we see that it is possible to constrain color-pattern clines to have equal width with only a modest reduction in fit (ΔAICc = 5.86; Table 2-S4 model f). Therefore, while the best supported model may be one where color-pattern metrics each have a unique width, constraining the widths to be equal still gives a reasonable fit.

Color-pattern and microsatellite cline analysis: spotted-striped transect

For the spotted-striped transect, only dorsal color and dorsal pattern showed sigmoidal variation along the sampling transect, whereas leg color and leg pattern were best described by a linear

model (Fig. 2-3, Table 2-S7). These results fit with our field observations, as there is a clear shift in dorsal color (green to yellow) and dorsal pattern (spotted to striped) between the two morphs, whereas the color and pattern of the legs show no clear transition. The center estimate for dorsal color was close to zero (-0.17 km), however, the center of the dorsal pattern cline appears to be shifted north towards the striped morph (center = 7.96 km). In addition, the width estimates between the dorsal color and dorsal pattern are distinct, with dorsal color showing a much narrower cline (2.94 km) than dorsal pattern (14.31 km). Therefore, moving along the transect towards the striped morph, our data suggest that dorsal color shifts earlier and more abruptly than dorsal pattern. This discordance between color and pattern clines is exemplified at the sampling locality "Upper Pongo 1" (km 1.44 on Fig. 2-3). We see that Upper Pongo 1 has a dorsal color value typical of the striped morph (i.e., yellow), however, the dorsal pattern value is more typical of the spotted morph.

Based on the mutual information plots (Fig. 2-S3) for the two variables showing sigmoidal variation, we see that variation in the red-green spectrum of the head (LM head) and body (LM body), and head hue (H head) all show high MI with the dorsal color metric. For dorsal pattern, most of the variables showing high MI were those extracted from the lower dorsum (prefix c3) whereas those extracted from the head (prefix c4) showed relatively low MI. Taken together, we can conclude that the clinal shift in color-pattern is primarily due to hue variation on the dorsum and patterning variation on the lower dorsum.

The factorial correspondence analysis on the microsatellite data generated four main axes which together described 77.5% of the genetic variation. Most of this variation was captured by axes 1 and 2, which accounted for 30.3% and 21.3% of the genetic variation, respectively. With

respect to axis 1, two clusters are evident: Micaela Bastidas (allopatric striped; Fig. 2-S1, population H) and all other populations. With respect to axis 2, the Pinto Recodo population forms its own cluster (allopatric spotted; Fig. 2-S1, population A). Thus, using axes 1 and 2 there are three clusters present: Micaela Bastidas, Pinto Recodo, and all other populations. These results are consistent with the Structure results (see below), where, when two populations are allowed, Micaela Bastidas is distinct from all other populations, and when three or more populations are allowed, Pinto Recodo is recovered as a distinct group. We performed clinal analyses on FCA axis 1 and found that the data best fit a sigmoidal model of variation along the transect (Table 2-S7, Fig. 2-3). However, the estimates for center (43.26 km) and width (29.93) were highly discordant relative to the color-pattern clines, and showed wide confidence intervals (Table 2-S7).

The global regression analysis shows that the best supported model was one where each variable (dorsal color, dorsal pattern, and microsatellite FCA axis 1) has a unique center and width parameter. However, by inspecting the AICc values of each model, we find that constraining the microsatellite cline along with the color-pattern clines (Table 2-S5, models b–d) greatly reduces model fit (ΔAICc = 74.0) compared to models where microsatellite clines were unconstrained (Table 2-S5, models –e–g). Also, constraining either center or width to be equal among the color-pattern clines substantially reduces model fit (ΔAICc when centers constrained = 19.83, ΔAICc when widths constrained = 15.27), indicating that the clines for these two metrics have distinct shapes and positions.

Landscape genetics

For the banded-striped transect, the best-supported number of populations in the STRUCTURE analysis was $K = 3$ (Delta K peak = 41.1). These three populations are Sauce (banded morph, pop. 1 on Fig. 2-2), Micaela Bastidas (striped morph, pop. 16 on Fig. 2-2), and a phenotypically heterogeneous central Huallaga group (populations 2–15 on Fig. 2-2). Most of the genetic structure appears to be associated with geographic distance. For example, Sauce is on the southern end of the transect, and separated from the nearest sampling locality (Vaquero) by a mountain range rising to approx. 1200 m elevation (for comparison, the highest recorded elevation for *R. imitator* is 945 m). Similarly, Micaela Bastidas is on the northern end of the transect, roughly 60 km from the nearest sampling locality. Therefore, the genetic distinctness of these two populations is likely due to isolation-by-distance. The central Huallaga group spans the banded-striped transition zone, yet there is no discernible genetic structure across the phenotypic transition (Fig. 2-2). Using multiple-matrix regression (Wang 2013), we determined that the geographic distance between populations was significantly correlated with genetic distance (β_D= 0.752, $P = 0.0018$), whereas color-pattern distance was not ($\beta_{CP} = 0.037$, $P = 0.421$). These results support an isolation-by-distance model of genetic divergence among populations, and provide quantitative support for the interpretation of the STRUCTURE analysis.

For the spotted-striped transect, the best-supported number of populations in the STRUCTURE analysis was $K = 2$ (Delta K peak = 329.1). The two populations correspond to Micaela Bastidas (striped morph, pop. H on Fig. 2-2) and a larger group containing populations A–G (Fig. 2-2). This latter group is distributed in the Cordillera Escalera mountains and surrounding lowlands, and includes pure spotted populations (e.g., San Jose, pop. C), and pure

striped populations (e.g. Pongo, pop. G). Increasing the values of *K* to 3 or 4 reveals additional structure in that the Pinto Recodo population is distinct, however, we did not find any genetic structure across the phenotypic transition zone. Again, genetic structure here seems to be associated with geographic isolation, in that the geographically distant populations Micaela Bastidas and Pinto Recodo are distinct. The multiple matrix regression analysis revealed a marginally significant correlation between geographic distance and genetic distance (β_D= 0.678, P = 0.054), but no correlation between color-pattern distance and genetic distance (β_{CP}= -0.317, P = 0.274). These results provide marginal support for an isolation-by-distance model of genetic divergence among populations, and reject the isolation-by-adaptation hypothesis.

Mate choice: Banded-striped

We had a total of 120 successful free-release trials across the four sampled populations (allopatric banded, n=35; transition zone banded, n=22; transition zone striped, n=24; allopatric striped, n=38). In total, we recorded 40.35 hours of courtship time across all trials, with an average of 20.3 minutes of total courtship time per trial. Comparing allopatric banded and allopatric striped populations (Fig. 2-4a), we found a significant effect of male origin on courtship time (χ^2 = 8.03, df = 1, P = 0.005), indicating that courtship preferences were significantly different between these two populations. This is likely due to the allopatric banded population's preference for its own morph, as the allopatric striped population did not seem to show a preference toward either morph, whereas the allopatric banded population showed a significantly higher courtship time with its own morph (Fig. 2-4a). We also found a significant effect of pair ID on courtship time (χ^2 = 8.65, df = 1, P = 0.003), indicating that female identity

influences courtship preferences exhibited by males. Comparing transition zone banded to transition zone striped populations (Fig. 2-4b), we found no significant effect of male origin on courtship time ($\chi^2 = 3.32$, df = 1, $P = 0.069$). In addition, we found no significant effect of pair ID on courtship time ($\chi^2 = 1.86$, df = 1, $P = 0.173$).

For the plastic model mate choice experiments, we had a total of 85 successful trials across the four sampled populations (allopatric banded, n=22; transition zone banded, n=21; transition zone striped, n=22; allopatric striped, n=20). We recorded a total of 18.39 hours of interaction time (i.e., time on platform), with an average of 18.4 minutes of interaction time per trial. We found no significant effect of male morph on interaction time ($\chi^2 = 0.01$, df = 1, $P = 0.929$) and no significant effect of transect position on interaction time ($\chi^2 = 2.27$, df = 1, $P = 0.132$) (Fig. 2-S4a). There was also no significant interaction between male morph and transect position with respect to interaction time ($\chi^2 = 0.46$, df = 1, $P = 0.495$) (Fig. 2-S4a). For calls (Fig. 2-S4b), there were fewer (total n=56) successful trials (allopatric banded, n=17; transition zone banded, n=11; transition zone striped, n=12; allopatric striped, n=16) because males interacted with plastic models but did not call in some trials. We recorded a total of 2,003 calls directed towards plastic models, with an average of 36 calls per trial. We found a significant effect of male morph on calling behavior ($\chi^2 = 5.43$, df = 1, $P = 0.0197$), with banded males directing more calls toward models with their own morphotype (Fig. 2-S4b). However, there was no significant effect of transect position ($\chi^2 = 0.192$, df = 1, $P = 0.661$). Finally, there was no significant interaction between the effects of male morph and transect position with respect to calling behavior ($\chi^2 = 0.926$, df = 1, $P = 0.336$).

Mate choice: Spotted-striped

For the spotted-striped mate choice experiment, we had a total of 94 successful free-release trials across the three sampled populations (spotted, n=40; transition zone striped, n=29; allopatric striped, n=25). We recorded a total of 34.92 hours of courtship time across all trials, with an average of 22.3 minutes of total courtship time per trial. We found a marginally significant effect of male origin on courtship time ($\chi^2 = 4.93$, $P = 0.0852$), and a strong effect of pair ID on courtship time ($\chi^2 = 10.16$, $P = 0.0014$). The marginal effect of male origin may be due to the apparent preference by each striped population for its own morph (Fig. 2-5). However, as the effect is not significant, we cannot conclude that mate preferences are different among populations tested in this experiment.

Discussion

Using quantitative analyses of color-pattern in combination with cline analyses, we have demonstrated the existence of two narrow mimetic transition zones in the poison frog *Ranitomeya imitator*. The banded-striped transition zone is characterized by abrupt shifts in color and pattern on both the dorsum and the hindleg. The widths of these shifts varied substantially among different color-pattern metrics. Dorsal color showed the narrowest cline width of 1.16 km, whereas leg pattern showed the widest cline width of 8.97 km. In the spotted-striped transition zone, the transition is characterized by an abrupt shift in dorsal color, a relatively gradual shift in dorsal pattern, and a smooth (linear) shift in both leg color and leg pattern. Dorsal color showed a relatively narrow cline of 2.94 km wide, whereas the cline in dorsal pattern was much wider at 14.31 km. One explanation for differences among cline widths on

phenotypic traits is that the strength of selection differs across the traits (Endler 1977; Haldane 1948; Mallet *et al.* 1990). For a cline stabilized by positive frequency dependent selection, the width of a cline can be expressed as where σ is dispersal distance, s is the selection coefficient, K is a constant (here, we used $K = \sqrt{12}$ following (Mallet & Barton 1989), and w is the width of a cline (Endler 1977). We can then calculate the strength of selection needed to produce a cline of the observed width, given our dispersal estimate (for *R. imitator*, 0.095 km/gen; see (Twomey *et al.* 2014). For the banded-striped transition zone, this gives a selection coefficient on dorsal color of 8.03%, compared to 0.13% for leg pattern. For the spotted-striped transition zone, dorsal color shows a narrower width (2.94 km) than dorsal pattern (14.31 km), giving selection coefficients on dorsal color and dorsal pattern of 1.25% and 0.05%, respectively. We caution that these estimates are highly contingent the dispersal estimate for *R. imitator*, which has never been measured directly, but only inferred via population genetic methods. For example, a two-fold increase in the dispersal estimate would increase the selection coefficient for dorsal color to 32%. Also, there is substantial uncertainty surrounding width estimates for many of the color-pattern metrics. Looking at the confidence intervals, the width for the dorsal color cline in the banded-striped transect could be as wide as 8.08 km, while the dorsal pattern cline in the spotted-striped transect could be as wide as 30.38 km.

Between the two transition zones, we generally see higher estimated selection coefficients for dorsal color than dorsal pattern, indicating that color may be under stronger selection for mimicry than pattern. For leg color and pattern, we get two distinct trends comparing the two transition zones. In the banded-striped transition zone, both leg color and leg pattern follow a clear sigmoidal cline, whereas in the spotted-striped transition zone, leg color and pattern follow

gradual, linear trends, and are only subtly different among the spotted and striped morphs of *R. imitator*. This can probably be explained in terms of model species variation, as the highland and lowland morphs of *R. variabilis* have similar leg coloration (greenish blue reticulation), whereas lowland *R. variabilis* and *R. summersi* have distinct leg coloration (greenish blue reticulation vs. orange bands). For the banded morph of *R. imitator*, it is perhaps surprising that the color-pattern of the legs so clearly corresponds to *R. summersi,* given that the legs are a seemingly small component of the overall mimetic signal. The most likely explanation is that the legs are important for mimicry, and thus subject to selection for mimetic resemblance. Indeed, leg color also seems to be important for mimicry in the striped-varadero transition zone, as does the color of the upper arms (Twomey *et al.* 2014). An alternative possibility is that the color-pattern of both the legs and dorsum in the banded morph are controlled by the same set of genes, or that the dorsal banding genes are epistatic to the leg color genes.

Despite strong phenotypic differences across mimetic transition zones, we found no associated genetic structure across these zones. Rather, population genetic structure was always associated with allopatric populations rather than distinct phenotypes (Fig. 2-2). This was also supported by the cline analysis of the microsatellite data, where we used a factorial correspondence analysis to reduce the microsatellite variation to a single descriptive axis. If mimetic morphs were reproductively isolated, we would expect to see sharp steps in the microsatellite clines at the same locations on the transects as the color-pattern clines. Such a pattern was not observed. Instead, the microsatellite clines had drastically different width and center estimates than any of the color-pattern metrics (Fig. 2-3, Tables 2-S6 & 2-S7). This lack of mimicry-associated genetic structure was also confirmed by our landscape genetic analyses.

For both transition zones we found that among-population genetic distance was correlated with geographic distance, but not color-pattern distance (Fig. 2-S5), suggesting that color-pattern divergence across these two transition zones has not resulted in a barrier to gene flow. These results stand in contrast to the results from the striped-varadero transition zone (Twomey *et al.* 2014), where genetic divergence was correlated with both color-pattern distance and geographic distance. In another dendrobatid species, *Oophaga pumilio*, (Wang & Summers 2010) found that genetic divergence was primarily associated with divergence in dorsal coloration. One mechanism for morph-associated genetic isolation is 'immigrant inviability', where immigrant individuals are at a selective disadvantage due to a phenotype-environment mismatch (Nosil *et al.* 2005). This could come about in two ways. First, because aposematic phenotypes are predicted to be under positive frequency dependent selection, individuals possessing the wrong phenotype should experience higher predation risk. The possibility of immigrant inviability via increased predation has been supported by a number of studies in dendrobatids (Chouteau & Angers 2011, 2012; Comeault & Noonan 2011; Noonan & Comeault 2009), where reciprocal transplants in colored clay models have generally found higher attack rates on foreign versus local phenotypes. Second, immigrants may be at a mating disadvantage, as several studies in dendrobatids have now found evidence for morph-based positive assortative mating (Maan & Cummings 2008; Reynolds & Fitzpatrick 2007; Richards-Zawacki *et al.* 2012; Richards-Zawacki & Cummings 2011; Summers *et al.* 1999; Twomey *et al.* 2014). However, in the case of *R. imitator*, both transition zones studied here are substantially wider (roughly 3–6 kilometers) than the species' estimated dispersal capabilities (0.095 km/generation), suggesting that for a given individual, dispersal across morph boundaries is not feasible. Rather, dispersal and gene

flow between morphs most likely occurs through the transition zone. Interestingly, in a clay model study of the spotted-striped transition zone, (Chouteau & Angers 2012), found that there was much lower predation overall (on any morph) in the transition zone sites than either of the pure striped or pure spotted populations. This reduced predation pressure may facilitate the persistence of nonmimetic hybrids.

The results from our free-release mate choice experiments indicate that most of the populations studied here do not show any discernible preference for their own morph versus an alternate morph. However, there was one exception: the allopatric banded morph demonstrated a significant preference for its own morph. Furthermore, as this preference for banded frogs was not seen in the allopatric striped population, we can conclude that this preference is not due to banded females simply being more attractive (for example, by virtue of their larger body size). However, we found no preferences in the transition zone populations, with both the transition zone banded and striped morphs showing random courtship with respect to morph type. This relaxed assortative mating at the transition zone is consistent with the population genetic results, which show a lack of genetic structure across morph boundaries. As neither the striped nor banded populations from the transition zone appear to be choosy, this may facilitate gene flow across morph boundaries. Similarly, for the spotted-striped transition zone, we found that none of the three populations tested showed a significant preference for either morph. There was a general trend for striped frogs to prefer their own morph (Fig. 2-5), but this preference was not statistically significant. This is also consistent with the lack of genetic structure across the mimetic transition zone, which may be due to random mating with respect to morph.

As the free-release trials allow frogs to make courtship decisions using all available cues (e.g., color-pattern, acoustic, behavioral, olfactory), these trials provide an assessment of the premating isolation between two morphs when all potential cues are present. However, as these trials do not isolate the cues used in mate choice, we conducted additional mate choice experiments using plastic frog models in order to determine whether assortative mating was specifically based on color-pattern. We found mixed results depending on our choice of response variable. When measuring interaction time, we found no evidence for preferences in any population. However, when measuring calling behavior, we found that banded males directed more calls towards banded models than striped models. Calling behavior is a more stringent response variable in that a male could "interact" with a model without necessarily showing courtship interest. Interaction time was simply scored as time spent on the model's platform, with no information on male orientation or interest, whereas calling is a behavior specifically associated with courtship. Based on these results, color-pattern may be one of the cues involved in mate choice in this system. Color-pattern may therefore be acting as a 'magic-trait', that is, a trait subject to divergent ecological selection that also mediates mate choice (Gavrilets 2004). Identifying such traits is of interest to speciation researchers in that they are thought to greatly facilitate speciation-with-gene flow. In *Heliconius* butterflies, wing color-pattern represents a good example of a magic trait (Jiggins *et al.* 2001; Merrill *et al.* 2012). Selection for Müllerian mimicry has caused wing color-patterns to diverge as different populations participate in distinct mimicry rings, and there is substantial evidence that wing color-pattern itself mediates assortative mating (Jiggins *et al.* 2001, 2004). Our mate choice experiment using plastic models provides evidence that the color-pattern in *R. imitator* may be acting in a similar way.

In a recent paper (Twomey *et al.* 2014), we presented a study on the striped-varadero transition zone, where we used quantification of color-pattern, cline analysis, landscape genetics, and mate choice experiments to infer levels of reproductive isolation between the striped and varadero morphs (Fig. 2-1). In that study, we found that color-pattern clines were narrower (approx. 1–4 km, see Fig. S2-6), there was strong genetic structuring between morphs at the transition zone, and the striped morph from the transition zone displayed a significant mating preference for its own morph (whereas the striped morph from an allopatric population did not) (Table 2-1). Thus, in the striped-varadero transition zone, it appears that mimetic divergence has led to a breakdown in gene flow between morphs, and this may represent a case of incipient ecological speciation. However, in the two transition zones studied here, there is no indication that mimetic divergence has led to morph-based reproductive isolation. Thus, among the three different transition zones in *R. imitator*, there is variation in the extent to which mimetic divergence has led to reproductive isolation among morphs (Table 2-1). The *R. imitator* mimicry system is useful for comparative analyses of the factors influencing progress toward speciation given that we observe one case of mimicry-associated reproductive isolation, and two cases without such isolation. Furthermore, as these mimetic populations are all members of the same species, we can identify isolating barriers that arise in the early stages of population divergence.

Hypotheses explaining differential progress towards ecological speciation often focus on the nature of environmental heterogeneity. For example, in cline theory, important distinctions have been made between gradual environmental changes (gradient model) versus abrupt, step-like environmental changes (ecotone model) (Endler 1977). In the case of a mimetic organism, the environmental context of a mimetic adaptation is dictated in part by the geographic distributions

of the model species. Thus, as the goal of this study is to understand conditions that promote ecological speciation across mimetic transition zones in *R. imitator*, we first discuss the geographical distributions of the model species in order understand the ecological context of each mimetic shift.

Model species

The distributions of the model species can be summarized as follows: *Ranitomeya summersi* has a restricted range, corresponding almost exactly to the distribution of the banded morph of *R. imitator* (Fig. 2-1) (Brown *et al.* 2008b). *Ranitomeya variabilis* is widespread throughout the Amazonian lowlands and east-Andean foothills of Peru (south to roughly 10.5° south latitude), Ecuador, and southern Colombia (to approximately 1.5° north latitude) (Brown *et al.* 2011). This species displays two primary color-morphs: a striped morph (with yellow dorsolateral stripes), which occurs in the lowlands (roughly 150–400 m elevation), and a spotted morph (with green dorsal reticulation), which occurs in the highlands (roughly 400–1600 m elevation). Finally, *R. fantastica* is known from a handful of localities the departments of San Martin (northern part) and Loreto (southwestern part). Its distribution appears to be bounded by the Huallaga River (to the east) and the Marañón River (to the north), and extends into the east-Andean foothills up to roughly 1000 m elevation in the southern margin of its range. This species is highly polytypic throughout its range (Brown *et al.* 2011), however, the morph that serves as the model for the varadero morph of *R. imitator* has only been found near the village of Varadero.

In *R. imitator*, the most widespread morph is the striped morph, which occurs broadly throughout the lowlands of northern San Martin and southern Loreto departments. We have

several lines of evidence that lead us to consider this as the ancestral *R. imitator* morph. First, the striped phenotype is found in other, closely related species (e.g., *R. sirensis*), indicating that this may be an ancestral phenotype. Second, from a biogeographical perspective, the two species in the clade sister to *R. imitator* (*R. vanzolinii* and *R. flavovittata*) occur in the lowlands of eastern Peru. This suggests possible colonization routes via the lowlands between the Ucayali and Huallaga Rivers, which is an area where the striped morph of *R. imitator* is known to occur. Thus, a likely colonization history was westward expansion across the lowlands of northern San Martin and southern Loreto, followed by dispersal into the highlands of northern San Martin. In this scenario, as the lowland striped morph colonized new habitats, it experienced novel selective pressures for mimetic shifts towards either the spotted, banded, or varadero morphs.

In the banded-striped transition zone, the mimetic shift in *R. imitator* can be explained by the parapatric distributions of the model species (*R. summersi* and *R. variabilis*). In vicinity of this transition zone, our collection records indicate that the striped morph of *R. variabilis* occurs from the lowlands as far south as Callanayacu (locality 10 in Fig. 2-S1). Similarly, for *R. summersi*, this species is relatively common near Sauce (locality 1, Fig. 2-S1), occurs as far north as the Callanayacu North site (locality 8, Fig. 2-S1). Therefore, at least at a local scale, the two model species have abutting distributions at Callanayacu, which is precisely where we observe a mimetic shift in *R. imitator*. Thus, as *R. imitator* radiated from the lowlands into this region, populations colonizing regions southwest of Callanayacu found themselves in a novel mimetic environment in that *R. summersi* was now the only model species present in the area.

In the spotted-striped transition zone, the mimetic shift in *R. imitator* can be explained by the recurrent shift in *R. variabilis* from the lowland striped morph to the highland spotted morph.

This shift from yellow, striped *R. variabilis* in the lowlands to green, spotted *R. variabilis* in the highlands has been observed in a number of sites from central Peru to Ecuador (Brown *et al.* 2011). These shifts are known to occur in areas far outside the geographic range of *R. imitator*, and furthermore, phylogenetic analyses strongly reject a monophyletic highland spotted clade. This indicates that that the highland/lowland polytypism in *R. variabilis* has evolved several times independently, and that its evolution is independent of the presence/absence of *R. imitator* populations. The phenotypic shift in *R. variabilis* is also apparent in the transition zone studied here, which occurs at the interface between highland and lowland habitats. Thus, the shift from the striped to the spotted morph in *R. imitator* appears to simply track local variation in the model species along the elevation gradient. The cause of the elevation-associated polytypism in *R. variabilis* remains unknown, and invites further investigation.

The striped-varadero transition zone cannot be explained by the distributions of model species. In this area, both *R. variabilis* and *R. fantastica* are broadly distributed and present on both sides of the cline. Although we do not know the geographic distribution of the "varadero morph" of *R. fantastica,* a population over 30 km to the south has a generally similar appearance (orange head with blue reticulation on body and legs), so it is likely widespread throughout this region. Thus, it appears that *R. imitator* has shifted from one model species to another, with no clear explanation such as model species presence/absence. There are a number of possible explanations, which should be investigated further. Theory predicts that when model species differ in their abundance or palatability, a mimic species will be under stronger selection to evolve to resemble the more abundant or better-protected model (Joron & Mallet 1998). Thus, differences in abundance and/or toxicity between *R. variabilis* and *R. fantastica* could cause

variation across the landscape in terms of which model species confers optimal protection. In terms of abundance, in most sites *R. variabilis* and *R. fantastica* are both rare relative to *R. imitator* (Twomey *et al.* 2013), but there is no evidence of any shift in relative abundances of either species near Varadero. In terms of toxicity, Varadero *R. fantastica* possess slightly higher quantities of alkaloids than two other analyzed populations of *R. variabilis* (Stuckert *et al.* 2014a), but again there is no evidence of any shift towards higher toxicity in Varadero. Thus, the shift in *R. imitator* from one model species to another remains poorly understood.

Differential progress along the 'speciation continuum'

Given the wide range of examples highlighting the variation in completeness of ecological speciation (Nosil *et al.* 2009), a key goal of speciation research to understand the reasons for this variation. Among the three transition zones in *R. imitator*, we see two transition zones with little or no reproductive isolation, and one transition zone with partial, mimicry-associated reproductive isolation. What explains variation in the amount of reproductive isolation across these three transition zones? Is there something unique about the striped-varadero transition zone that explains its greater progress toward speciation than the other transition zones?

Factors influencing progress toward speciation generally fall into two classes: nonselective and selective (Nosil 2012; Nosil *et al.* 2009). Here, we address two nonselective factors (geographic barriers and secondary contact) and two selective factors ("stronger selection" and "multifarious selection") to understand differential progress toward speciation in *R. imitator*. First, geographic isolation may facilitate ecological speciation when geographic barriers occur between ecomorphs. Populations that are divergent ecologically and also separated by some

geographical barrier should diverge more readily than if that geographic barrier were absent, as in the latter case divergent selection is no longer counteracted by gene flow (Nosil 2012). This may be particularly relevant for clines stabilized by positive frequency dependent selection, in which clines can form in seemingly arbitrary geographic areas, but can shift position over time, and may eventually become "trapped" at geographical barriers (Bazykin 1969; Blum 2002; Mallet 1986; Mallet & Barton 1989). In *R. imitator*, we might therefore expect to see geographic barriers separating mimetic morphs, especially at Varadero, where isolation is strongest. However, this is not what we observe, as the transition zone at Varadero occurs across approximately 1–2 km of unimpeded lowland rainforest habitat. While there is a small river nearby, this river is roughly 1 km south from the transition zone, and striped populations on either side of the river are not genetically isolated (Twomey *et al.* 2014). In the banded-striped transition zone, the Huallaga River appears to be a potential barrier between morphs, although this is not what we observe. Instead we see no associated genetic structure or pronounced morphological differences across the river. For example, on Figure 2-3a, populations 6–7, 9–10, 11–12, and 14–15 are populations on opposite sides of the Huallaga River. However, further upstream, this river may be a barrier between the spotted and banded populations, as we have found pure spotted and pure banded morphs occurring on north and south sides of the river, respectively. We have not found intermediates between the banded and spotted morphs, which may suggest these two morphs are somewhat isolated.

Second, historical allopatry between ecomorphs may also promote speciation (Rundle & Nosil 2005). In this view, previously isolated populations with subsequent secondary contact should exhibit greater reproductive isolation than populations that were isolated for a shorter

time, or were never isolated at all. As isolated populations are no longer subject to the homogenizing effects of gene flow, over time, these populations may diverge in seemingly arbitrary characters. Thus, a signature of secondary contact between populations is congruent clines in many characters, not just those subject to divergent ecological selection (Coyne & Orr 2004). The striped-varadero transition zone is unique in that, in addition to clines in color-pattern elements, there are congruent clines in advertisement calls, body size, and microsatellites (Twomey *et al.* 2014). Compared to the two other transition zones, which show no clines in non-mimicry traits, the striped-varadero zone is therefore the most suggestive of secondary contact. However, distinguishing whether hybrid zones are due to primary divergence or secondary contact is difficult (Barton & Hewitt 1985), and late-stage primary transition zones could show clines at neutral loci (and presumably neutral traits) provided divergent selection is strong enough to pose a barrier to gene flow (Feder & Nosil 2010). In terms of biogeographical evidence, there is nothing to suggest any historical isolation at the striped-varadero transition zone as there are no major rivers or mountains in the area. Thus, whether the striped-varadero transition zone is a result of primary divergence or secondary contact remains unknown.

We focus on two selective hypotheses explaining progress toward speciation. The "stronger selection" hypothesis predicts that speciation is facilitated by strong selection on a single trait, whereas the "multifarious selection" hypothesis predicts that speciation is facilitated by divergent selection on multiple, independent traits (Nosil *et al.* 2009; Nosil & Harmon 2009). In the latter case, the total strength of divergent selection should be higher when adaptation occurs along multiple independent niche axes, in which case multifarious selection more readily drives speciation by increasing the total strength of selection (Nosil & Harmon 2009). Assuming a

constant strength of selection, multifarious divergence may drive widespread genomic divergence as multiple loci are "pulled apart" by divergent selection, whereas strong selection on a single trait increases per-trait selection coefficients, facilitating divergence with gene flow (Mallet 2006). The "stronger selection" hypothesis has received support in studies finding positive correlations between reproductive isolation and divergence in one phenotypic trait, for example, body size (Funk *et al.* 2006; McKinnon *et al.* 2004) and color-pattern (Jiggins *et al.* 2004; Maan & Cummings 2008). In the context of the current study, the "stronger selection" hypothesis generates a clear prediction, which is that the magnitude of mimetic divergence will be positively correlated to levels of reproductive isolation. We can use the difference between *ymax* and *ymin* parameters from our clinal analyses of color-pattern to obtain an overall magnitude of mimetic shift across each transition zone. The difference between these parameters represents the overall cline height and therefore can be thought of as a mimicry effect size. Doing so, the banded-striped transition zone has the largest magnitude of mimetic divergence of 1.80, followed by the spotted-striped transition zone (1.35), and the striped-varadero transition zone (0.45). Thus, under the "stronger selection" hypothesis, we would predict the banded-striped and spotted-striped transition zones to be the most reproductively isolated, and the striped-varadero transition zone to be the least reproductive isolated, which is directly opposite from our observations. It is important to point out that this assumes mimetic color-pattern is effectively a single trait, which may not be the case. If different components of the mimetic color pattern are under independent genetic control, they may be better thought of as several different traits all undergoing divergent selection. Under the "multifarious selection" hypothesis, we would therefore predict that reproductive isolation should be positively correlated to the number

of color-pattern traits under divergent selection. If we treat each color-pattern metric as an independent trait, in the spotted-striped transition zone divergent selection is acting on two traits (dorsal color and dorsal pattern), in the striped-varadero transition zone on three traits (arm color, leg color, and body pattern), and in the banded-striped transition zone on four traits (dorsal color, dorsal pattern, leg color, and leg pattern). From this, we would predict the banded-striped transition zone to show the strongest reproductive isolation, which is again not what we observed. However, as we do not know how many independent genetic loci are involved in the color-pattern shift at each transition zone, we do not know if the shifts should be treated as one or many axes of divergence.

One observation related to multifarious selection is that there is evidence of a shift in body size at the striped-varadero transition zone (Twomey *et al.* 2014), while the other two transition zones show no such shift (Fig. 2-S7). Thus, at the striped-varadero transition zone, divergent selection may be acting both on mimetic color-pattern and body size, versus color-pattern only at the banded-striped and spotted-striped transition zones. Divergence in body size is frequently implicated in the evolution of reproductive isolation (Bolnick *et al.* 2006; Funk *et al.* 2006; McKinnon *et al.* 2004), and thus the observed reproductive isolation at the striped-varadero transition zone may have arisen from the combined effects of divergence in both mimicry and body size. While we do not yet understand the adaptive significance (if any) of the body size shift at the striped-varadero transition, there are several intriguing possibilities that warrant further research. First, (Twomey *et al.* 2014) suggested that body size could be related to mimicry, as *R. fantastica* (the model species of the larger varadero morph of *R. imitator*) is somewhat larger than *R. variabilis* (the model species of the smaller striped morph). Another

possibility is that size represents an adaptation to distinct habitat types present in this area. During our field work in the Varadero region, we noticed that the smaller striped morph is found in disturbed habitats while the larger varadero morph occurs in undisturbed primary rainforest. While this could be a coincidence (given that the striped morph occurs in closer proximity to a human settlement), there are at least two ecological explanations for a habitat-related shift in body size. First, if disturbed habitats contain a poorer available diet relative to undisturbed habitats, the body size shift may simply represent a plastic developmental response due to nutritional differences between the two habitats. Second, as *R. imitator* uses small phytotelmata (water-bodies formed in plants) for reproduction, if breeding-plant communities differ between the two habitats such that smaller plants (and phytotelmata) are found in the disturbed habitat while larger plants are found in the undisturbed habitat, body size could represent an adaptation for using smaller versus larger phytotelmata. While we do not have data on phytotelm size across the striped-varadero transition, it appears that smaller phytotelmata (e.g. *Calathea* spp.) predominate on the striped-side of the cline while larger phytotelmata (e.g. *Heliconia* spp.) seem to be more common on the varadero-side of the cline. Future studies on the striped-varadero transition should investigate the possibility of body size as an adaptive trait subject to divergent selection, and whether color-pattern and/or body size are cues that mediate assortative mating.

In summary, we conducted a study integrating several lines of evidence (Table 2-1) to infer levels of mimicry-associated reproductive isolation across two transition zones in *Ranitomeya imitator*, and found no evidence to suggest that these mimetic morphs are reproductively isolated. These results stand in contrast to the striped-varadero transition zone, where mimetic divergence appears to be driving reproductive isolation between two mimetic morphs of *R.*

imitator (Twomey *et al.* 2014). Whereas in the two transition zones studied here there is no evidence for assortative mating at either transition zone, the striped-varadero transition does show evidence of assortative mating at the transition zone. Furthermore, we find that, unlike the striped-varadero transition, the two transition zones studied here do not show any evidence of a shift in body size across the phenotypic transition. Given these results, we can conclude that the striped-varadero transition zone is unique in at least two ways: (1) it is the only transition zone with evidence of assortative mating at the interface between distinct morphs, and (2) it is the only transition zone showing a shift in a trait possibly unrelated to mimicry (body size). The latter result lends support to the "multifarious selection" hypothesis, which stipulates that ecological speciation is facilitated when divergent selection acts on multiple independent traits. These results also underscore the importance of assortative mating as an early-evolving barrier that may be important in the early stages of population divergence.

Acknowledgments

We thank Anders BL Larsen for discussion on image texture quantification; Morgan Kain for help with multiple-matrix regression; and Justin Touchon for help with generalized linear mixed models. We thank Jason Brown, Steven Chen, Victor Holmes, Tiffany Kosch, César Lopez, Manuel Guerra-Panaifo, Michael Mayer, Mark Pepper, Ronald Mori Pezo, Manuel Sanchez-Rodriguez, Lisa Schulte, Adam Stuckert, James Tumulty, and Justin Yeager for help in the field. This research was funded by a NSF DDIG (1210313) grant awarded to ET and KS, a National Geographic Society grant awarded to KS (8751-10) and the NCCB scholarship (2012) at East Carolina University awarded to ET. Research permits were obtained from the Ministry of

Natural Resources (DGFFS) in Lima, Peru (Authorizations No. 050-2006-INRENA-IFFS-DCB, No. 067-2007-INRENA-IFFS-DCB, No. 005-2008-INRENA-IFFS-DCB). Tissue exports were authorized under Contrato de Acceso Marco a Recursos Geneticos No. 0009-2013-MINAGRI-DGFFS/DGEFFS, with CITES permit number 003302. All research was conducted following an animal use protocol approved by the Animal Care and Use Committee of East Carolina University.

Figures and tables

Figure 2-1. Müllerian mimicry in *Ranitomeya imitator*. (a) Four mimetic morphs of *R. imitator* and corresponding model species. (b) Map of central Peru showing the geographical distribution of mimetic morphs of *R. imitator* (colored polygons). The numbered grey boxes show the three mimetic transition zones studied herein. Box numbers as follows: (1) banded-striped transition zone, (2) spotted-striped transition zone, (3) striped-varadero transition zone. The black dotted line separating the spotted and banded morphs indicates a putative riverine barrier. These two morphs come into close contact but appear to be isolated by the Huallaga River. Inset map (upper right) showing study area within Peru.

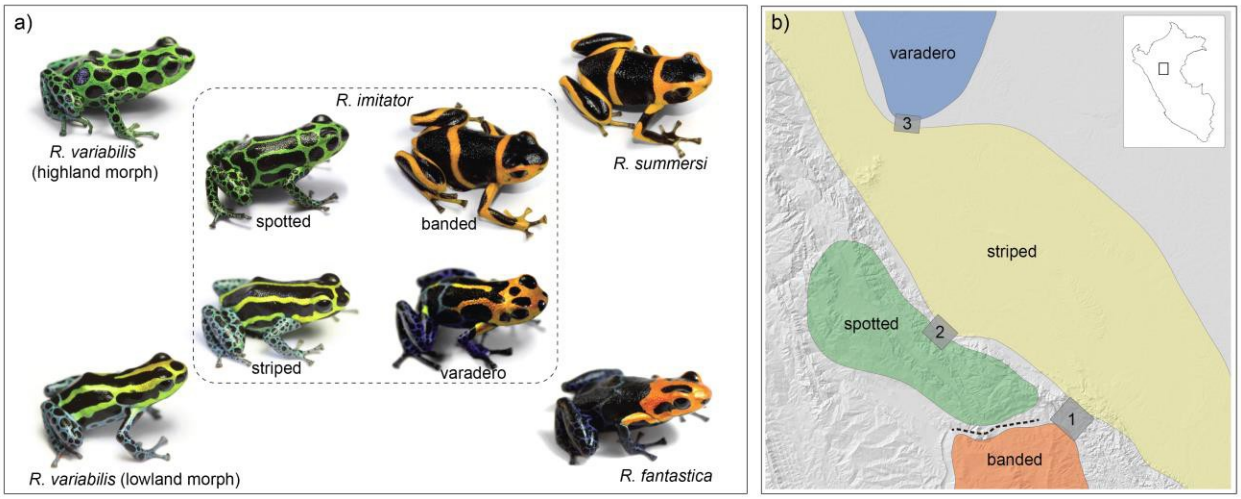


Figure 2-2. Structure plots based on microsatellite data for (a) banded-striped transition zone, and (b) spotted-striped transition zone. Vertical bars represent single individuals. Bar colors represent posterior probabilities of assignment to inferred genotypic group, where the number of groups (*K)* was allowed to vary from 2–4. For the banded-striped transition, the best-supported number of groups was *K*=3; for the spotted-striped transition, *K*=2. Horizontal bars indicate mimetic morph (upper) and sampling locality (lower). Sampling locality codes correspond to Tables 2-S1 and 2-S2 and Fig. 2-S1.

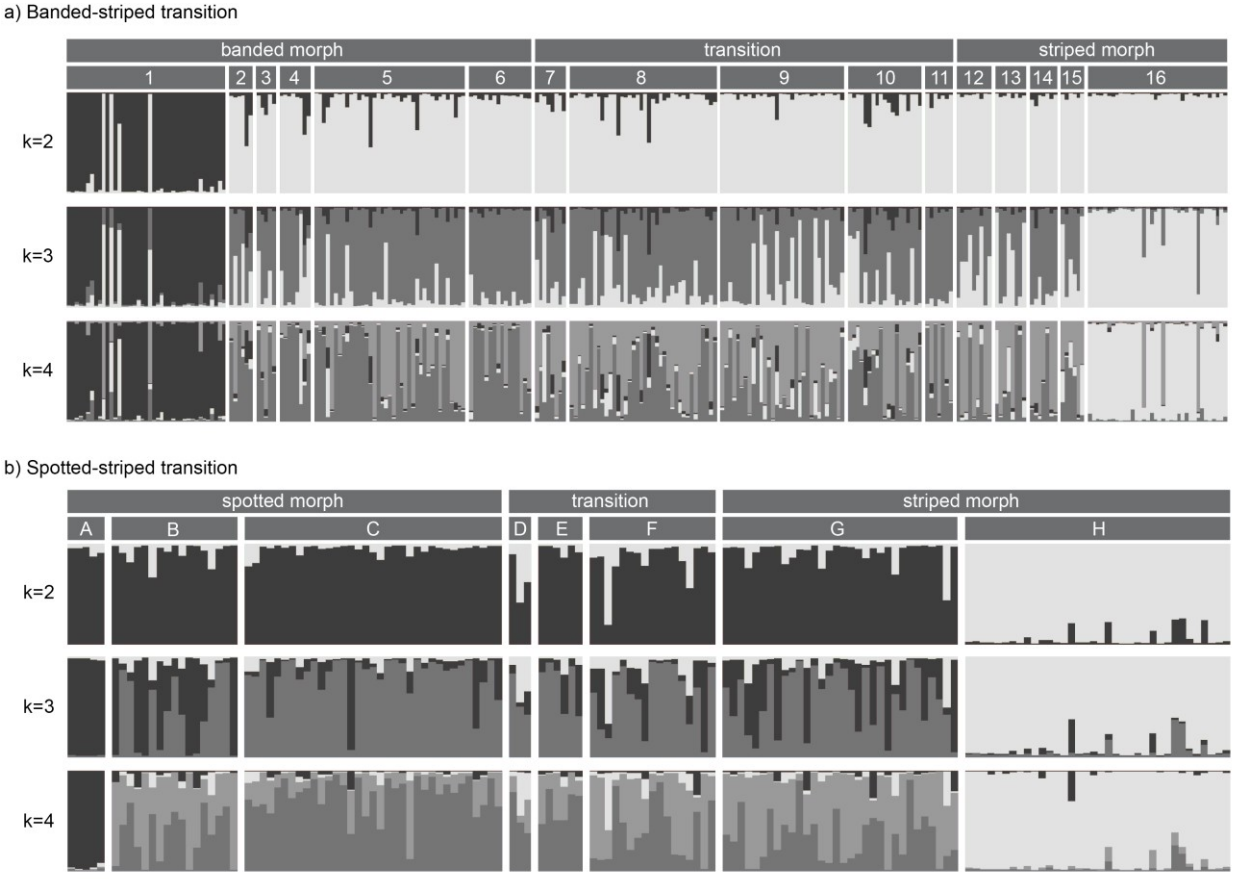
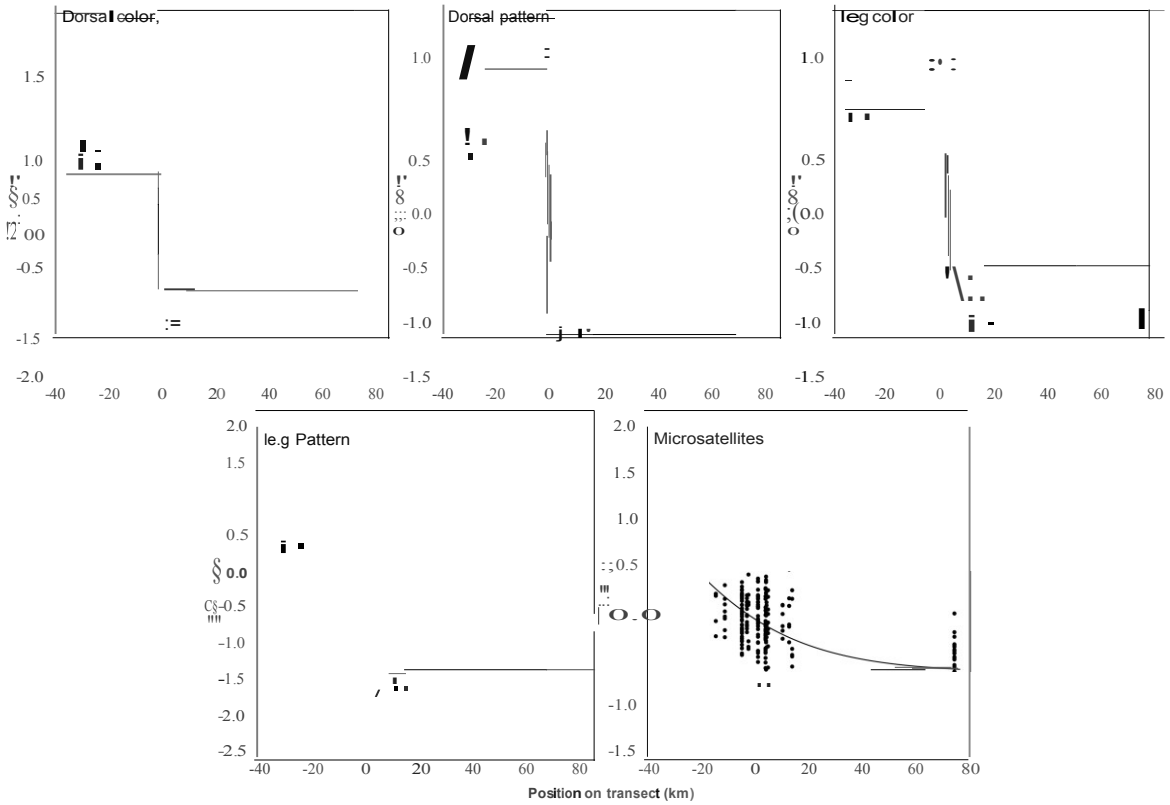
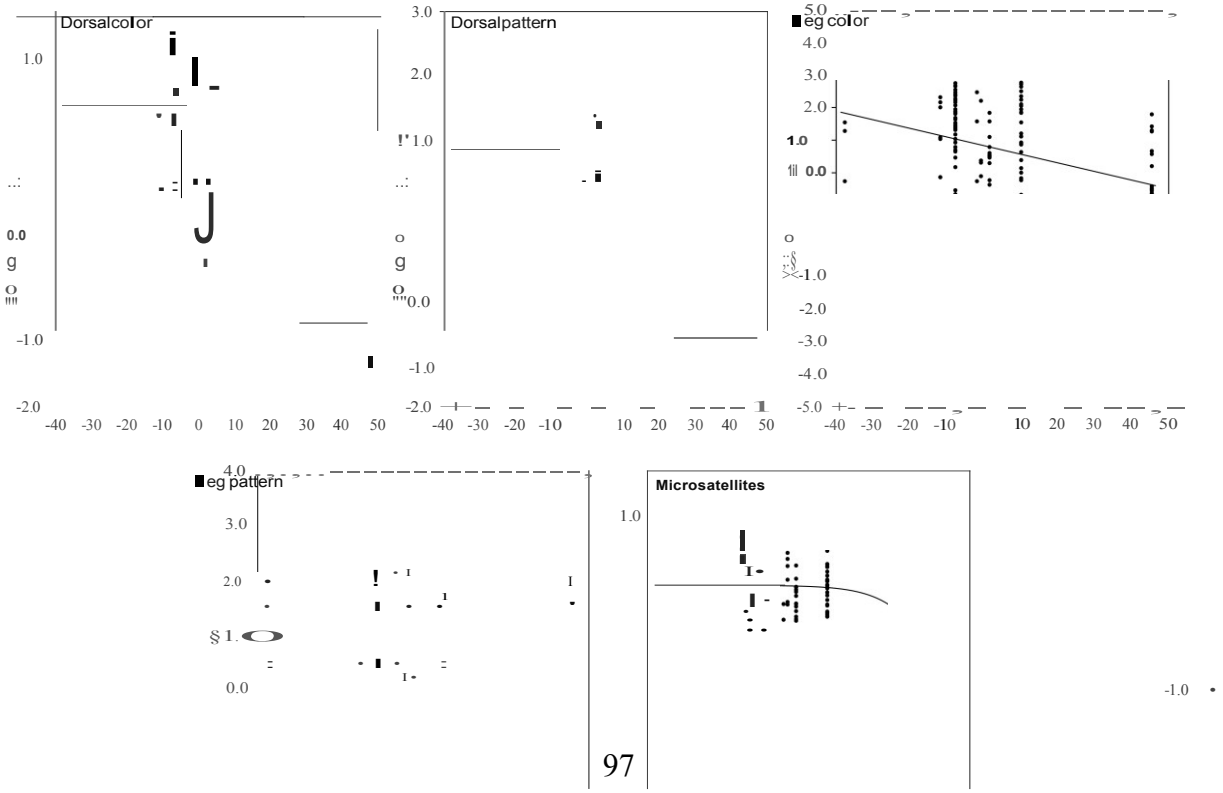


Figure 2-3. Clines in color-pattern metrics and microsatellites. (a) Banded-striped transition clines. (b) Spotted-striped transition clines. In all panels, trait values for individual *R. imitator* (represented by dots) are plotted along the sampling transect (x-axis, given in km). Color pattern metrics (dorsal color, dorsal pattern, leg color, and leg pattern) were derived from a kernel discriminant analysis. For the banded-striped transition (a), values closer to +1 indicate closer similarity to *R. summersi* and closer to -1 lowland *R. variabilis*. For the spotted-striped transition (b), values closer to +1 indicate closer similarity to highland *R. variabilis* and closer to -1 lowland *R. variabilis.* In the microsatellite panels, the y-axis is defined as the first axis from a factorial correspondence analysis. For all panels, the fit line represents the best-supported model according to the AICc. Fit lines were plotted using best-fit parameter estimates.

a) Banded-striped transition



b) Spotted-striped transition



-1.0

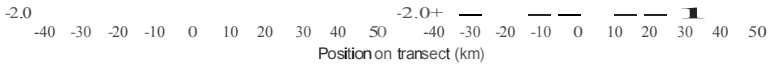

-2.0
-2.0+
-40 -30 -20 -10 0 10 20 30 40 50
-40 -30 -20 -10 0 10 20 30 40 50
Position on transect (km)

Figure 2-4. Boxplots showing courtship times in the banded-striped free-release mate choice experiment. (a) Allopatric populations experiment, where banded and striped frogs originated from the ends of the sampling transect. (b) Transition populations experiment, where banded and striped frogs originated from near the phenotypic transition zone. The asterisk indicates a statistically significant effect ($P < 0.05$) of male origin on courtship times from the GLMM.

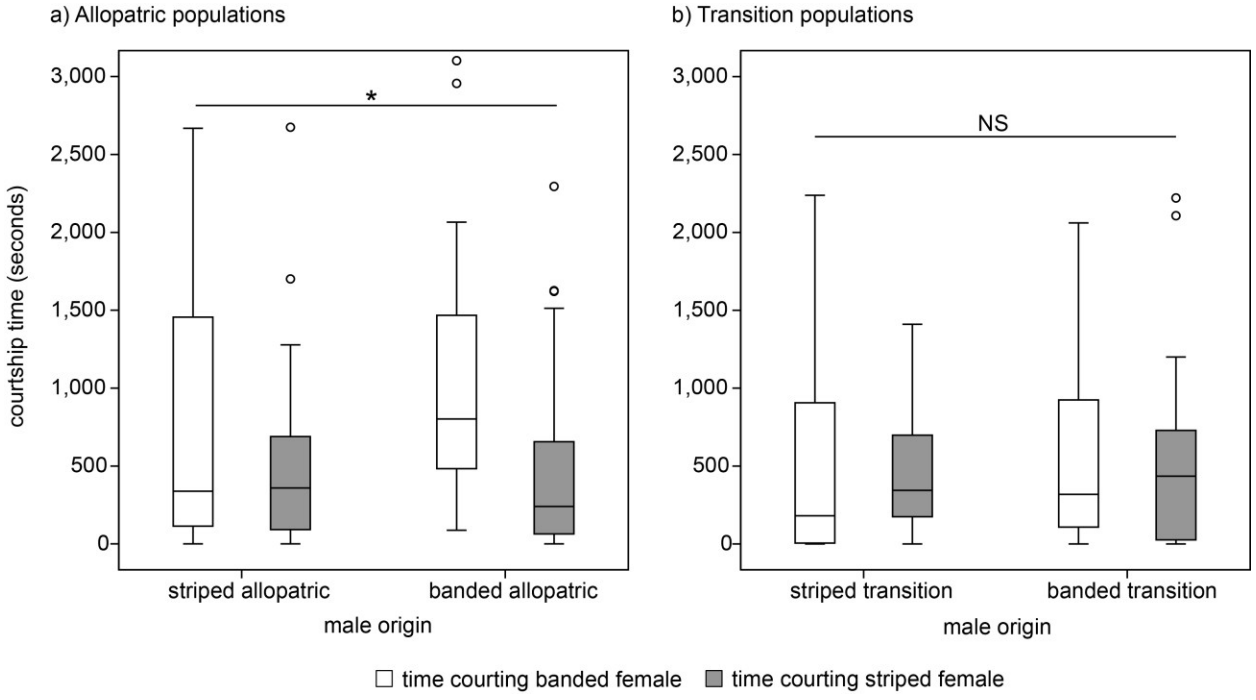


Figure 2-5. Boxplot showing courtship times in the spotted-striped free release mate choice experiment. Significance test is based on a GLMM testing for the effect of male origin on courtship time.

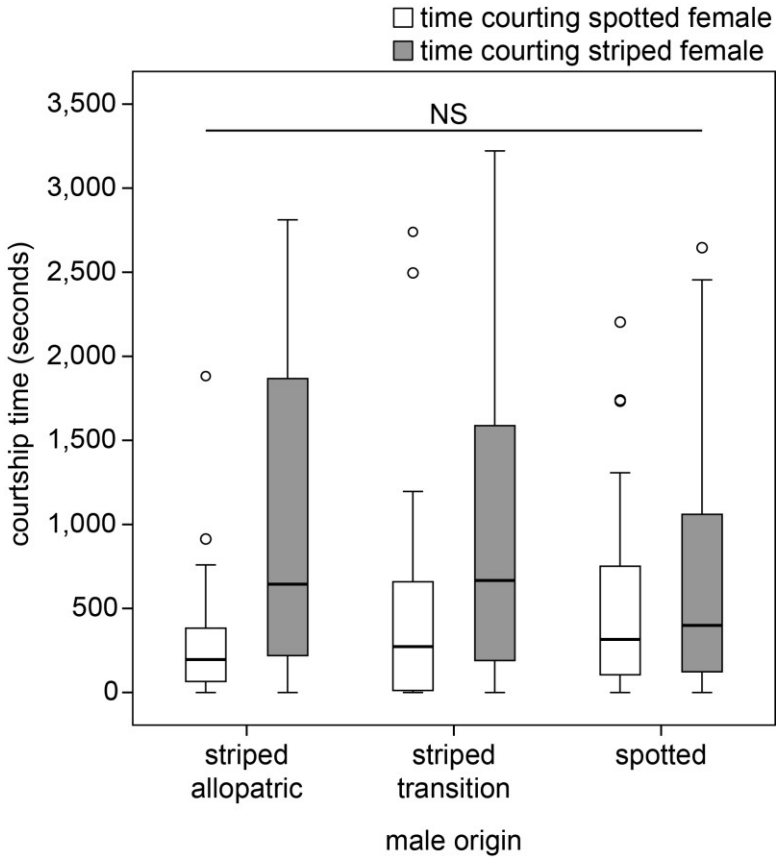


Figure 2-S1. Sampling localities and variation in *Ranitomeya imitator*. Orange dots show *R. imitator* populations included on the banded-striped transect; green dots show *R. imitator* populations included on the spotted-striped transect. Population numbers/letters correspond to Tables 2-S1 and 2-S2, respectively. Note that the Micaela Bastidas population (pop. H/16) is included in both transects.

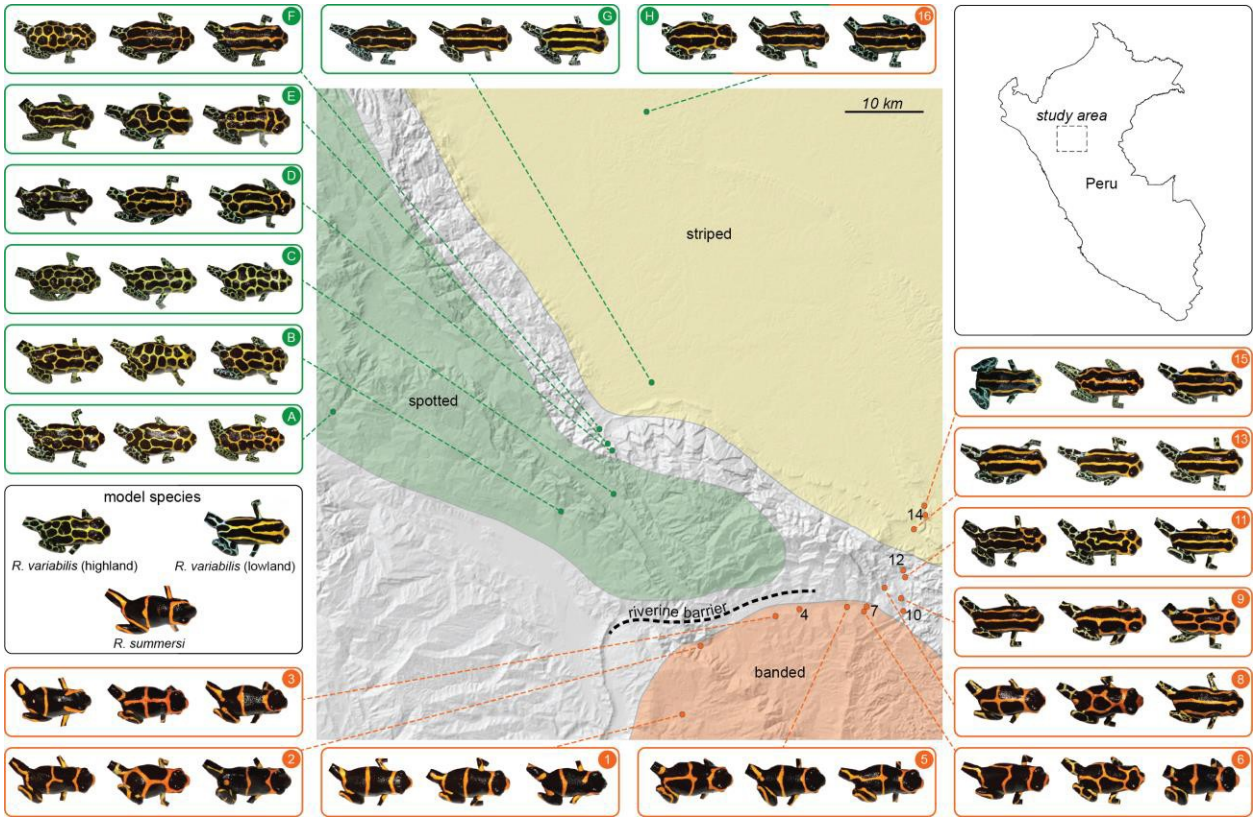


Figure 2-S2. Mutual information plot for color-pattern variables, banded-striped transect. Larger bars indicate greater variable contribution to the discriminant function. Panels are split by body region (dorsum/leg) and variable type (color/pattern). For color variables, prefixes are as follows: Qt = brightness, MS = brightness in yellow/blue range, LM = brightness in red/green range, C = chroma, H = hue. Pattern variables use the following naming convention: body region, metric, and associated parameters. For example, c3goh, sigma=2, b=0 indicates that the body region of interest was c3, the metric was goh (gradient orientation histogram), and the associated parameters for its extraction were sigma=2 and b=0. Body regions are as follows: c3=lower dorsum, c4=head, c1=right leg, c2=left leg. Extracted metrics are as follows: goh=gradient orientation histogram, sih=shape index histogram, bwratio=color/non-color ratio.

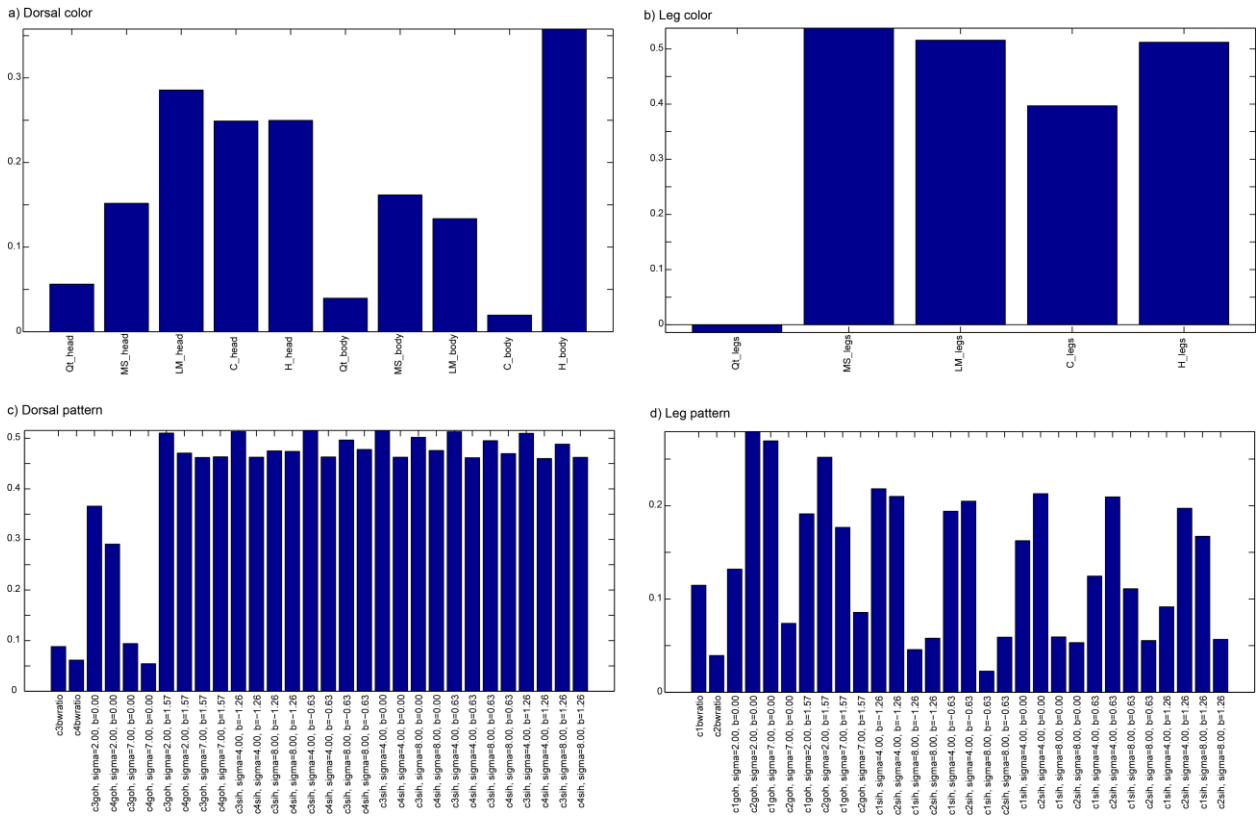


Figure 2-S3. Mutual information plot for color-pattern variables, spotted-striped transect.

Variable naming convention follows Figure 2-S2.

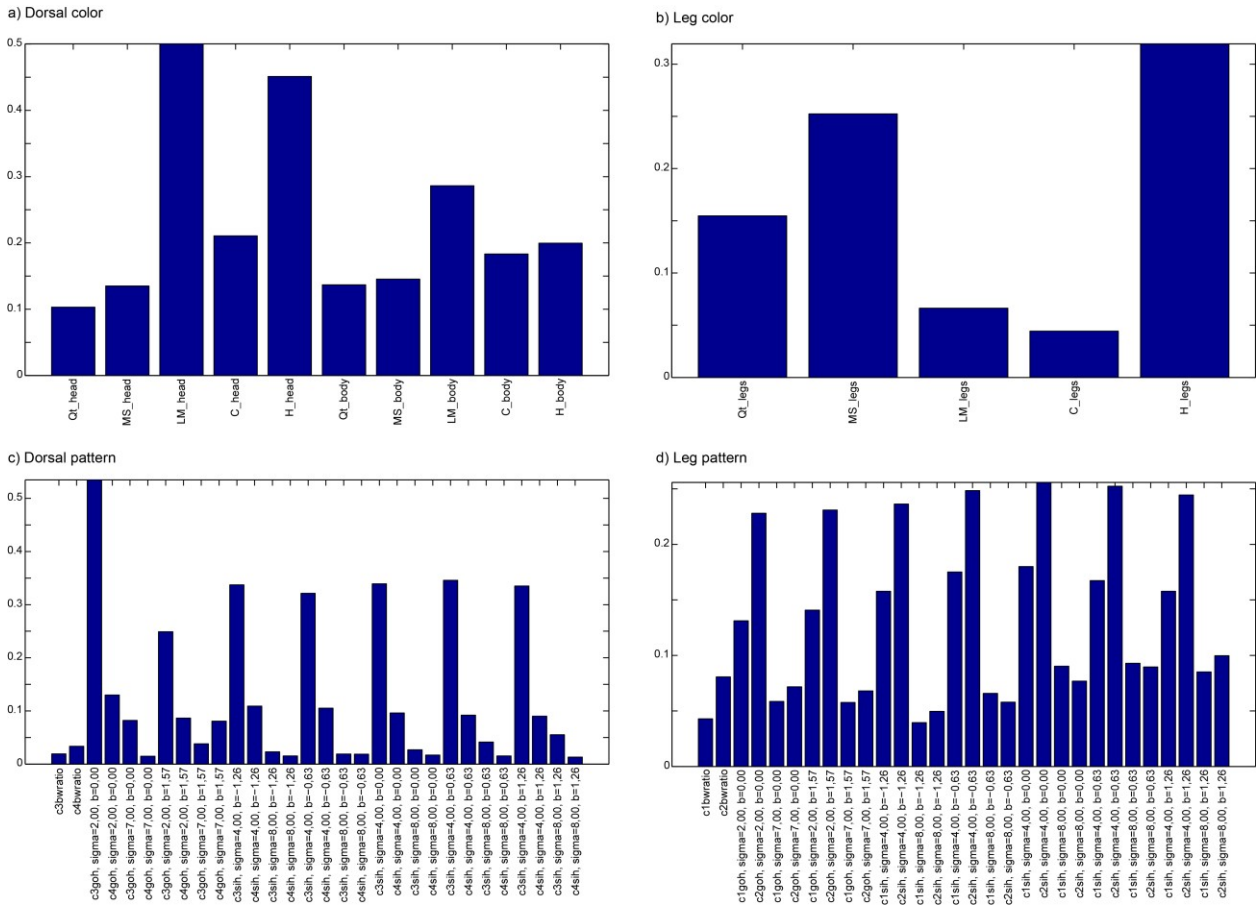


Figure 2-S4. Boxplots showing (a) association time and (b) number of calls from males directed towards striped and banded plastic models. Significance tests are based on a GLMM testing for the effect of male origin on response variables.

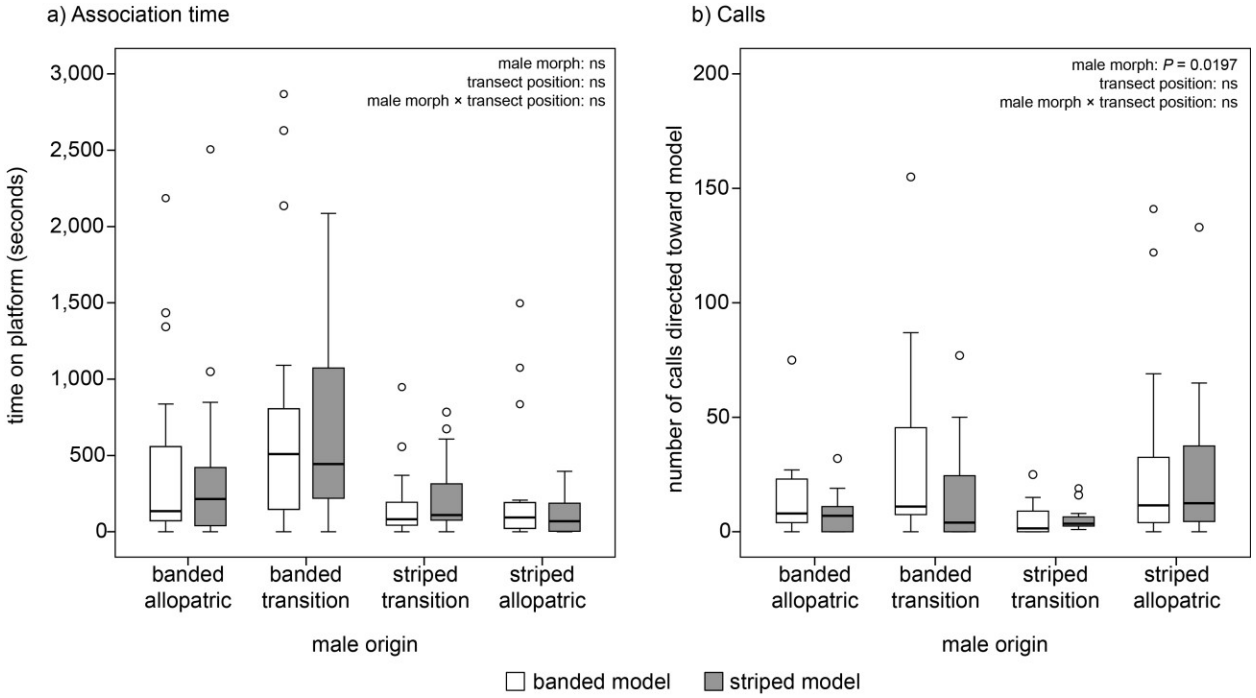


Figure 2-S5. Proportions of among-population genetic divergence explained by geographic distance and color pattern distance for each of the three mimetic transition zones in *R. imitator*. For each transition zone, the height of the bars correspond to the regression coefficient β between genetic distance/geographic distance (light grey bar) and genetic distance/color pattern distance (dark grey bar). Thus, the light grey and dark grey bars represent effects of isolation-by-distance and isolation-by-adaptation, respectively. Note that regression coefficients are given as absolute values (for the spotted-striped transition, the regression coefficient for color pattern distance is negative, although not statistically significant). Regression coefficients and associated *P*-values (shown inside bars) were calculated using multiple matrix regression (see methods).

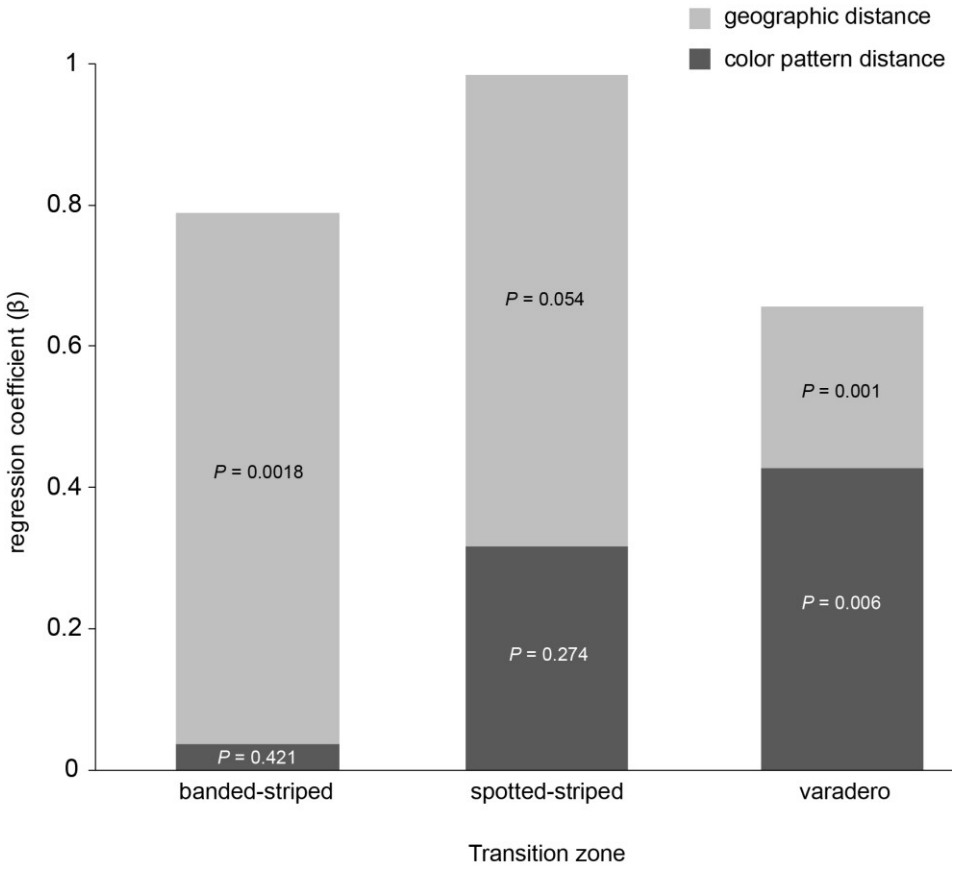


Figure 2-S6. Cline comparison between three known mimetic transition zones in *Ranitomeya imitator*. (a) Banded-striped transition zone, (b) Spotted-striped transition zone, and (c) Striped-varadero transition zone. Data from transition zone (c) adapted from Twomey et al. 2014 (data available at Dryad doi:10.5061/dryad.rd586).

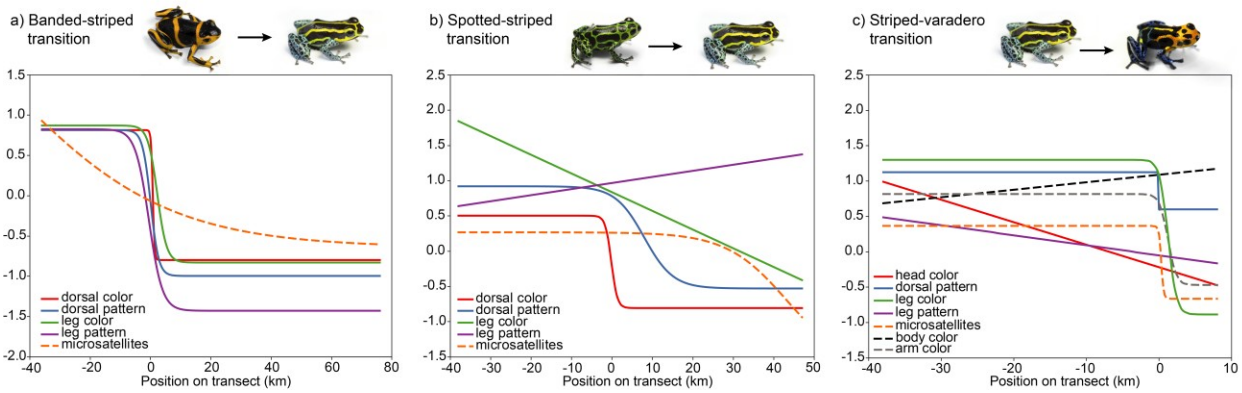


Figure 2-S7. Differences in average male mass across mimetic transition zones. (a) Banded-striped transition; populations included were Curiyacu (transition banded) and Achinamisa (transition striped), (b) Spotted-striped transition; populations included were San Jose (spotted) and Pongo de Cainarachi (transition striped), (c) Striped-varadero transition, populations included were Varadero South Bank (transition striped) and Varadero Forest site 1 (varadero). See Twomey et al. (2014) for details on collection localities for the striped-varadero transition zone.

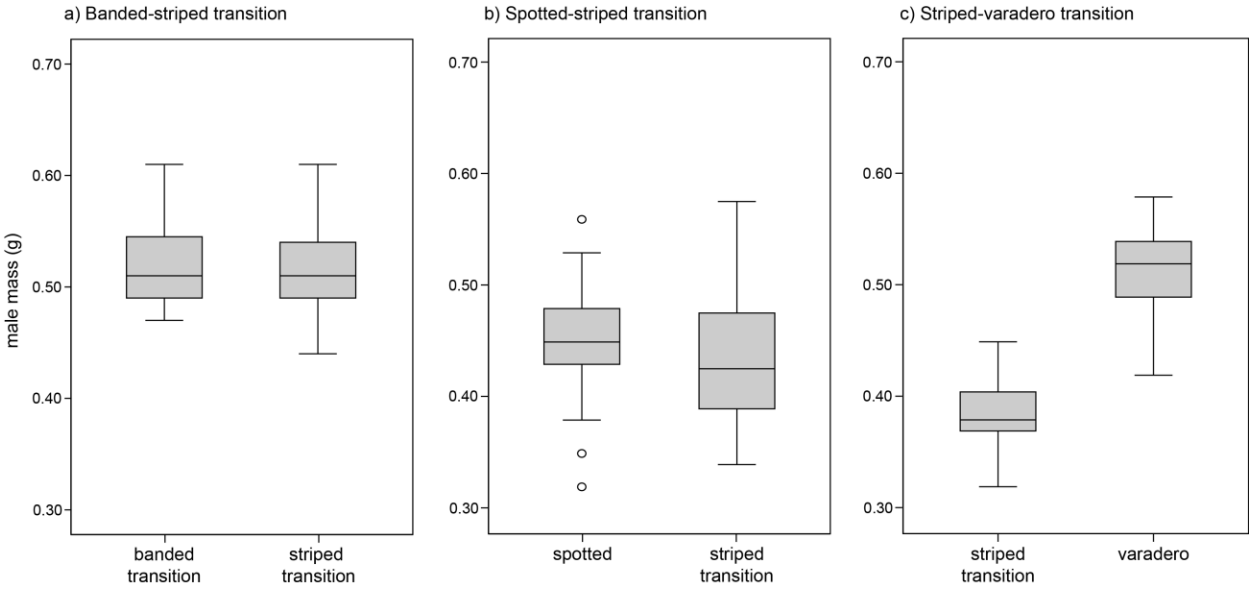


Table 2-1. Transition zone comparison.

Transition zone	Color pattern clines	Color-pattern cline width	Microsatellite structure	Landscape genetics	Assortative mating	Model species
Banded-striped	Sigmoidal shift in dorsal color, dorsal pattern, leg color, and leg pattern. Centers nearly coincident	5.35 km	Associated with allopatric populations. No structure across mimetic transition zone.	Genetic distance correlated with geographic distance, not color pattern distance	Associated with allopatric populations. Some evidence for color pattern-based assortative mating (plastic model study)	Mimetic shift coincides with parapatric distributions of *R. summersi*and *R. variabilis*
Spotted-striped	Sigmoidal shift in dorsal color and dorsal pattern. Centers offset by approximately 8 km.	2.95 km	Associated with allopatric populations. No structure across mimetic transition zone.	Genetic distance correlated (marginally) with geographic distance, not color pattern distance	Weak or absent in all tested populations	*R. variabilis* is polymorphic along an elevational gradient; this variation is tracked by *R. imitator*
Striped-varadero*	Sigmoidal shift in arm color, leg color, and body pattern. Cline centers are statistically coincident.	1.45 km	Structure coincides with mimetic transition zone.	Genetic distance correlated with both color pattern distance and geographic distance.	Preference in striped morph stronger in transition population; varadero population no preference (i.e. asymmetric preferences at transition zone)	Both model species (*R. fantastica* and *R. variabilis*) present on both sides of the mimetic transition zone.

* Data from Twomey *et al.* 2014

Table 2-S1. Sampling localities and sample sizes for banded-striped transect. For microsatellites, the number of samples taken from a previous study is shown in parentheses.

locality	number on Fig. 2-S1	position on transect (km from center)	latitude	longitude	elevation (m)	color pattern sample size	microsatellites sample size
Sauce	1	-31.71	-6.7075	-76.1967	669	40	41 (12)
Vaquero	2	-25.62	-6.6222	-76.1744	208	6	6 (0)
Chipaota	3	-14.57	-6.5851	-76.0808	209	5	5 (0)
Lower Chipaota	4	-11.1	-6.5760	-76.0506	216	—	8 (8)
Curiyacu	5	-4.77	-6.5741	-75.9920	214	4	39 (4)
Malpaso south bank	6	-2.96	-6.5784	-75.9696	204	6	16 (4)
Malpaso north bank	7	-2.47	-6.5741	-75.9676	187	—	8 (8)
Chipesa	8	1.21	-6.5499	-75.9449	204	8	38 (14)
Callanayacu north bank	9	4.01	-6.5630	-75.9240	372	6	32 (0)
Callanayacu south bank	10	4.04	-6.5787	-75.9216	264	—	19 (19)
Ricardo Palma south bank	11	4.32	-6.5364	-75.9194	193	7	7 (0)
Ricardo Palma north bank	12	4.77	-6.5279	-75.9212	184	—	9 (9)
Achinamisa	13	10.26	-6.4762	-75.9087	174	11	8 (2)
Aguas Termales	14	12.67	-6.4592	-75.8946	183	—	7 (7)
Lower Achinamisa	15	13.78	-6.4476	-75.8949	189	2	6 (4)
Micaela Bastidas	16	74.04	-5.9554	-76.2424	177	30	36 (36)

Table 2-S2. Sampling localities and sample sizes for spotted-striped transect. For microsatellites, the number of samples taken from a previous study is shown in parentheses.

locality	letter on Fig. 2-S1	position on transect (km from center)	latitude	longitude	elevation (m)	color pattern sample size	microsatellites sample size
Pinto Recodo	A	-37.71	-6.3301	-76.6331	349	5	5 (0)
Bocatoma	B	-11.85	-6.4547	-76.3489	498	7	17 (10)
San Jose	C	-7.85	-6.4328	-76.2829	495	40	35 (12)
Upper Pongo 2	D	-1.96	-6.3779	-76.2845	477	3	3 (0)
Sisayacu	E	-0.91	-6.3705	-76.2908	377	6	6 (0)
Upper Pongo 1	F	1.44	-6.3520	-76.3009	327	17	17 (0)
Pongo	G	10.1	-6.2927	-76.2354	209	32	32 (11)
Micaela Bastidas	H	45.42	-5.9554	-76.2424	177	30	36 (36)

Table 2-S3. Parameters used in the kernel discriminant analysis.

Transition zone		*Regularization* λ	*Kernel width* γ
Banded-striped	dorsal color	0.0004	5.7294
	leg color	0.0003	2.7526
	dorsal pattern	0.0004	8.8596
	leg pattern	0.0011	7.0909
Spotted-striped	dorsal color	0.00481	3.41665
	leg color	0.0001	3.67380
	dorsal pattern	0.0001	12.15990
	leg pattern	0.00020	9.22890

Table 2-S4. Banded-striped transect global regression results for common center and width parameters. Models were compared using AICc. The best supported model is shown in bold.

model		dorsal color	dorsal pattern	leg color	leg pattern	microsatellites	total RSS	parameters	AICc	ΔAICc
	n	124	124	124	124	285				
a) No constraints	**center**	**0.72**	**0.02**	**2.23**	**-0.56**	**-47.68**				
	width	**-1.20**	**-5.37**	**-6.25**	**-8.97**	**-106.97**				
	RSS	**45.26**	**10.24**	**13.16**	**21.20**	**30.47**	**120.33**	**20**	**-1419.62**	**0.00**
b) Shared center	center	0.46	0.46	0.46	0.46	0.46				
	width	-1.84	-4.54	-9.41	-8.21	-60.78				
	RSS	45.26	10.36	15.26	21.67	34.88	127.43	16	-1383.27	36.34
c) Shared width	center	-0.22	0.02	2.28	-0.05	-16.04				
	width	-5.41	-5.41	-5.41	-5.41	-5.41				
	RSS	46.36	10.24	13.22	21.93	40.84	132.59	16	-1352.23	67.39
d) Shared center and width	center	0.18	0.18	0.18	0.18	0.18				
	width	-8.63	-8.63	-8.63	-8.63	-8.63				
	RSS	48.78	11.06	15.91	21.45	50.48	147.68	12	-1276.38	143.23
e) Microsatellites unconstrained; color pattern shared center	center	0.99	0.99	0.99	0.99	-47.68				
	width	-0.54	-0.95	-8.43	-7.88	-106.97				
	RSS	45.26	10.53	14.32	22.34	30.47	122.92	17	-1409.32	10.30
f) Microsatellites unconstrained; color pattern shared width	center	-0.22	0.04	2.28	-0.03	-47.68				
	width	-5.35	-5.35	-5.35	-5.35	-106.97				
	RSS	46.32	10.24	13.24	21.96	30.47	122.22	17	-1413.76	5.86
g) Microsatellites unconstrained; color pattern shared center and width	center	0.50	0.50	0.50	0.50	-47.68				
	width	-5.93	-5.93	-5.93	-5.93	-106.97				
	RSS	47.14	10.50	16.09	21.96	30.47	126.15	14	-1395.29	24.33

Table 2-S5. Spotted-striped transect global regression results for common center and width parameters. Models were compared using AICc. The best supported model is shown in bold.

model		dorsal color	dorsal pattern	microsatellites	total RSS	parameters	AICc	ΔAICc
	n	140	148	151				
a) No constraints	**center**	**-0.17**	**7.96**	**36.92**				
	width	**-2.94**	**-14.31**	**-6.02**				
	RSS	**34.09**	**52.62**	**13.00**	**99.71**	**12**	**-625.98**	**0.00**
b) Shared center	center	8.01	8.01	8.01				
	width	0.00	-14.26	-269.73				
	RSS	52.34	52.62	18.89	123.84	10	-535.03	90.95
c) Shared width	center	-0.17	2.78	0.20				
	width	-4.03	-4.03	-4.03				
	RSS	34.26	56.63	34.25	125.14	10	-530.48	95.51
d) Shared center and width	center	1.08	1.08	1.08				
	width	-4.84	-4.84	-4.84				
	RSS	36.93	59.42	33.28	129.64	8	-519.14	106.84
e) Microsatellites unconstrained; color pattern shared center	center	1.44	1.44	36.92				
	width	0.00	-24.84	-6.02				
	RSS	34.66	57.16	13.00	104.82	11	-606.15	19.83
f) Microsatellites unconstrained; color pattern shared width	center	-0.17	2.41	36.92				
	width	-2.95	-2.95	-6.02				
	RSS	34.09	56.64	13.00	103.73	11	-610.72	15.27
g) Microsatellites unconstrained; color pattern shared center and width	center	0.99	0.99	36.92				
	width	-4.43	-4.43	-6.02				
	RSS	36.54	59.78	13.00	109.33	10	-589.77	36.21

Table 2-S6. Banded-striped transect model fit results. In all cases, the best supported model is shown in bold. For cases where the sigmoid model was best supported, point estimates and 95% Monte Carlo confidence intervals on center and width parameters are given.

variable	model	AICc	ΔA ICc	Akaike weight	center	center 95% CI	width	width 95% CI
dorsal color	flat	-102.5	117.7	0.000	–	–	–	–
	linear	-158.6	61.6	0.000	–	–	–	–
	sigmoid	**-220.1**	**0.0**	**1.000**	**0.74**	**-1.08 – 1.18**	**-1.16**	**-8.08 – -0.02**
dorsal pattern	flat	-23.7	302.0	0.000	–	–	–	–
	linear	-139.7	186.0	0.000	–	–	–	–
	sigmoid	**-325.6**	**0.0**	**1.000**	**0.02**	**-0.64 – 1.14**	**-5.37**	**-7.51 – -0.36**
leg color	flat	-44.2	270.8	0.000	–	–	–	–
	linear	-175.8	139.2	0.000	–	–	–	–
	sigmoid	**-315.0**	**0.0**	**1.000**	**2.23**	**1.52 – 2.84**	**-6.25**	**-9.18 – -4.26**
leg pattern	flat	35.4	211.7	0.000	–	–	–	–
	linear	-71.5	104.8	0.000	–	–	–	–
	sigmoid	**-176.3**	**0.0**	**1.000**	**-0.56**	**-1.69 – 1.14**	**-8.97**	**-13.68 – -0.34**
microsatellites FCA axis 1	flat	-519.5	109.5	0.000	–	–	–	–
	linear	-556.4	72.6	0.000	–	–	–	–
	sigmoid	**-629.1**	**0.0**	**1.000**	**-47.68**	**-186.17 – -7.98**	**-106.97**	**-164.18 – -47.62**

Table 2-S7. Spotted-striped transect model fit results. In all cases, the best supported model is shown in bold. For cases where the sigmoid model was best supported, point estimates and 95% Monte Carlo confidence intervals on center and width parameters are given.

variable	model	AICc	ΔA ICc	Akaike weight	center	center 95% CI	width	width 95% CI
dorsal color	flat	-62.8	126.7	0.000	–	–	–	–
	linear	-119.3	70.1	0.000	–	–	–	–
	sigmoid	**-189.5**	**0.0**	**1.000**	**-0.17**	**-1.20 – 1.30**	**-2.94**	**-7.00 – -0.07**
dorsal pattern	flat	-51.0	91.8	0.000	–	–	–	–
	linear	-120.4	22.3	0.000	–	–	–	–
	sigmoid	**-142.8**	**0.0**	**1.000**	**7.96**	**1.49 – 10.09**	**-14.31**	**-30.38 – -0.11**
leg color	flat	124.2	19.1	0.000	–	–	–	–
	linear	**105.0**	**0.0**	**0.805**	–	–	–	–
	sigmoid	107.9	2.8	0.195	–	–	–	–
leg pattern	flat	25.6	2.9	0.111	–	–	–	–
	linear	**22.7**	**0.0**	**0.480**	–	–	–	–
	sigmoid	23.0	0.3	0.409	–	–	–	–
microsatellites FCA axis 1	flat	-230.1	132.0	0.000	–	–	–	–
	linear	-284.3	77.7	0.000	–	–	–	–
	sigmoid	**-362.0**	**0.0**	**1.000**	**43.26**	**1.53 – 92.37**	**-29.93**	**-65.5 – -0.01**

Chapter 3: Intraspecific call variation in the mimic poison frog *Ranitomeya imitator*

Abstract

Calls are important premating isolating barriers in frogs; thus, studying intraspecific variation in calls may allow us to understand whether call divergence is important during the early stages of speciation. *Ranitomeya imitator* is a species of poison frog (family Dendrobatidae) that has undergone extensive color-pattern diversification through a Müllerian mimetic radiation, establishing four distinct morphs in north-central Peru (striped, banded, varadero, and spotted). Previous studies have determined partial reproductive isolation between certain color morphs, although the specific mechanisms responsible for this isolation are poorly understood. In this study, we conduct a species-wide analysis of variation in advertisement calls to investigate whether distinct mimetic morphs also show advertisement call differences. We also conduct phonotaxis experiments to investigate whether call differences among mimetic morphs cause reproductive isolation via female choice. We find that different color morphs generally show weak or no differences in advertisement calls, with two exceptions. First, there is a shift in call pulse rate between the striped and banded morphs, which occurs at the same geographic location as the color-pattern shift. Second, between the striped and varadero morphs, there is a shift in both note length and dominant frequency, occurring at the same geographic location as the color-pattern shift. Despite the observed call differences among certain populations, our phonotaxis experiments show that females from the striped and banded populations show no preference with respect to pulse rate or dominant frequency, and females from the striped and varadero

populations show no preference with respect to note length. Thus, while certain mimetic morphs show call differences, calls differences alone are unlikely to cause reproductive isolation.

Introduction

In frogs, advertisement calls play an important role in reproduction, carrying information about mate location, mate quality, and species identity (Wells 2007). Because advertisement calls are often species-specific signals, call differences between populations or species may prevent recognition of potential mates, and, therefore, may result in reproductive isolation (Backwell & Jennions 1993; Gerhardt 1974; Lemmon 2009; Oldham & Gerhardt 1975; Ryan & Rand 1993; Telford & Passmore 1981). Call differences may also be relevant to speciation, as there are several examples where speciation in frogs appears to have been driven by call divergence (Boul *et al.* 2007; Funk *et al.* 2009; Guerra & Ron 2008; Platz 1989). By examining call variation among populations of a single species, we can understand how potential reproductive barriers evolve during early stages of population divergence and speciation. However, while intraspecific call variation in frogs is relatively common (Wilczynski & Ryan 1999), it is often unknown whether such variation results in reproductive isolation.

Ranitomeya imitator is a species of poison frog that has undergone a Müllerian mimetic radiation, that is, different populations have evolved distinct color-patterns in order to mimic several different congeneric species (Symula *et al.* 2001; Twomey *et al.* 2013; Yeager 2009). There are four mimetic morphs of *R. imitator*, all of which occur in different geographic regions in north-central Peru. Where different morphs come into contact, "transition zones" are formed, which are characterized by the presence of phenotypic intermediates (Twomey *et al.* submitted,

2014). This system is ideal for studying how divergence in an ecologically relevant trait (mimetic color pattern) can drive speciation in that there are three different mimetic transition zones, thus, we can study each transition zone separately and identify any reproductive barriers present, allowing us to investigate which barriers arise during the process of ecological speciation. In a previous study, Twomey *et al.* (2014) found that mimetic divergence may be driving incipient speciation between two mimetic morphs, and furthermore that these two morphs appear to show significant call differences despite occurring less than 4 km apart. However, whether or not these call differences cause reproductive isolation between the two morphs remains unknown. Furthermore, call variation across the other two mimetic transition zones has not yet been investigated.

While mimetic divergence itself should not drive call differentiation (most divergence should occur only in traits related to mimicry), there are a few processes by which call differentiation between mimetic morphs might arise. First, if divergent selection is strong enough to reduce gene flow between morphs, divergence in traits unrelated to mimicry may occur via genetic drift (Feder & Nosil 2010). In this view, divergence in most traits occurs as a consequence of reduced gene flow between morphs. A similar pattern could arise via divergence in allopatry followed by secondary contact. Second, call divergence could occur if traits subject to divergent ecological selection have pleiotropic effects on calls, for example body size, which can influence call frequency and pulse rate (Duellman & Trueb 1986; Wilczynski & Ryan 1999). Third, because aposematic traits are subject to positive frequency dependent selection (Endler & Greenwood 1988; Greenwood *et al.* 1989), different morphs should become fixed in different geographic areas, creating parapatry between morphs. Thus, as different morphs will often occupy different

geographic areas, they may also occupy distinct habitats, which could drive call divergence directly (e.g., if different calls were better suited to different local conditions), or indirectly (e.g., if different habitats caused divergence in body size, indirectly driving call divergence). Finally, it is possible that reproductive character displacement could cause call divergence in areas where different lineages come into contact (Höbel *et al.* 2003). As cross-morph matings may result in the production of non-mimetic offspring, individuals that mate with their own mimetic morph could have higher fitness than individuals that mate at random (Noor 1999). Thus, call differences between morphs could arise as an adaptation to avoid cross-morph matings. A key prediction of this hypothesis is that call differences between morphs should be greater between sympatric (or parapatric) populations versus allopatric populations.

In this study, we investigate call variation in *R. imitator* to address whether shifts in call parameters (if any) coincide with shifts in mimetic color-pattern. We also conduct phonotaxis experiments to determine whether any of the observed call differences result in reproductive isolation between morphs. Finally, we investigate the influence of body-size variation on call variation in order to understand potential mechanisms explaining call variation across the striped-varadero transition zone.

Methods

Call recording and analysis

We recorded a total of 75 individual *R. imitator* from 13 localities in the departments of San Martin and Loreto, Peru. We included recordings from an additional 58 *R. imitator* from 7 localities from a previously published paper (Twomey *et al.* 2014) corresponding to the striped-

varadero transect. Our sampling localities span the three mimetic transition zones in *R. imitator* (banded-striped, 9 localities, 61 recordings; spotted-striped, 6 localities, 36 recordings, striped-varadero, 7 localities, 58 recordings), allowing us to investigate whether aspects of the advertisement call shift across mimetic transition zones. See Table 3-1 for details on sampling localities and sample sizes. Transect position was calculated as the distance from a fixed point representing the estimated transition zone center. See Twomey *et al.* (2014) and Twomey *et al.* (submitted) for details on transect position calculations.

Call recordings were obtained by searching for calling *R. imitator* males in the field. *Ranitomeya imitator* has two call types: a longer advertisement call (Fig. 3-1), and a shorter call used for close-range courtship. For this study, we focused on advertisement calls, which can be recognized by having longer note lengths and inter-note intervals than courtship calls. Calls were recorded with a Marantz PMD660 solid state digital recorder with a Sennheiser ME66-K6 microphone. We extracted three bioacoustic variables commonly used in dendrobatid studies (Brown *et al.* 2011): dominant frequency (frequency at which peak amplitude is reached), note length (length in seconds of a note measured from the start of the first pulse to the end of the last pulse, where a note is described as a discrete bundle of pulses), and pulse rate (pulses per second, calculated by dividing the number of pulses contained within a note by the note length) (Fig. 3-1). Call measurements were taken in Raven Pro version 1.3 (Charif *et al.* 2008). For each male, a recording generally consisted of several notes. Measurements were always taken on at least 3 notes and averaged for each male. To account for any influence of temperature on calls, we took temperature measurements alongside each call recording in the microhabitat of the calling male. We then standardized each of the three call variables (dominant frequency, note length, and pulse

rate) by calculating regression residuals. Further statistical analyses were done on these temperature-standardized variables rather than on raw measurements. Residuals were calculated in SPSS version 20 on the entire dataset simultaneously (i.e., not separated per transect). Comparisons of call parameters between the populations used in the phonotaxis experiments (Figures 3-3 and 3-4) were done with a one-way ANOVA in SPSS version 20.

To describe spatial variation (i.e., variation along each transect) in call parameters, we fit three candidate models to the data and evaluated each using ΔAIC_c. The three models were a flat model, a linear model, and a sigmoid model. A flat model describes a scenario where a call parameter is constant across the transect, and consists of a single parameter, the population mean, which is defined as the grand mean of all individuals. A linear model describes a constant, smooth shift across the transect, and has two parameters, slope and y-intercept. The linear model was fit with linear regression. Finally, a sigmoid model describes a scenario where a call parameter abruptly shifts along the transect. This is expected, for example, if two mimetic morphs sharing a transition zone have distinct calls. However, it is possible that an abrupt shift could occur anywhere along the transect, and not just at a mimetic transition zone. Thus, we imposed a constraint on the position of the cline such that the center parameter was constrained to be equal to the center of the color-pattern clines found in previous studies (Twomey *et al.* submitted, Twomey *et. al.* 2014). For the banded-striped transect the center was constrained to 0.99 km, for the spotted-striped transect the center was constrained to 1.44 km, and for the striped-varadero transect the center was constrained to 0.52 km. Thus, by constraining call-centers to color-pattern centers, we are testing the hypothesis that there is an abrupt shift in call

parameters that occurs at the same geographical location as the shift in color-pattern. For the sigmoidal model, we performed nonlinear regression using a 4-parameter tanh function

$$y = \frac{1 + \tanh\left(\frac{2x - c}{w}\right)}{2\left(\frac{1}{y_{\max} - y_{\min}}\right)} + y_{\min}$$

where *c* is the center of the cline, *w* is the width of the cline, and *ymax* and *ymin* are the maximum and minimum trait values. Parameter searches were done using the solver function in EXCEL using a least-squares optimality criterion. Solver was run using the GRG nonlinear algorithm with the following settings: convergence = 0.0001; central derivatives; multistart on; population size = 100. To evaluate which of the three models (1-parameter flat, 2-parameter linear, or 4-parameter sigmoid) was a better fit to the data, we calculated ΔAIC_c and Akaike weights (w_i) for each model (Burnham & Anderson 2002) using the residual sum of squares divided by the sample size as the likelihood criterion.

To describe relationships between call parameters and male body mass, for each population where we had male body mass data (mass data from Twomey *et al.* submitted, Twomey *et al.* 2014), we took the population average male body mass, and regressed this on the population average of each call parameter. We used population-level data (i.e., averaging across individuals) as we did not have corresponding mass data for each call recording. We also regressed average male mass on elevation to test for a relationship between mass and elevation. Linear regressions were calculated in SPSS version 20.

Phonotaxis experiments

We used phonotaxis trials to determine whether call differences among populations of *R. imitator* might cause reproductive isolation. We conducted two different phonotaxis experiments: one with striped and varadero frogs, where we manipulated a single call variable (note length), and another experiment with banded and striped frogs, where we manipulated two call variables independently (dominant frequency and pulse rate). For the former experiment, we chose to test a striped-transition and varadero population (Varadero South Bank and Varadero Forest 1, respectively, see Table 3-1). These two populations have significant call differences with respect to note length and dominant frequency (Fig. 3-2). However, due to time limitations, we only conducted phonotaxis trials manipulating note length. For the latter experiment, we chose to test allopatric banded and allopatric striped populations (Sauce and Micaela Bastidas, respectively, see Table 3-1), as these two populations have significant call differences with respect to both pulse rate and dominant frequency (Fig. 3-4).

We collected female *R. imitator* in the field and kept them in captivity in Tarapoto, Peru. Females were identified by their slightly larger size (Brown *et al.* 2008a), as well as on the basis of observations made while collecting (e.g. engaged in courtship with a male). Frogs were housed individually in glass terraria (dimensions 50 cm x 30 cm x 30 cm) when not being used in phonotaxis trials. Terraria had roughly two inches of washed gravel and leaf litter as a substrate, with two bromeliads in each terrarium to provide hiding places. Water and food (wild fruit flies) were both constantly available.

To test for call preferences of females, we conducted phonotaxis trials in which we introduced a female into an experimental arena, played experimental calls on either end of the

arena, and measured the amount of time the female spent within a 20 cm interaction zone with each speaker. The arena was a large plastic box, 85 cm long x 50 cm wide x 40 cm tall. On opposing ends of the box, holes were cut in the plastic (which were then covered by screen) to facilitate sound transmission. Speakers for call playback were placed behind these holes, outside the arena. Around each speaker-hole, we drew a 20-cm radius circle representing an interaction zone threshold. In the center of the arena, there was a neutral-zone, in which we placed leaf litter and a 10 cm length of PVC tube so that the female could remain hidden. In this way, a female that was uninterested in the calls could remain hidden outside each interaction zone. A trial was started by placing the female inside the PVC tube section in the center of the arena, covering the entire arena with a piece of glass (to prevent escape), beginning the call playback (details on call manipulations below), and filming the trial for one hour. We randomized the call playback with respect to both arena side and speaker identity in order to account for side preferences and inherent differences between speakers. For example, in an experiment to test for preference for long versus short note length, we generated an experimental playback call where a long call was played on one channel while a short call was played on the alternate channel, thus, during playback, the long and short calls would be emitted from different speakers. However, if one speaker systematically differed from the other, this would confound call type with speaker identity (e.g., if the short call was on channel 1, then speaker 1 would always play the short call). To address this, we made a second version of the playback track where the channels were swapped. Prior to each trial, we used a coin flip to determine which experimental track would be played (thus randomizing call type with respect to speaker identity), and a second coin flip to

determine which side of the arena each speaker would be placed (thus randomizing speaker identity with respect to arena side).

For the striped-transition/varadero phonotaxis experiment where we tested note length preferences, we generated experimental calls by editing a recording of a *R. imitator* advertisement call from a "neutral" site (San Jose, Cainarachi valley, Peru). We selected a single note that had an original length of 0.692 s. To create the short call, two pulses were deleted from the middle of the note, so that the new note length was 0.63 s, corresponding to the average note length of the striped-transition population. The long call was created by duplicating 10 pulses from the middle of the call, to reach a final note length of 0.99 s, which corresponds to the average note length of the varadero population. This ensures that both calls were manipulated, accounting for any effect of the manipulation process on preference, and that the manipulations were both done in the middle of the call. Notes were then copied in their respective channels such that one note was played every 10 seconds. Notes in each channel were staggered by 5 seconds to avoid interference. This call series was then looped to play continuously for 1 hour.

For the banded/striped phonotaxis experiment, we had two call manipulations: one where we manipulated dominant frequency, and one where we manipulated pulse rate. For the pulse rate manipulation, to make a low-pulse-rate call, we isolated a single pulse, added 0.0193 seconds of silence to the end, and copied the pulse until a call length of 0.65 seconds was reached. For the high-pulse-rate call, we added only 0.0143 seconds of silence to the end of the pulse. As the intervening silence between pulses was lower, this resulted in a higher pulse rate. The two experimental calls thus had an identical total note length (0.65 seconds), but differed in pulse rate, with the low-pulse call with a rate of 32 pulses per second, and the high-pulse call with a

rate of 38 pulses per second. These pulse rates correspond to the pulse rate averages in the Sauce population (banded allopatric) and Micaela Bastidas population (striped allopatric), respectively. For the dominant frequency manipulation, we made a low-frequency call by isolating a single pulse of 5550 Hz, shifting the pitch down to 5124 Hz, and copying the pulse 23 times. We did the same thing to make the high-frequency call except that we shifted the pitch of the initial pulse up to 5700 Hz. The two experimental calls thus had identical length and pulse rate, but with dominant frequencies of 5124 Hz, corresponding to the Sauce population, or 5700 Hz, corresponding to the Micaela Bastidas population. For both the dominant frequency and pulse-rate experimental calls, notes were copied and staggered in different channels as described above. All call manipulations were done in Audacity 2.0 (audacity.sourceforge.net).

To analyze phonotaxis data, we used generalized linear mixed models using the glmmADMB package (Skaug *et al.* 2011) with an underlying beta-binomial error distribution to test whether the time females spent in the interaction zone of each call type was influenced by female population origin. We used "female origin" as a fixed effect and individual "female identity" as a random effect. The significance of female origin was determined using χ^2 tests comparing models where female origin was either included or excluded as a parameter in the model.

Results

For the banded-striped transect, we found that note length showed no consistent pattern of variation along the transect (Fig. 3-2a), whereas there was a gradual, linear increase in dominant frequency moving from the banded toward the striped morph (Fig. 3-2a). Pulse rate variation was best explained by a sigmoidal model, even with the center of the cline constrained *a priori* to

0.99 km. Thus, there was evidence of a stepped increase in pulse rate at the same geographic location as color-pattern shift (Fig. 3-2a, Table 3-2), although the width of the cline was quite wide (11.63 km, Table 3-2). For the spotted-striped transition, we found that both pulse rate and note length showed no consistent pattern of variation along the transect (Fig. 3-2b). There was a weak, linear increase in dominant frequency moving from the spotted toward the striped morph (Fig. 3-2b), although the relationship between transect position and dominant frequency was weak (R^2 = 0.09). Finally, for the striped-varadero morph, variation in both note length and dominant frequency was best explained by a sigmoidal model (Fig. 3-2c), even when the center of the cline was constrained *a priori* to 0.52 km. Thus, moving along the transect in the direction of the varadero morph, there was evidence of a stepped increase in note length and decrease in dominant frequency at the same geographic location as the color-pattern shift. The cline widths were relatively narrow: 4.89 km for note length and 3.52 km for dominant frequency (Table 3-2). Variation in pulse rate was best described by a linear model (Fig. 3-2c), showing a slight decrease in pulse rate along the transect (i.e., moving from the striped to the varadero morph).

We found that average male mass was not significantly correlated to elevation (R^2= 0.106, P = 0.218; Fig. 3-6a). This lack of a mass-elevation relationship is exemplified by lowland populations (elevation < 200 m), showing a range of mass variation exceeding variation shown across all other populations. Neither note length nor pulse rate were significantly correlated to average male mass (note length, R^2= 0.044, P = 0.438, Fig. 3-6b; pulse rate, R^2= 0.028, P = 0.538, Fig. 3-6c). However, dominant frequency was negatively correlated to average male mass (R^2= 0.311, P = 0.025; Fig. 3-6d).

When comparing calls between populations used in the phonotaxis experiments, we found that the striped transition and varadero populations are significantly different with respect to note length ($F_{1,38} = 8.91$, $P = 0.005$) and dominant frequency ($F_{1,37} = 17.55$, $P = 0.00017$), but not pulse rate ($F_{1,35} = 0.255$, $P = 0.617$) (Fig. 3-3). Comparing the banded allopatric and striped allopatric populations (Fig. 3-3), we found significant differences in pulse rate ($F_{1,20} = 11.67$, $P = 0.003$) and dominant frequency ($F_{1,20} = 10.16$, $P = 0.005$), but not note length ($F_{1,20} = 0.135$, $P = 0.718$).

For the striped transition/varadero phonotaxis experiment where we tested for note length preferences, we had a total of 24 successful trials across the two sampled populations (striped transition, n=12; varadero, n=12). In total, we recorded 87.8 minutes of interaction time (i.e., time within interaction zone), for an average of 3.7 minutes of interaction time per trial. Comparing the preferences in the two populations (Fig. 3-5a), we found no significant effect of female origin on interaction time ($\chi^2 = 2.27$, df = 1, $P = 0.1319$), indicating that call preferences were not significantly different between these two populations. Furthermore, the regression coefficients for each level of female origin were not significantly different from zero, indicating that neither population showed a significant preference for either call type. Thus, based on these results, neither the striped transition females nor varadero females appear to show any preference for long versus short advertisement calls.

For the banded allopatric/striped allopatric phonotaxis experiment where we tested for dominant frequency preferences, we had a total of 17 successful trials across the two sampled populations (banded allopatric, n=10, striped allopatric, n=7). In total, we recorded 57.5 minutes of interaction time, for an average of 3.4 minutes of interaction time per trial. Comparing the

preferences in the two populations (Fig. 3-5b), we found no significant effect of female origin on interaction time ($\chi^2 = 1.35$, df = 1, $P = 0.246$), indicating that call preferences with respect to dominant frequency were not significantly different between these two populations. Regression coefficients for levels of female origin were not significantly different from zero, indicating that neither population showed a significant preference for high or low frequency calls. For the pulse-rate experiment, we had only 12 successful trials across the two populations (banded allopatric, n=7, striped allopatric, n=5). We recorded a total of 37.9 minutes of interaction time, for an average of 3.2 minutes of interaction time per trial. Comparing preferences in the two populations (Fig. 3-5c), we found no significant effect of female origin on interaction time ($\chi^2 = 0.054$, df = 1, $P = 0.816$), indicating that call preferences with respect to pulse rate were not significantly different between populations. Regression coefficients for levels of female origin were not significantly different from zero, indicating that neither population showed a significant preference for high or low pulse rate calls.

Discussion

We found intraspecific variation in a number of call traits among populations of *R. imitator*. The most notable variation was in dominant frequency, which varied significantly across all three transects (Fig. 3-2). However, there was also evidence for intraspecific variation in both pulse rate and note length (Fig. 3-2). Our sampling revealed clinal variation in call traits across two mimetic transition zones. Clinal variation in advertisement calls has been found in a number of frogs (Bernal *et al.* 2005; Hasegawa *et al.* 1999; Narins & Smith 1986; Ryan *et al.* 1996), which is often associated with an underlying environmental gradient such as elevation or latitude. In *R.*

imitator, an underlying elevation gradient could explain some of the variation in dominant frequency across the banded-striped and spotted-striped transition zones. In these transition zones, the striped morph occurring in the lowlands generally has a higher dominant frequency than either the banded or spotted morphs, both of which occur in highland habitats. Thus, along the elevation gradient, there appears to be a general shift towards lower dominant frequency, even when controlling for recording temperature. One possible explanation for this elevation-associated shift would be if there was a similar shift in body mass, for example, if frogs in highland habitats were larger and thus produced lower-frequency calls. However, we found no significant correlation between body size and elevation (Fig. 3-7), indicating that the shift toward a lower frequency in the highland populations is likely not driven by an elevation effect on body size. In addition, we found that clinal variation in advertisement calls tracks shifts in mimetic color-pattern in two of the three transition zones. In the banded-striped transition zone, we found a stepped shift in pulse rate corresponding to the shift in mimetic color-pattern. The shift occurs over a relatively wide area (11.63 km), but involves a significant increase in pulse rate seen in the striped morph. Thus, these two mimetic morphs appear to be somewhat distinct with respect to pulse rate. However, our phonotaxis experiments show that differences in pulse rate and dominant frequency across this transect are unlikely to generate any reproductive isolation between mimetic morphs. For this experiment, we selected a banded and striped population, each from the ends of the sampling transect. Thus, these populations are well-differentiated with respect to dominant frequency and pulse rate (Fig. 3-4). However, females from both populations showed no preferences for high- versus low-frequency calls, nor for high- versus low-pulse rate calls (Fig. 3-5). Therefore, call variation among these populations does not appear to mediate

assortative mating, which has been observed in the banded-allopatric population studied here (Twomey *et al.* submitted).

In the striped-varadero transition zone, we found a shift in both note length and dominant frequency. Both of these shifts occur in the same location as the mimetic transition zone, and are relatively narrow (< 5 km). Thus, at the striped-varadero transition zone, the two mimetic morphs are well-differentiated in two aspects of the advertisement calls. In addition to a shift in call parameters across this transition zone, there is also a strong shift in body size, with striped males being on average much smaller than varadero males (Twomey *et al.* 2014). Thus, the shift in dominant frequency is likely due to this shift in mass, as there is a well-established negative relationship between body size and call frequency (Duellman & Trueb 1986). This relationship was also confirmed for *R. imitator* in the current study (Fig. 3-6d). However, we found no such relationship between male mass and note length (Fig. 3-6b), which suggests that the shift in note length cannot be explained by differences in body size between the two morphs. In a previous study, Twomey *et al.* (2014) found that the striped and varadero morphs are somewhat genetically isolated, and that there is evidence of morph-based assortative mating present at the transition zone. Thus, one possibility, related to the observed call differences, is that this reproductive isolation is mediated by differences in advertisement calls. However, our phonotaxis trials do not support this hypothesis, as females from both populations were equally likely to approach long-note calls as short-note calls (Fig. 3-5a). Whether differences in dominant frequency generate any reproductive isolation between these populations remains untested.

Advertisement calls in the *vanzolinii* clade of *Ranitomeya* (a species group containing *R. cyanovittata, R. flavovittata, R. imitator, R. sirensis, R. vanzolinii,* and *R. yavaricola*) are remarkably conserved (Brown *et al.* 2011): all species have a short, musical trill. For example, the breadth of variation in calls within *R. imitator* meets or exceeds the variation seen across the entire species group. Across all recorded *R. imitator* individuals, note length ranged from 0.27–1.38 seconds, pulse rate ranged from 29–47 calls per second, and dominant frequency ranged from 4703–5977 Hz. Looking at the calls of other species in the clade, note length ranges from 0.57 seconds (in *R. vanzolinii*) to 2.2 seconds (in *R. sirensis*); pulse rate ranges from 24 pulses per second (in *R. sirensis*) to 32 pulses per second (*R. yavaricola*), and dominant frequency ranges from 5010 Hz (in *R. sirensis*) to 6000 Hz (in *R. yavaricola*) (Brown *et al.* 2011). Thus, in many cases, the variation in call parameters across the entire clade is less than the variation seen among individual *R. imitator*. The exception is that *R. sirensis* can produce a longer call (up to 2.2 seconds in length), and a lower pulse rate (down to 24 pulses per second), than has been observed in *R. imitator.* Overall, this suggests that speciation in the *vanzolinii* clade has taken place without substantial interspecific call divergence. Color-pattern, however, varies drastically among many of the species in this clade. Color-pattern divergence is frequently implicated as a driver of population divergence in dendrobatids (Rudh *et al.* 2011; Twomey *et al.* 2014; Wang & Summers 2010), but research in this group has generally focused on how color pattern divergence among populations of a single species can lead to within-species reproductive isolation. Thus far, no studies have attempted to address whether a known radiation of species (i.e., a scenario where speciation is already complete) was driven by color-pattern divergence. As most species in the *vanzolinii* group show similar advertisement calls, other factors, such as

color-pattern divergence or biogeographic history, may have been important during the diversification of this group.

Acknowledgments

We thank Justin Touchon for help with generalized linear mixed models. We thank César Lopez, Manuel Guerra-Panaifo, Ronald Mori Pezo, Lisa Schulte, Adam Stuckert, and James Tumulty for help in the field. This research was funded by a NSF DDIG (1210313) grant awarded to ET and KS, a National Geographic Society grant awarded to KS (8751-10) and the NCCB scholarship (2012) at East Carolina University awarded to ET. Research permits were obtained from the Ministry of Natural Resources (DGFFS) in Lima, Peru (Authorizations No. 050-2006-INRENA-IFFS-DCB, No. 067-2007-INRENA-IFFS-DCB, No. 005-2008-INRENA-IFFS-DCB). All research was conducted following an animal use protocol approved by the Animal Care and Use Committee of East Carolina University.

Figures and tables

Figure 3-1. Advertisement call of *R. imitator* and depiction of parameters extracted for bioacoustic analysis. Call recorded at 22°C.

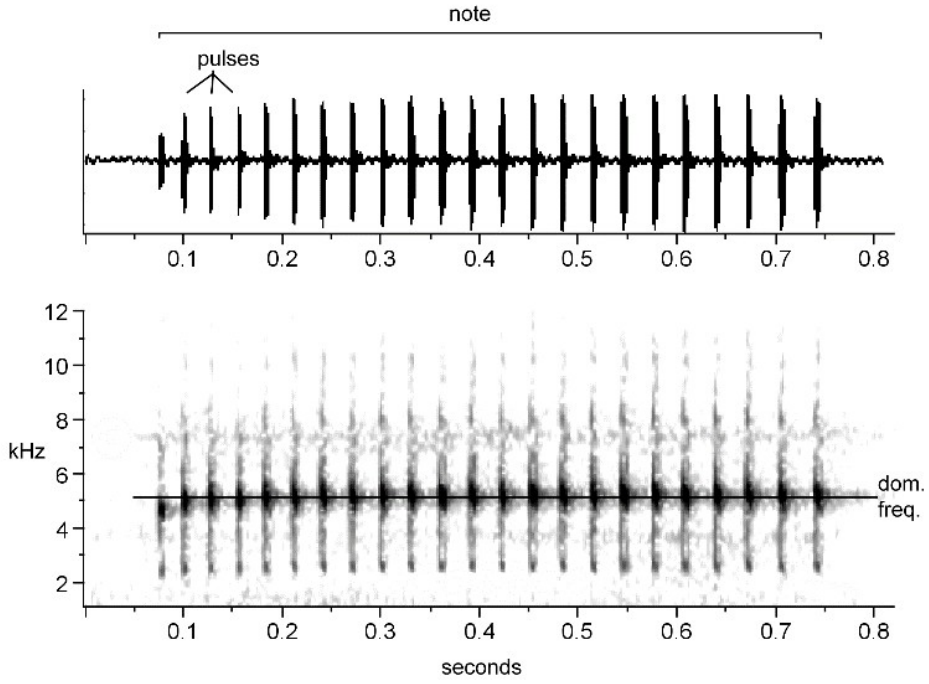


Figure 3-2. Clines in advertisement call parameters across the three mimetic transition zones. (a) Banded-striped transition zone, (b) spotted-striped transition zone, (c) striped-varadero transition zone. In all panels, call parameter values for individual *R. imitator* (represented by dots) are plotted along the sampling transect (x-axis, given in km). For all panels, the fit line represents the best-supported model according to the AICc. Fit lines were plotted using best-fit parameter estimates.

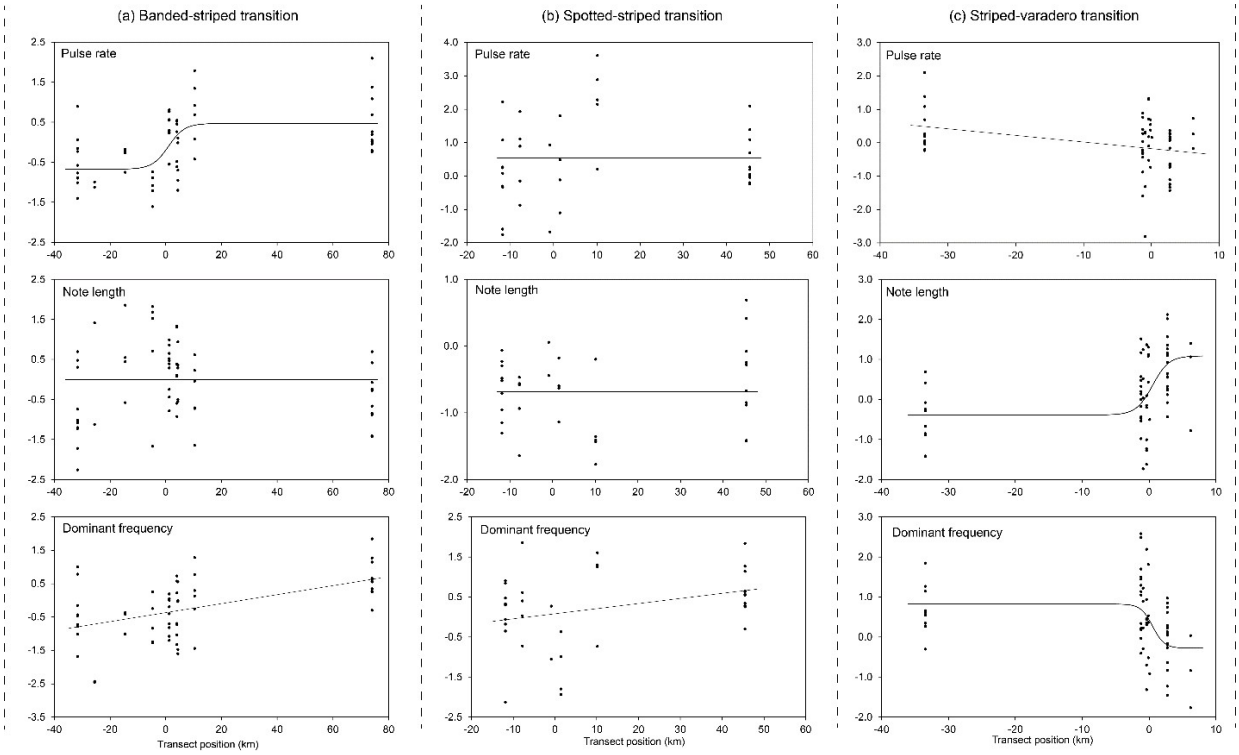


Figure 3-3. Comparison of call parameters between populations used in the striped/varadero phonotaxis experiment. Error bars represent one standard error. Asterisks indicate significant differences following the results of a one-way ANOVA.

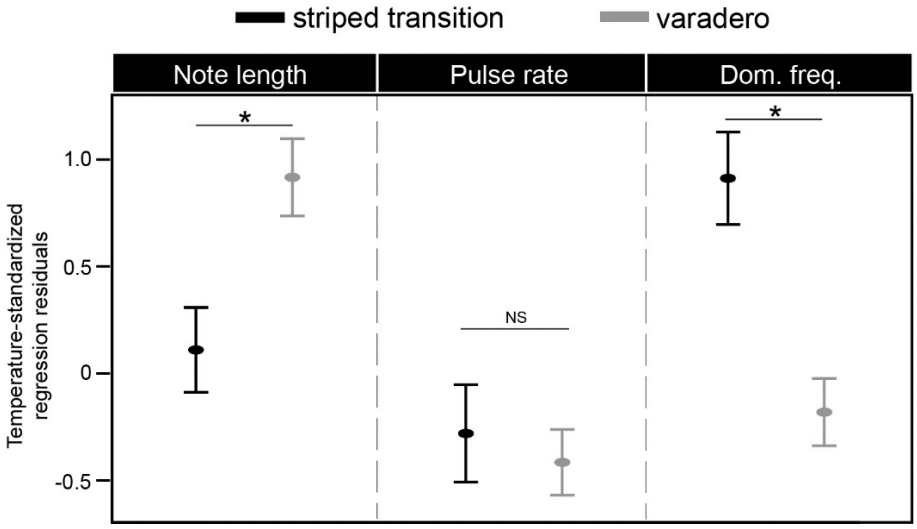


Figure 3-4. Comparison of call parameters between populations used in the striped/banded phonotaxis experiment. Error bars represent one standard error. Asterisks indicate significant differences following the results of a one-way ANOVA.

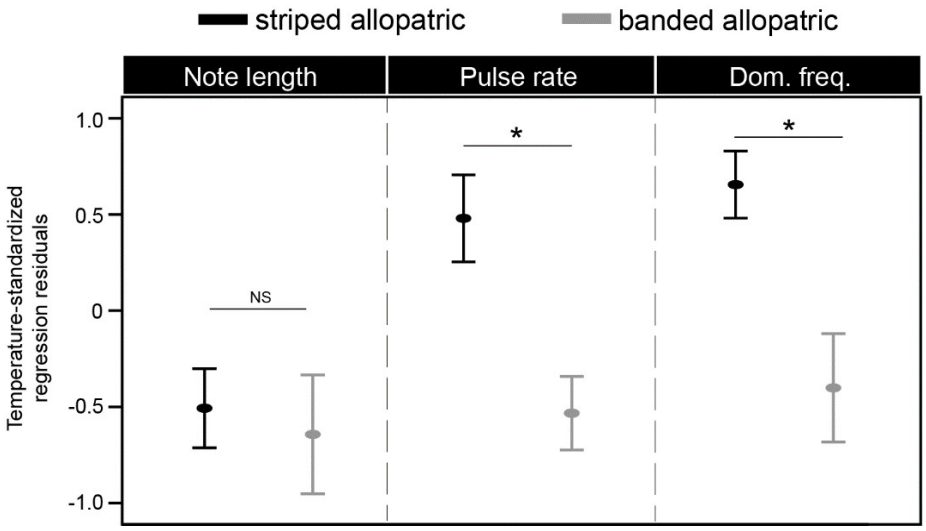


Figure 3-5. Phonotaxis results. For display on the figure, raw interaction times for each trial were converted to a "response index", which was calculated by dividing the time in the interaction zone of one call by the total interaction time. This index therefore ranges from 0 (indicating all interaction was done with one call) to 1 (indicating all interaction was done with the alternate call), with a value of 0.5 (indicated by the dotted line) indicating no preference. Open circles show the mean response index for each population; error bars represent 95% confidence intervals. (a) Striped-varadero phonotaxis experiment. In this panel, response index of 1 indicates preference for long call, value of 0 indicates preference for short call. (b) Striped-banded phonotaxis experiment: dominant frequency manipulation. In this panel, response index of 1 indicates preference for high frequency call, value of 0 indicates preference for low frequency call. (c) Striped-banded phonotaxis experiment: pulse rate manipulation. In this panel, response index of 1 indicates preference for high pulse rate call, value of 0 indicates preference for low pulse rate call. Significance tests based on GLMM detailed in methods.

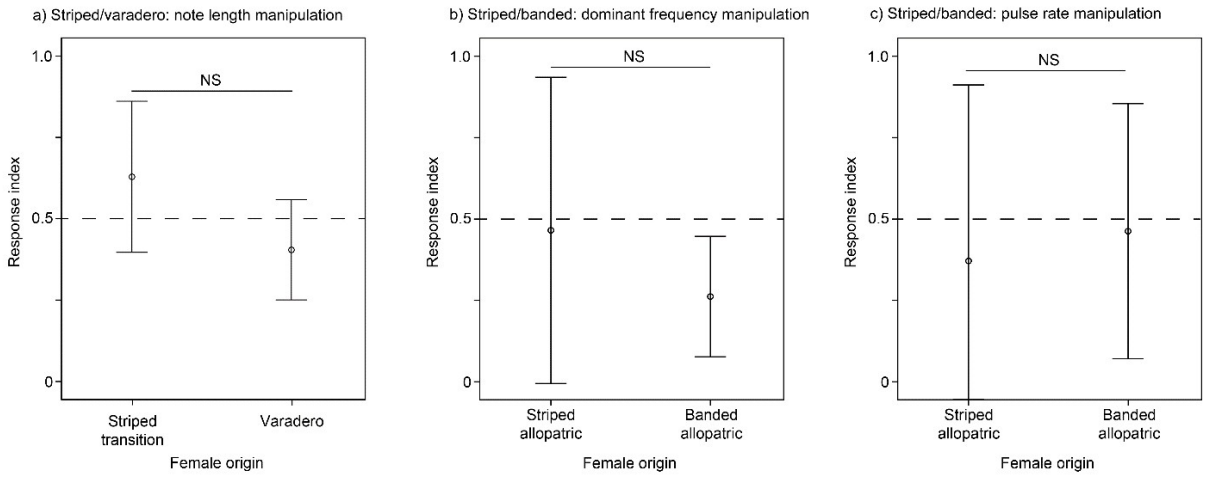


Figure 3-6. Regressions between elevation and average male mass (a), average male mass and note length residuals (b), average male mass and pulse rate residuals (c), and average male mass and dominant frequency residuals (d). Individual dots represent population averages.

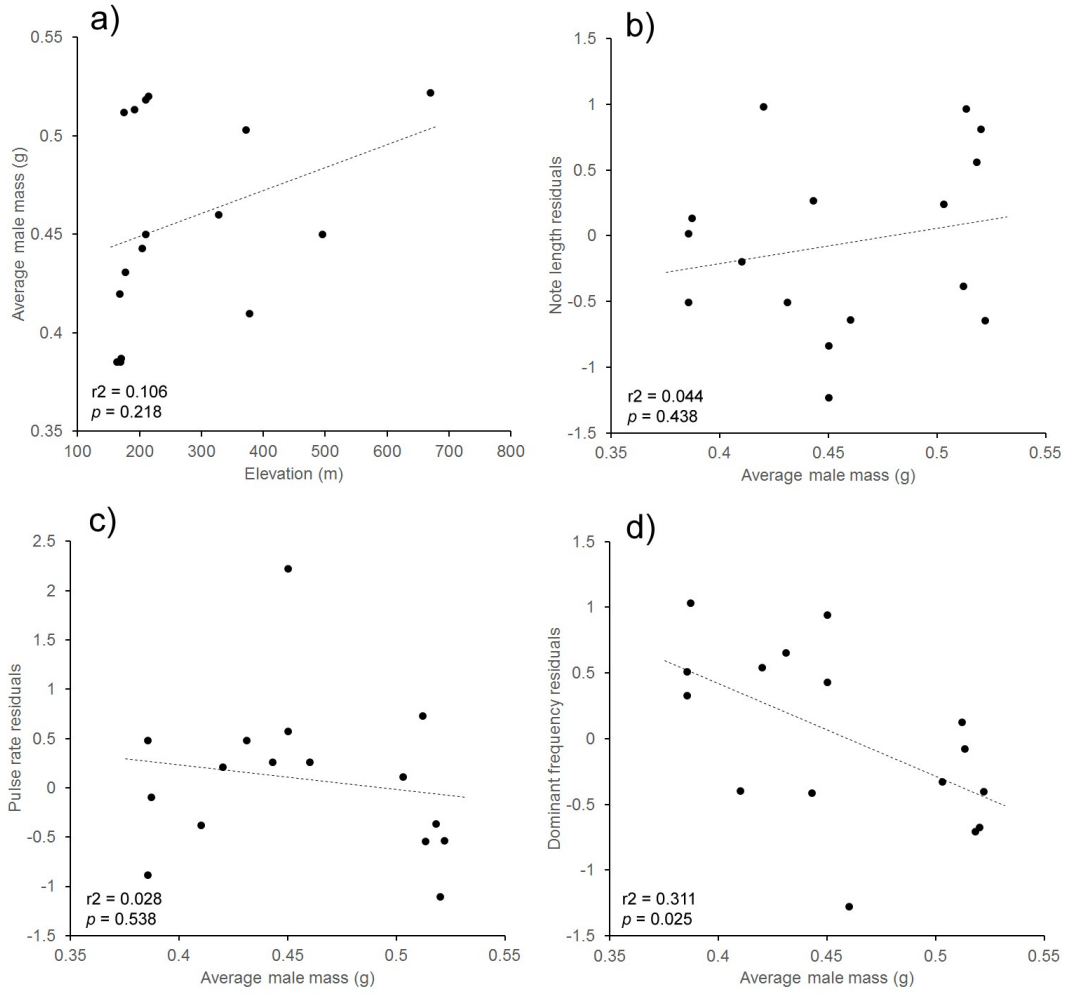


Table 3-1. Sampling localities, transect positions, and call sample sizes for each of the three mimetic transition zones.

transect	site	position on transect	lat	long	call sample size
Banded-striped	Sauce	-31.71	-6.7075	-76.1967	11
	Vaquero	-25.62	-6.6222	-76.1744	2
	Chipaota	-14.57	-6.5851	-76.0808	4
	Curiyacu	-4.77	-6.5741	-75.9920	5
	Chipesa	1.21	-6.5499	-75.9449	10
	Callanayacu North	4.01	-6.5630	-75.9240	7
	Ricardo Palma	4.32	-6.5364	-75.9194	5
	Achinamisa	10.26	-6.4762	-75.9087	6
	Micaela Bastidas*	74.04	-5.9554	-76.2424	11
Spotted-striped	Bocatoma	-11.85	-6.4547	-76.3489	9
	San Jose	-7.85	-6.4328	-76.2829	5
	Sisayacu	-0.91	-6.3705	-76.2908	2
	Upper Pongo	1.44	-6.3520	-76.3009	4
	Pongo	10.1	-6.2927	-76.2354	5
	Micaela Bastidas*	45.42	-5.9554	-76.2424	11
Striped-varadero	Micaela Bastidas*	-33.46	-5.9554	-76.2424	11
	Varadero - South Bank*	-1.25	-5.7177	-76.4163	13
	Varadero - Bridge*	-0.92	-5.7142	-76.4178	4
	Varadero - Stream*	-0.39	-5.7084	-76.4174	6
	Varadero - Transition 1*	-0.16	-5.7073	-76.4161	4
	Varadero - Forest 1*	2.72	-5.6821	-76.4171	17
	Varadero - Forest 3*	6.03	-5.6515	-76.4241	3

* Call data taken from Twomey *et al.* 2014

Table 3-2. Model fit results for call clines. In all cases, the best supported model is shown in bold. For cases where the sigmoid model was best supported, point estimates for cline parameters are given.

transition zone	variable	model	AICc	Δ AICc	Akaike weight	center	width
banded-striped	pulse rate	flat	-24.0	17.1	0.000	–	–
(*n* = 61)		linear	-35.9	5.2	0.068	–	–
		sigmoid	**-41.2**	**0.0**	**0.932**	**0.99**	**11.63**
	note length	**flat**	**-3.1**	**0.0**	**0.635**	–	–
		linear	-1.8	1.3	0.332	–	–
		sigmoid	2.8	6.0	0.032	–	–
	dominant frequency	flat	-9.2	15.9	0.000	–	–
		linear	**-25.1**	**0.0**	**0.959**	–	–
		sigmoid	-18.7	6.3	0.040	–	–
spotted-striped	pulse rate	**flat**	**17.9**	**0.0**	**0.605**	–	–
(*n* = 36)		linear	19.7	1.8	0.241	–	–
		sigmoid	20.6	2.7	0.154	–	–
	note length	**flat**	**-38.7**	**0.0**	**0.685**	–	–
		linear	-37.1	1.7	0.294	–	–
		sigmoid	-31.8	7.0	0.021	–	–
	dominant frequency	flat	0.7	1.3	0.323	–	–
		linear	**-0.6**	**0.0**	**0.620**	–	–
		sigmoid	4.2	4.8	0.057	–	–
striped-varadero	pulse rate	flat	-18.8	4.3	0.100	–	–
(*n* = 64)		**linear**	**-23.1**	**0.0**	**0.844**	–	–
		sigmoid	-17.7	5.4	0.056	–	–
	note length	flat	1.1	13.0	0.001	–	–
		linear	-7.8	4.1	0.115	–	–
		sigmoid	**-11.9**	**0.0**	**0.884**	**0.52**	**4.89**
	dominant frequency	flat	-9.7	9.0	0.011	–	–
		linear	-10.7	8.1	0.017	–	–
		sigmoid	**-18.7**	**0.0**	**0.972**	**0.52**	**3.52**

References

Agapow, P.-M. & Burt, A. (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1, 101–102.

Alexandrou, M.A., Oliveira, C., Maillard, M., McGill, R.A., Newton, J., Creer, S. & Taylor, M.I. (2011) Competition and phylogeny determine community structure in Müllerian co-mimics. *Nature* 469, 84–88.

Arias, C.F., Munoz, A.G., Jiggins, C.D., Mavarez, J., Bermingham, E. & Linares, M. (2008) A hybrid zone provides evidence for incipient ecological speciation in *Heliconius* butterflies. *Molecular ecology* 17, 4699–4712.

Arias, C.F., Rosales, C., Salazar, C., Castaño, J., Bermingham, E., Linares, M. & McMillan, W. (2012) Sharp genetic discontinuity across a unimodal *Heliconius* hybrid zone. *Molecular ecology* 21, 5778–5794.

Backwell, P. & Jennions, M. (1993) Mate choice in the Neotropical frog, *Hyla ebraccata*: sexual selection, mate recognition and signal selection. *Animal Behaviour* 45, 1248–1250.

Barton, N.H. & Gale, K.S. (1993) Genetic analysis of hybrid zones. In: *Hybrid zones and the evolutionary process*. , pp. 13–45.

Barton, N.H. & Hewitt, G. (1985) Analysis of hybrid zones. *Annual review of Ecology and Systematics* 16, 113–148.

Bates, H.W. (1862) Contributions to an Insect Fauna of the Amazon Valley. Lepidoptera: Heliconidae. *Transactions of the Linnean Society of London* 23, 495–566.

Bazykin, A. (1969) Hypothetical mechanism of speciaton. *Evolution*, 685–687.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. (1996) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. *Laboratoire génome, populations, interactions, CNRS UMR* 5000.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 289–300.

Bernal, X.E., Guarnizo, C. & Lüddecke, H. (2005) Geographic variation in advertisement call and genetic structure of *Colostethus palmatus* (Anura, Dendrobatidae) from the Colombian Andes. *Herpetologica* 61, 395–408.

Berner, D., Grandchamp, A.-C. & Hendry, A.P. (2009) Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. *Evolution* 63, 1740–1753.

Blum, M.J. (2002) Rapid movement of a *Heliconius* hybrid zone: evidence for phase III of Wright's shifting balance theory?. *Evolution* 56, 1992–1998.

Bolnick, D.I., Near, T.J. & Wainwright, P.C. (2006) Body size divergence promotes post-zygotic reproductive isolation in centrarchids. *Evolutionary Ecology Research* 8, 903–913.

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.

Brown, J.L., Chouteau, M., Glenn, T. & Summers, K. (2009)a) The development and analysis of twenty-one microsatellite loci for three species of Amazonian poison frogs. *Conservation Genetics Resources* 1, 149–151.

Brown, J.L., Morales, V. & Summers, K. (2009)b) Home range size and location in relation to reproductive resources in poison frogs (Dendrobatidae): a Monte Carlo approach using GIS data. *Animal Behaviour* 77, 547–554.

Brown, J.L., Morales, V. & Summers, K. (2010) A key ecological trait drove the evolution of biparental care and monogamy in an amphibian. *The American Naturalist* 175, 436–446.

Brown, J.L., Twomey, E., Amézquita, A., Barbosa de Souza, M., Caldwell, J.P., Lötters, S., Von May, R., Melo-Sampaio, P.R., Mejia-Vargas, D., Perez-Pena, P. & others (2011) A taxonomic revision of the Neotropical poison frog genus *Ranitomeya* (Amphibia: Dendrobatidae). *Zootaxa* 3083, 1–120.

Brown, J.L., Twomey, E., Morales, V. & Summers, K. (2008)a) Phytotelm size in relation to parental care and mating strategies in two species of Peruvian poison frogs. *Behaviour* 145, 1139–1165.

Brown, J.L., Twomey, E., Pepper, M. & Sanchez-Rodriguez, M. (2008)b) Revision of the *Ranitomeya fantastica* species complex with description of two new species from Central Peru (Anura: Dendrobatidae). *Zootaxa* 1823, 1–24.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York.

Chamberlain, N.L., Hill, R.I., Kapan, D.D., Gilbert, L.E. & Kronforst, M.R. (2009) Polymorphic butterfly reveals the missing link in ecological speciation. *Science* 326, 847–850.

Charif, R., Waack, A. & Strickman, L. (2008) Raven Pro 1.3 user's manual. *Ithaca, NY: Cornell Laboratory of Ornithology*.

Chouteau, M. & Angers, B. (2011) The role of predators in maintaining the geographic organization of aposematic signals. *The American Naturalist* 178, 810–817.

Chouteau, M. & Angers, B. (2012) Wright's shifting balance theory and the diversification of aposematic signals. *PloS one* 7, e34028.

Comeault, A. & Noonan, B. (2011) Spatial variation in the fitness of divergent aposematic phenotypes of the poison frog, *Dendrobates tinctorius*. *Journal of Evolutionary Biology* 24, 1374–1379.

Coyne, J.A. & Orr, H.A. (2004) *Speciation*. Sinauer Associates Sunderland, MA.

Cushman, S.A., McKelvey, K.S., Hayden, J. & Schwartz, M.K. (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist* 168, 486–499.

Dalal, N. & Triggs, B. (2005) Histograms of oriented gradients for human detection. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society*. IEEE, pp. 886–893.

Duellman, W.E. & Trueb, L. (1986) *Biology of Amphibians*. McGraw-Hill, New York.

Earl, D. & vonHoldt, B. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4, 359–361.

Endler, J.A. (1977) *Geographic variation, speciation, and clines*. Princeton University Press.

Endler, J.A. (1990) On the measurement and classification of colour in studies of animal colour patterns. *Biological Journal of the Linnean Society* 41, 315–352.

Endler, J. & Greenwood, J. (1988) Frequency-dependent predation, crypsis and aposematic coloration [and discussion]. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 319, 505–523.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14, 2611–2620.

Feder, J.L. & Nosil, P. (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64, 1729–1747.

Funk, D.J., Nosil, P. & Etges, W.J. (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America* 103, 3209–3213.

Funk, W., Cannatella, D. & Ryan, M. (2009) Genetic divergence is more tightly related to call variation than landscape features in the Amazonian frogs *Physalaemus petersi* and *P. freibergi*. *Journal of Evolutionary Biology* 22, 1839–1853.

Gavrilets, S. (2004) *Fitness landscapes and the origin of species*. Princeton University Press.

Gay, L., Crochet, P.-A., Bell, D.A. & Lenormand, T. (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution* 62, 2789–2806.

Gerhardt, H.C. (1974) Behavioral isolation of the tree frogs, *Hyla cinerea* and *Hyla andersonii*. *American Midland Naturalist*, 424–433.

Gomez, D. (2006) AVICOL, a program to analyse spectrometric data. Last update october 2011. Free executable available at http://sites.google.com/site/avicolprogram/ or from the author at dodogomez@yahoo.fr.

Greene, H.W. & McDiarmid, R.W. (1981) Coral snake mimicry: does it occur? *Science* 213, 1207–1212.

Greenwood, J.J., Cotton, P.A. & Wilson, D.M. (1989) Frequency-dependent selection on aposematic prey: some experiments. *Biological Journal of the Linnean Society* 36, 213–226.

Guerra, M.A. & Ron, S.R. (2008) Mate choice and courtship signal differentiation promotes speciation in an Amazonian frog. *Behavioral Ecology* 19, 1128–1135.

Haldane, J. (1948) The theory of a cline. *Journal of Genetics* 48, 277–284.

Hasegawa, Y., Ueda, H. & Sumida, M. (1999) Clinal geographic variation in the advertisement call of the wrinkled frog, *Rana rugosa*. *Herpetologica* 55, 318–324.

Hatfield, T. & Schluter, D. (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* 53, 866–873.

Hegna, R.H., Saporito, R.A. & Donnelly, M.A. (2012) Not all colors are equal: predation and color polytypism in the aposematic poison frog *Oophaga pumilio*. *Evolutionary Ecology* 27, 1–15.

Hendry, A., Bolnick, D., Berner, D. & Peichel, C. (2009) Along the speciation continuum in sticklebacks. *Journal of Fish Biology* 75, 2000–2036.

Höbel, G., Carl Gerhardt, H. & Noor, M. (2003) Reproductive character displacement in the acoustic communication system of green tree frogs (*Hyla cinerea*). *Evolution* 57, 894–904.

Jiggins, C.D. (2008) Ecological speciation in mimetic butterflies. *Bioscience* 58, 541–548.

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

Jiggins, C., Estrada, C. & Rodrigues, A. (2004) Mimicry and the evolution of premating isolation in *Heliconius melpomene* Linnaeus. *Journal of Evolutionary Biology* 17, 680–691.

Jiggins, C., McMillan, W., King, P. & Mallet, J. (1997) The maintenance of species differences across a *Heliconius* hybrid zone. *Heredity* 79.

Joron, M. & Mallet, J.L. (1998) Diversity in mimicry: paradox or paradigm? *Trends in Ecology & Evolution* 13, 461–466.

Koenderink, J.J. & van Doorn, A.J. (1992) Surface shape and curvature scales. *Image and vision computing* 10, 557–564.

Kronforst, M.R., Young, L.G., Kapan, D.D., McNeely, C., O'Neill, R.J. & Gilbert, L.E. (2006) Linkage of butterfly mate preference and wing color preference cue at the genomic location of *wingless*. *Proceedings of the National Academy of Sciences* 103, 6575–6580.

Kronforst, M., Young, L. & Gilbert, L. (2007) Reinforcement of mate preference among hybridizing *Heliconius* butterflies. *Journal of evolutionary biology* 20, 278–285.

Larsen, A.B., Vestergaard, J.S. & Larsen, R. (2014) HEp-2 cell classification using shape index histograms with donut-shaped spatial pooling. *IEEE Transactions on Medical Imaging* PP, 1–1.

Lemmon, E.M. (2009) Diversification of conspecific signals in sympatry: geographic overlap drives multidimensional reproductive character displacement in frogs. *Evolution* 63, 1155–1170.

Lindeberg, T. (1996) Scale-space: A framework for handling image structures at multiple scales. *CERN European Organization for Nuclear Research-Reports*, 27–38.

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

Maan, M.E. & Cummings, M.E. (2009) Sexual dimorphism and directional sexual selection on aposematic signals in a poison frog. *Proceedings of the National Academy of Sciences* 106, 19072–19077.

Mallet, J. (1986) Hybrid zones of *Heliconius* butterflies in Panama and the stability and movement of warning colour clines. *Heredity* 56, 191–202.

Mallet, J. (2006) What does< i> Drosophila</i> genetics tell us about speciation? *Trends in ecology & evolution* 21, 386–393.

Mallet, J. & Barton, N.H. (1989) Strong natural selection in a warning-color hybrid zone. *Evolution*, 421–431.

Mallet, J., Barton, N., Lamas, G., Santisteban, J., Muedas, M. & Eeley, H. (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics* 124, 921–936.

Mallet, J., Beltrán, M., Neukirchen, W. & Linares, M. (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology* 7, 28.

Marek, P.E. & Bond, J.E. (2009) A Müllerian mimicry ring in Appalachian millipedes. *Proceedings of the National Academy of Sciences* 106, 9755–9760.

McKinnon, J.S., Mori, S., Blackman, B.K., David, L., Kingsley, D.M., Jamieson, L., Chou, J. & Schluter, D. (2004) Evidence for ecology's role in speciation. *Nature* 429, 294–298.

McMillan, W.O., Jiggins, C.D. & Mallet, J. (1997) What initiates speciation in passion-vine butterflies? *Proceedings of the National Academy of Sciences* 94, 8628–8633.

Merrill, R.M., Gompert, Z., Dembeck, L.M., Kronforst, M.R., McMillan, W.O. & Jiggins, C.D. (2011) Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution* 65, 1489–1500.

Merrill, R.M., Wallbank, R.W., Bull, V., Salazar, P.C., Mallet, J., Stevens, M. & Jiggins, C.D. (2012) Disruptive ecological selection on a mating cue. *Proceedings of the Royal Society B: Biological Sciences* 279, 4907–4913.

Mika, S., Rätsch, G., Weston, J., Schölkopf, B. & Müller, K. (1999) Fisher discriminant analysis with kernels. In: *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.* IEEE, pp. 41–48.

Myers, C.W. & Daly, J.W. (1976) Preliminary evaluation of skin toxins and vocalizations in taxonomic and evolutionary studies of poison-dart frogs (Dendrobatidae). *Bulletin of the American Museum of Natural History* 157, 177–262.

Narins, P.M. & Smith, S.L. (1986) Clinal variation in anuran advertisement calls: basis for acoustic isolation? *Behavioral Ecology and Sociobiology* 19, 135–141.

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

Noor, M.A. (1999) Reinforcement and other consequences of sympatry. *Heredity* 83, 503–508.

Nosil, P. (2012) *Ecological speciation*. Oxford University Press, Oxford, UK.

Nosil, P., Crespi, B. & Sandoval, C. (2003) Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270, 1911–1918.

Nosil, P. & Harmon, L. (2009) Niche dimensionality and ecological speciation. In: *Speciation and Patterns of Diversity*. Cambridge University Press, Cambridge, UK, pp. 127–154.

Nosil, P., Harmon, L.J. & Seehausen, O. (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* 24, 145–156.

Nosil, P., Vines, T.H. & Funk, D.J. (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* 59, 705–719.

Oldham, R.S. & Gerhardt, H.C. (1975) Behavioral isolating mechanisms of the treefrogs *Hyla cinerea* and *H. gratiosa*. *Copeia*, 223–231.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P. & Shipley, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4, 535–538.

Peakall, R. & Smouse, P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6, 288–295.

Platz, J.E. (1989) Speciation within the chorus frog *Pseudacris triseriata*: morphometric and mating call analyses of the boreal and western subspecies. *Copeia*, 704–712.

Plowright, R. & Owen, R.E. (1980) The evolutionary significance of bumble bee color patterns: a mimetic interpretation. *Evolution* 34, 622–637.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Puebla, O., Bermingham, E., Guichard, F. & Whiteman, E. (2007) Colour pattern as a single trait driving speciation in *Hypoplectrus* coral reef fishes?. *Proceedings of the Royal Society B: Biological Sciences* 274, 1265–1271.

R Development Core Team (2005) R: A language and environment for statistical computing. *R foundation for Statistical Computing*.

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

Richards-Zawacki, C.L. (2009) Effects of slope and riparian habitat connectivity on gene flow in an endangered Panamanian frog, *Atelopus varius*. *Diversity and Distributions* 15, 796–806.

Richards-Zawacki, C.L. & Cummings, M.E. (2011) Intraspecific reproductive character displacement in a polymorphic poison dart frog, *Dendrobates pumilio*. *Evolution* 65, 259–267.

Richards-Zawacki, C.L., Wang, I.J. & Summers, K. (2012) Mate choice and the genetic basis for colour variation in a polymorphic dart frog: inferences from a wild pedigree. *Molecular ecology* 21, 3879–3892.

Richards-Zawacki, C.L., Yeager, J. & Bart, H.P. (2013) No evidence for differential survival or predation between sympatric color morphs of an aposematic poison frog. *Evolutionary Ecology* 27, 783–795.

Rosenblum, E.B. (2006) Convergent evolution and divergent selection: lizards at the White Sands ecotone. *The American Naturalist* 167, 1–15.

Rosenblum, E.B. & Harmon, L.J. (2011) "Same same but different": replicated ecological speciation at White Sands. *Evolution* 65, 946–960.

Rudh, A., Rogell, B., Hastad, O. & Qvarnström, A. (2011) Rapid population divergence linked with co-variation between coloration and sexual display in strawberry poison frogs. *Evolution* 65, 1271–1282.

Rundle, H.D. & Nosil, P. (2005) Ecological speciation. *Ecology Letters* 8, 336–352.

Ryan, M.J. & Rand, A.S. (1993) Phylogenetic patterns of behavioral mate recognition systems in the *Physalaemus pustulosus* species group (Anura: Leptodactylidae): the role of ancestral and derived characters and sensory exploitation. In: *Evolutionary pattern and processes*. Linnean Society Symposium Series, New York, USA, pp. 251–267.

Ryan, M.J., Rand, A.S. & Weigt, L.A. (1996) Allozyme and advertisement call variation in the túngara frog, *Physalaemus pustulosus*. *Evolution* 50, 2435–2453.

Schluter, D. (1996) Ecological Speciation in Postglacial Fishes. *Royal Society of London Philosophical Transactions Series B* 351, 807–814.

Seehausen, O. (2009) Progressive levels of trait divergence along a 'speciation transect' in the Lake Victoria cichlid fish *Pundamilia*. In: *Speciation and Patterns of Diversity*. Cambridge University Press, Cambridge, UK, pp. 155–176.

Seehausen, O., Terai, Y., Magalhaes, I.S., Carleton, K.L., Mrosso, H.D., Miyagi, R., van der Sluijs, I., Schneider, M.V., Maan, M.E., Tachida, H. & others (2008) Speciation through sensory drive in cichlid fish. *Nature* 455, 620–626.

Silverstone, P.A. (1975) A revision of the poison-arrow frogs of the genus *Dendrobates* Wagler. *Natural History Museum of Los Angeles County, Science Bulletin* 21, 1–55.

Skaug, H., Fournier, D., Nielsen, A., Magnusson, A. & Bolker, B. (2011) glmmADMB: generalized linear mixed models using AD Model Builder. *R package version 0.6* 5, r143. Stuckert, A.M., Saporito, R.A., Venegas, P.J. & Summers, K. (2014)a) Alkaloid defenses of co-mimics in a putative Müllerian mimetic radiation. *BMC Evolutionary Biology* 14,76.

Stuckert, A.M., Venegas, P.J. & Summers, K. (2014)b) Experimental evidence for predator learning and Müllerian mimicry in Peruvian poison frogs (*Ranitomeya*, Dendrobatidae). *Evolutionary Ecology* 28, 413–426.

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.

Summers, K., Cronin, T.W. & Kennedy, T. (2003) Variation in spectral reflectance among populations of *Dendrobates pumilio*, the strawberry poison frog, in the Bocas del Toro Archipelago, Panama. *Journal of Biogeography* 30, 35–53.

Summers, K., Symula, R., Clough, M. & Cronin, T. (1999) Visual mate choice in poison frogs. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266, 2141–2145.

Symula, R., Schulte, R. & Summers, K. (2001) Molecular phylogenetic evidence for a mimetic radiation in Peruvian poison frogs supports a Müllerian mimicry hypothesis. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268, 2415–2421.

Szymura, J.M. & Barton, N.H. (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution* 40, 1141–1159.

Tazzyman, S.J. & Iwasa, Y. (2010) Sexual selection can increase the effect of random genetic drift—a quantitative genetic model of polymorphism in *Oophaga pumilio*, the Strawberry poison-dart frog. *Evolution* 64, 1719–1728.

Telford, S. & Passmore, N. (1981) Selective phonotaxis of four sympatric species of African reed frogs (Genus *Hyperolius*). *Herpetologica*, 29–32.

Tumulty, J., Morales, V. & Summers, K. (2014) The biparental care hypothesis for the evolution of monogamy: experimental evidence in an amphibian. *Behavioral Ecology* 25, 262–270.

Turner, J. (1970) Studies of Müllerian mimicry and its evolution in burnet moths and heliconid butterflies. In: *Ecological genetics and evolution*. Springer, pp. 224–260.

Twomey, E., Vestergaard, J.S. & Summers, K. (2014) Reproductive isolation related to mimetic divergence in the poison frog *Ranitomeya imitator*. *Nature Communications* 5, 1–8.

Twomey, E., Vestergaard, J.S., Venegas, P. & Summers, K. (submitted) Mimetic divergence and the speciation continuum in the mimic poison frog *Ranitomeya imitator*.

Twomey, E., Yeager, J., Brown, J.L., Morales, V., Cummings, M. & Summers, K. (2013) Phenotypic and genetic divergence among poison frog populations in a mimetic radiation. *PloS one* 8, e55443.

Via, S. (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences* 106, 9939–9946.

Wang, I.J. (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* 67, 3403–3411.

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.

Wells, K.D. (2007) *The Ecology and Behavior of Amphibians*. University of Chicago Press.

Wilczynski, W. & Ryan, M.J. (1999) Geographic variation in animal communication systems. In: *Geographic Diversification of Behavior: Perspectives on Evolutionary Mechanisms*. Oxford University Press, New York, pp. 234–261.

Yeager, J. (2009) *Quantification of resemblance in a mimetic radiation*. MS thesis, East Carolina University, Dept. of Biology. Thesis. East Carolina University

Yeager, J., Brown, J.L., Morales, V., Cummings, M. & Summers, K. (2012) Testing for selection on color and pattern in a mimetic radiation. *Current Zoology* 58, 667–675.

Appendix A: IACUC approval letter

Animal Care and Use Committee
East Carolina University
212 Ed Warren Life Sciences Building
Greenville, NC 27834
252-744-2436 office • 252-744-2355 fax

November 11, 2008

Kyle Summers, Ph.D.
Department of Biology
Howell Science Complex
East Carolina University

Dear Dr. Summers:

Your Animal Use Protocol entitled, "The Evolution of Mimicry in Peruvian Poison Frogs," (AUP #D225) was reviewed by this institution's Animal Care and Use Committee on 11/11/08. The following action was taken by the Committee:

"Approved as submitted"

A copy is enclosed for your laboratory files. Please be reminded that all animal procedures must be conducted as described in the approved Animal Use Protocol. Modifications of these procedures cannot be performed without prior approval of the ACUC. The Animal Welfare Act and Public Health Service Guidelines require the ACUC to suspend activities not in accordance with approved procedures and report such activities to the responsible University Official (Vice Chancellor for Health Sciences or Vice Chancellor for Academic Affairs) and appropriate federal Agencies.

Sincerely yours,

Robert G. Carroll, Ph.D.
Chairman, Animal Care and Use Committee

RGC/jd

enclosure