

Genetic basis of aposematic coloration in a mimetic radiation of poison frogs

Authors: Tyler Linderoth^{1,2,3*}, Diana Aguilar-Gómez⁴, Emily White⁵, Evan Twomey⁶, Adam Stuckert⁷, Ke Bi², Amy Ko^{1,4}, Natalie Graham⁸, Joana L. Rocha¹, Jason Chang^{1,9,10}, Matthew D. MacManes¹¹, Kyle Summers^{5*†}, Rasmus Nielsen^{1,2,4,12,13*†}

Affiliations:

¹Department of Integrative Biology, University of California Berkeley; Berkeley, CA 94720, United States of America.

²Museum of Vertebrate Zoology, University of California Berkeley; Berkeley, CA 94720, United States of America.

³W.K. Kellogg Biological Station, Michigan State University; Hickory Corners, MI 49060, United States of America.

⁴Center for Computational Biology, University of California Berkeley; Berkeley, CA 94720, United States of America.

⁵Department of Biology, East Carolina University; Greenville, NC 27858, United States of America.

⁶Faculty of Biological Sciences, Goethe University Frankfurt; Frankfurt am Main, 60438, Germany.

⁷Department of Biology and Biochemistry, University of Houston; Houston, TX 77204, United States of America.

⁸Department of Environmental Science, Policy, and Management, University of California Berkeley; Berkeley, CA 94720, United States of America.

⁹Institute for Computational Medicine, Johns Hopkins University; Baltimore, MD 21218, United States of America.

¹⁰Department of Biomedical Engineering, Johns Hopkins University; Baltimore, MD 21218, United States of America.

¹¹Molecular, Cellular, and Biomedical Sciences, University of New Hampshire; Durham, NH 03824, United States of America.

¹²Department of Statistics, University of California Berkeley; Berkeley, CA 94720, United States of America.

¹³Globe Institute, University of Copenhagen; Copenhagen, 1350, Denmark.

†These authors contributed equally to this work.

*Corresponding author. Email: lindero1@msu.edu, summersk@ecu.edu, rasmus_nielsen@berkeley.edu

Abstract: The evolution of mimicry in a single species or population has rippling inter and intraspecific effects across ecological communities, providing a fascinating mechanism of phenotypic diversification. In this study we present the first identification of genes underlying

Müllerian mimicry in a vertebrate, the Peruvian mimic poison frog, *Ranitomeya imitator*. We sequenced 124 *R. imitator* exomes and discovered loci with both strong divergence between different mimetic morphs and phenotypic associations within an intraspecific admixture zone, implicating *mc1r*, *asip*, *bsn*, *retsat*, and *krt8.2* in the evolution of mimetic color phenotypes. We confirmed these associations for most candidate genes through linkage mapping in a lab-reared pedigree. We also sequenced transcriptomes from the model species, allowing tests for introgression and revealing that the mimetic resemblance between *R. imitator* and the models evolved independently. Selection analyses of the candidate genes show that the mimicry phenotypes likely have evolved through selective sweeps acting on polygenic variation. Our results suggest that the evolutionary origins and molecular mechanisms underlying mimicry phenotypes in vertebrates may be radically different from those previously documented in invertebrates such as the iconic *Heliconius* butterfly mimicry complex.

One Sentence Summary: Müllerian mimicry evolved through independent selective sweeps on color and pattern loci in the mimic poison frog.

Main Text: Mimetic radiations, in which different populations of a species diverge to resemble multiple models or co-mimics, have provided compelling examples of adaptation since the work of Wallace and Müller (1). Butterflies of the genus *Heliconius* have been the main focus of genomic analyses of mimetic radiations, illuminating the genes and gene pathways controlling the color and pattern variation involved in mimicry (2–4). With a few exceptions (e.g. (5)), much of the resemblance between mimics and co-mimics in these radiations appears to stem from the exchange of mimicry genes through introgression (6–8), but it is unclear if this conclusion extends to other species, in particular vertebrates. In fact, there is no vertebrate system for which the molecular mechanism of mimicry has been determined. However, there is a vertebrate that shows remarkable parallels to the *Heliconius* mimicry system: the Peruvian mimic poison frog (*Ranitomeya imitator*) has undergone a mimetic radiation to resemble several existing morphs of congeners (i.e., model species) across northern Peru. We focused on a banded morph that mimics *R. summersi* and a striped morph that mimics *R. variabilis* (Fig. 1) (9, 10).

Amphibians have notoriously large, complex genomes, and poison frogs are no exception (11), making genome sequencing and assembly extremely challenging. To overcome this, we used a combination of exome capture, divergence mapping, association testing within admixed populations, and pedigree linkage analyses to identify genes underlying the divergence of the mimetic morphs of *R. imitator*. We then used population genomic analyses of transcriptome sequence data to determine whether the resemblance between the mimetic populations of *R. imitator* and their model species is due to introgression or independent evolution. Finally, we identified patterns of selection at the molecular level to elucidate the mechanisms of molecular adaptation.

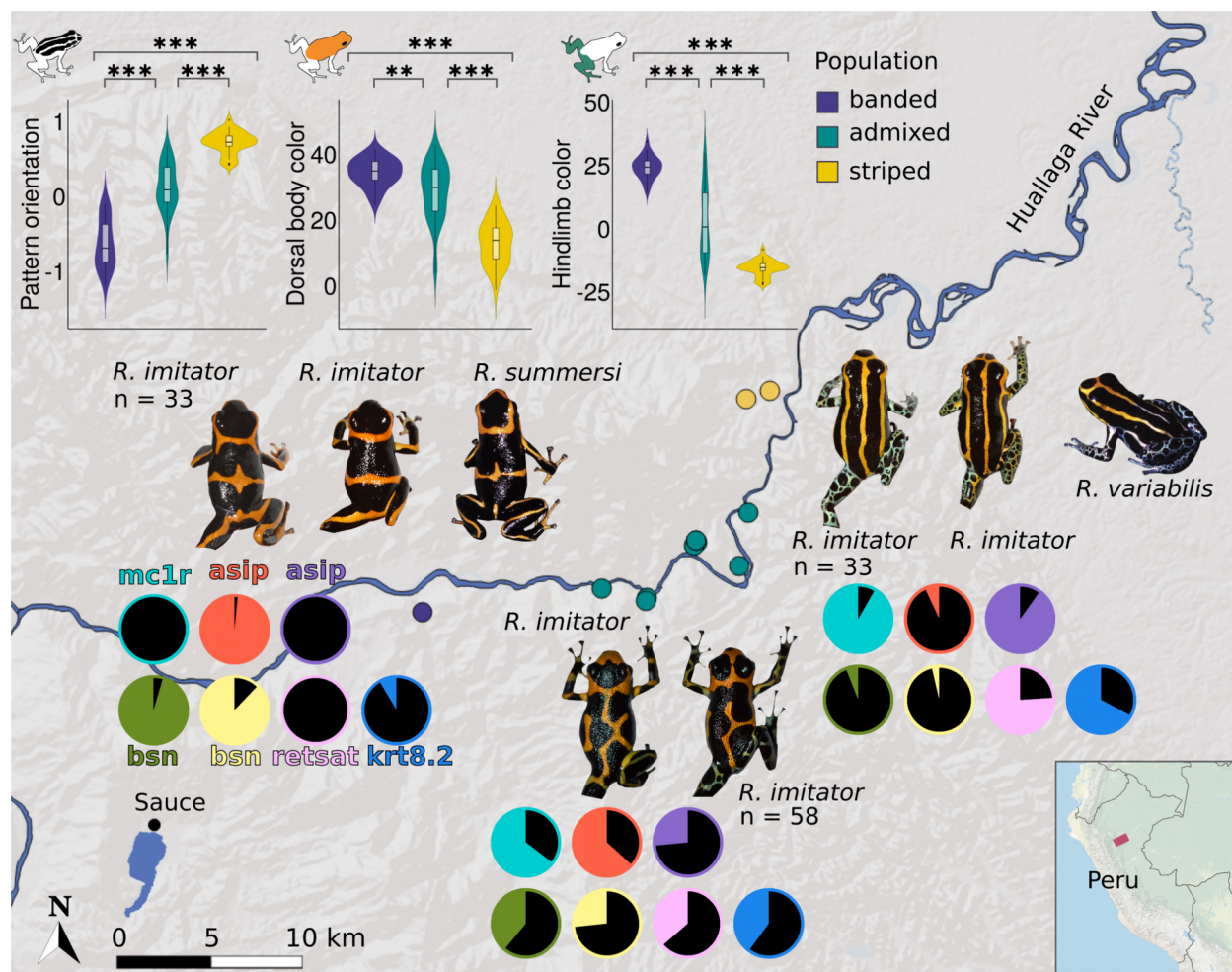


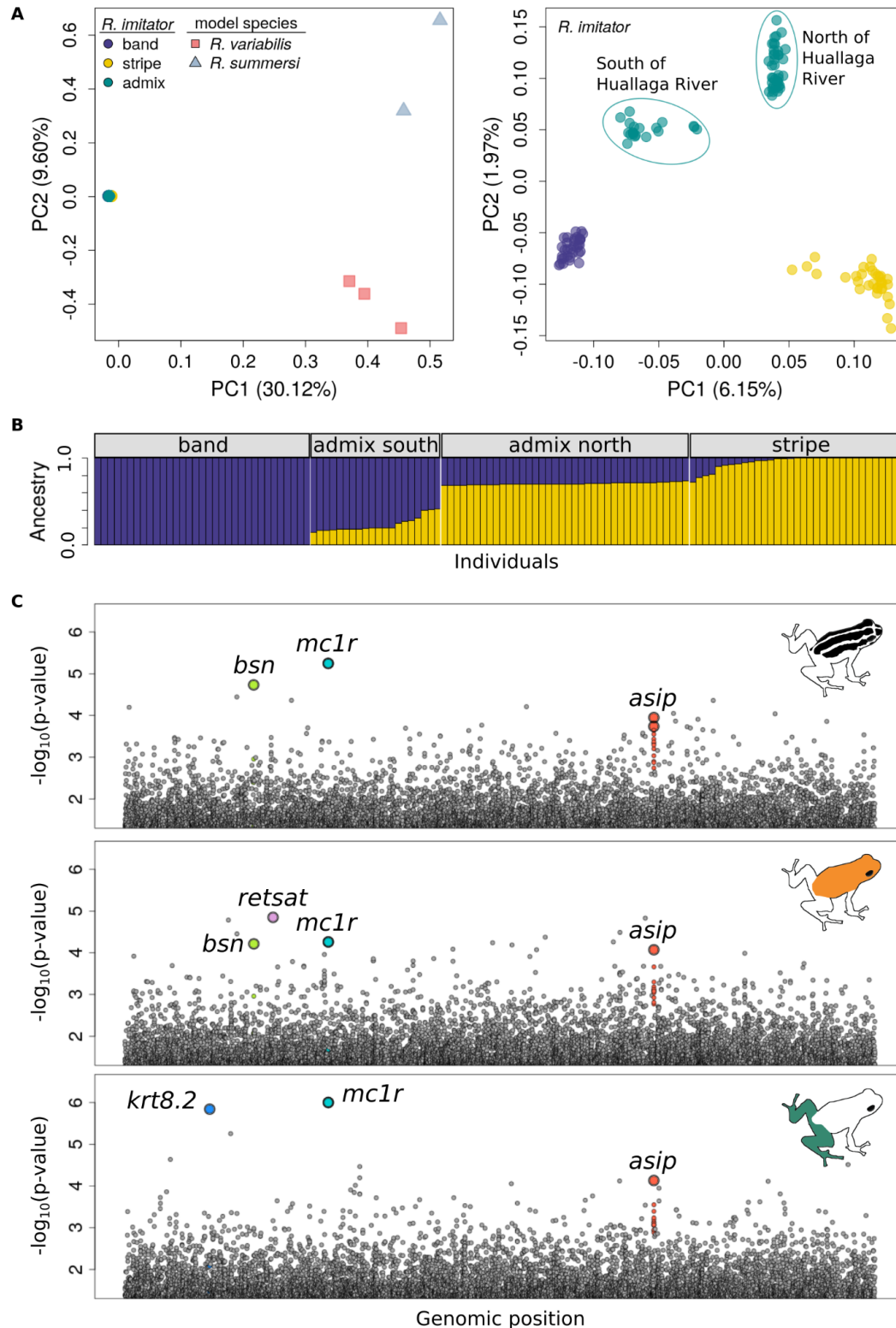
Fig. 1. *Ranitomeya imitator* phenotypes and allele frequencies at candidate color genes. Map showing the sampling localities (small colored dots, red rectangle in the inset map) and sizes of the three main morphs of *R. imitator* in San Martin province, Peru. Traversing northeast along the Huallaga river region, *R. imitator* switches from a banded morph resembling sympatric *R. summersi* to a striped morph that mimics sympatric *R. variabilis* after passing through a 7 km-wide transition zone where *R. imitator* exhibits a spectrum of intermediate phenotypes. Differences in coloration across this mimicry gradient are significant as shown by the violin plots (two-tailed p-value for Mann-Whitney U: ** p < 0.01, *** p < 0.001). Pie charts depict the frequencies of reference (black) and alternate (non-black) alleles of different candidate gene SNPs that have the strongest association with at least one of the coloration phenotypes.

Population genomic characterization of the *Ranitomeya* complex

To enable population-scale genomic analyses in *Ranitomeya imitator*, which has an estimated genome size of 6.8 Gb (12), we captured and sequenced exons from 13,084 genes in 124 individuals representing three distinct phenotypes (Fig. 1). We also sequenced transcriptomes

from the models *R. summersi* and *R. variabilis*. We mapped reads to a *de novo* exome assembly and identified 928,822 high-quality genetic variants across the *Ranitomeya* complex. Most of this genetic variation is attributable to differences between *R. imitator* and the model species, *R. summersi* and *R. variabilis*, represented by PC1 in the principal component analysis (PCA) of exome-wide genetic variation (Fig. 2A, left panel). The tight cluster of *R. imitator* in the middle of PC2, which represents differences from the model species, hints that *R. imitator*'s phenotypic likeness to the models may not derive from introgression.

Among 295,279 SNPs that segregate within *R. imitator*, most genetic differences are between the banded and striped morph populations, defining PC1 of *R. imitator* variation (Fig. 2A, right panel). The exome-wide F_{ST} between the striped and banded morphs is 0.17, which is appropriate for identification of candidate regions underlying color differences through scans of divergence (13). The admixed individuals are positioned between the pure morphs on PC1, confirming that this population is genetically intermediate (10), and are further separated into two clusters coinciding with subpopulations north and south of the Huallaga River (Fig. 1). There is greater genetic similarity among admixed individuals to the pure morph population from the same side of the river, suggesting that the river is a barrier to gene flow that favored higher migration from the morph residing on the like side. This is consistent with relative amounts of striped/banded ancestry inferred in admixture analyses (Fig. 2B). Importantly for the purposes of trait mapping, the levels of ancestry are fairly uniform across individuals within the north-of-river and south-of-river admixed subpopulations (Fig. 2B), suggesting that the admixture zone is sufficiently old (14) for recombination to break up genetic linkage. This is also consistent with the observation that admixed individuals are separated from the pure morph populations along PC2 of the PCA (Fig. 2A, right panel).



within *R. imitator* places individuals from the introgression zone between the striped and banded populations along PC1. The second axis of genetic variation, PC2, distinguishes the admixed frogs from the striped and banded populations. **(B)** NGSadmix (57) analysis showing different intermediate proportions of striped versus banded ancestry (colored bars) among admixed frogs sampled north and south of the Huallaga River. **(C)** Manhattan plots of $-\log_{10}$ -scaled Fisher's combined p-values from the test for excessive allele frequency differences between banded versus striped morph populations and tests of association between genotypes of admixed individuals and different aspects of coloration. Scans for SNPs associated with pattern orientation, dorsal body color, and hindlimb color are shown in the top, middle, and bottom panels, respectively. Exons are arbitrarily ordered along the x-axes. SNPs (points) in candidate coloration genes are displayed in color.

Genetic basis of mimetic phenotypes

Divergence mapping

We performed a scan across 266,944 SNPs distributed in 12,972 genes in 33 pure striped and 33 banded *R. imitator* to identify loci with higher allele frequency differences (δ) than expected given the background divergence between these two morphs. We assigned approximate p-values to outliers using simulations based on the Balding-Nichols model (15, 16) for the joint distribution of allele frequencies between the two morphs. Even very divergent SNPs (several with $\delta > 90\%$) were not significant after Bonferroni corrections to account for the number of tested SNPs. However, these corrections are very conservative and among the 20 most divergent SNPs (table S1) across the exome (> 99 δ percentile) we identified 11 SNPs in three genes (*mc1r*, *asip*, *bsn*) with relevant functions for influencing color phenotypes.

The most diverged SNP in the *R. imitator* exome ($\delta = 92\%$) was located in the melanocortin 1 receptor gene, *mc1r*, which encodes a protein that operates in the synthesis of melanin pigments and melanosome aggregation. This gene has been extensively documented as affecting pigmentation in various taxa (17–20). The next most diverged SNP ($\delta = 91\%$) is located in the agouti-signaling protein gene, *asip*, which produces a protein that competes with melanocortins to bind Mc1r, thereby inhibiting melanin dispersion in melanocytes. Increased *asip* expression has been linked to lighter coat colors in mammals (21, 22) and the countershading patterning of fish (23). Among the 20 most divergent SNPs across the exome, nearly half (eight) were located in *asip*. A third candidate gene, *bsn*, harbors the seventh most diverged SNP ($\delta = 89\%$). This gene has not previously been implicated in color phenotypes, but its product, Bassoon, is a scaffolding protein which could play a role in spatially organizing organelles involved in pigmentation such as melanosomes within chromatophores.

Dissecting the genetic basis for different features of mimetic phenotypes

Next, we tested for associations between 160,285 SNPs from 12,739 genes and quantitative measures of pattern orientation, dorsal body color, and hindlimb color in 58 admixed individuals from an introgression zone between the striped and banded morphs of *R. imitator* (Fig. 1). We combined the p-values from this association test with those of the divergence test using Fisher's Method (24) since SNPs with the largest combined test statistic, X^2 , are both highly divergent between the striped and banded morphs and highly correlated with a continuous color feature in admixed individuals, consistent with a causal color locus. Though no SNPs remained significant after a Bonferonni correction, all three candidate loci from the divergence mapping (*mc1r*, *asip*, *bsn*) were among the top 20 (> 99.9 percentile) highest X^2 values for either pattern orientation, dorsal body color, and/or hindlimb color (Fig. 2C, tables S5-7).

Combining evidence across mapping experiments also revealed two additional strong candidate color genes, *retsat* and *krt8.2*. *Retsat*, the highest-ranked X^2 SNP for dorsal color (table S6), is involved in retinol metabolism which could influence the orange hue of the dorsum through regulation of carotenoid levels. *Krt8.2*, the third-ranking X^2 SNP for hindlimb color, which encodes a type II basic intermediate filament protein that polymerizes with other keratins to form a cytoskeletal component of epithelial cells, could play a role in generating the blue/green hues of the hindlimbs since keratins are involved in structurally generating these colors in birds (25, 26).

SNPs with the largest X^2 values (strongest association) in candidate genes for each respective coloration phenotype on average explained 7%, 6%, and 9% of the pattern, dorsal color, and hindlimb color variation, respectively, in the admixed frogs (table S8, see "Phenotypic variance explained by candidate SNPs" of supplementary text), with the amount of phenotypic variance explained by any single candidate SNP ranging from 1.2-17.6%.

Linkage analyses in pedigrees

To validate the candidate genes identified by divergence and association mapping, we performed linkage analyses in a captive pedigree of seven *R. imitator* families. Inter-morph crosses between three banded and three striped parents produced seven families with nine F2 individuals on average (1-15 F2s per family), totalling 14 F1 and 62 F2 individuals across the entire pedigree (Fig. 3A).

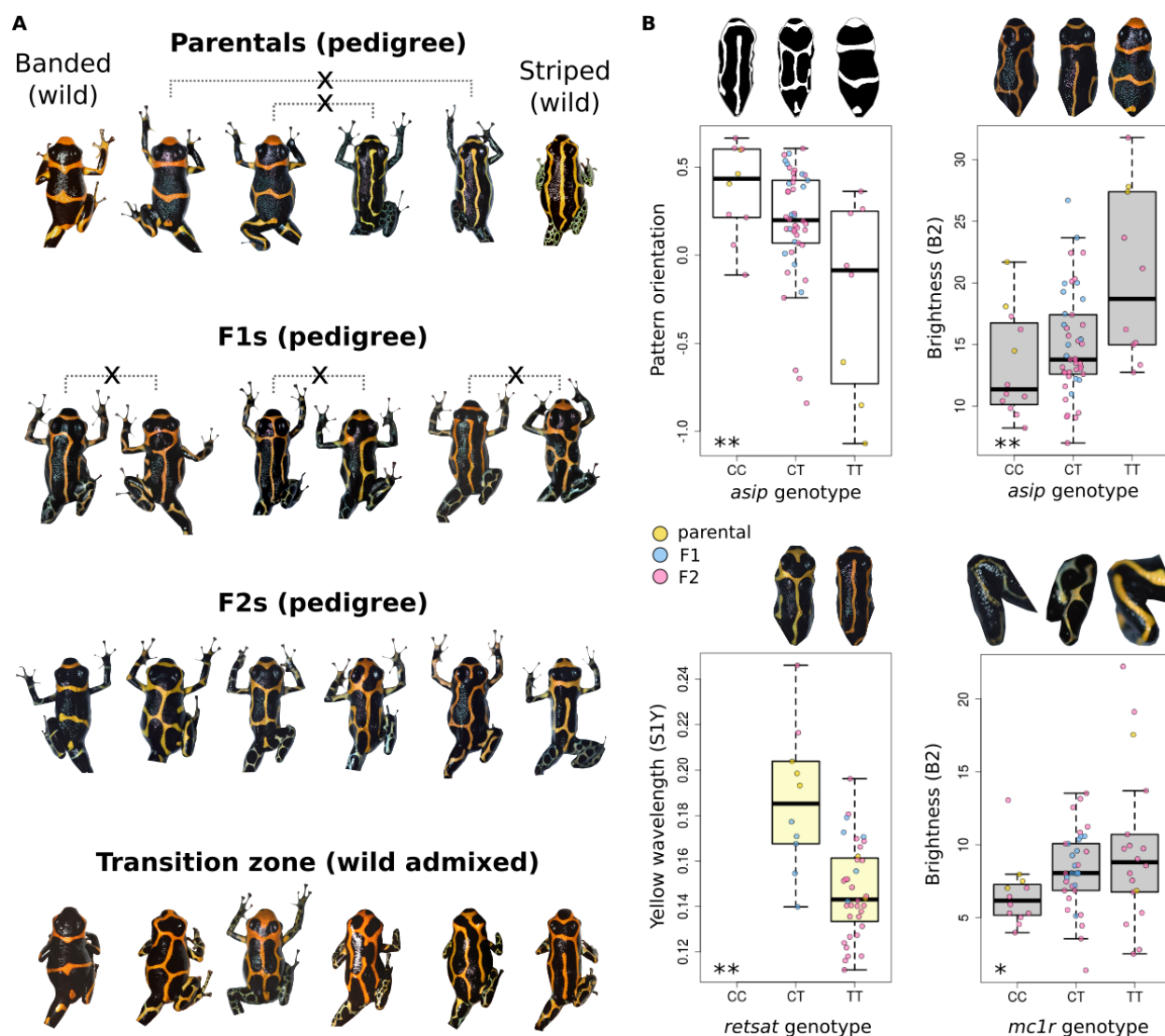


Fig. 3. Relationship between candidate SNP genotypes and phenotypes in a captive pedigree of *R. imitator*. (A) Crosses initiated between captive striped and banded *R. imitator* produced F1 and subsequent F2 generation offspring exhibiting phenotypes that closely mirror those of wild frogs from the admixture zone. (B) Boxplots show significant (LOD score q -value $< 0.1^*$, q -value $< 0.01^{**}$) relationships between allelic dosage at candidate gene SNPs and different color or pattern phenotypes of *R. imitator* crossed in captivity for three generations. The FDR-corrected q -scores are calculated for a total of 14 tests involving one SNP from each of four candidate genes and five phenotypic features.

The spectrum of hues and patterns on the body and legs of the pedigree population closely resembled that of wild frogs (Fig. 3A). Each of the genotyped SNPs from the putative pattern candidate genes, *mc1r*, *asip*, and *bsn*, showed significant associations with pattern orientation among pedigree frogs (table S9, LOD range 0.93-2.54, q -value range 0.003 - 0.091). Of the

candidate genes for dorsal body color, the genotyped *asip* SNP was significantly associated with body brightness (table S9, LOD = 2.47, q-value = 0.003) while *retsat* had a highly significant association with yellow intensity on the body (LOD = 3.49, q-value = 8.54×10^{-4}). Neither *mc1r* or *bsn* were significantly associated with body brightness or yellow intensity in the pedigree (LOD scores ranging 0.09-0.48, q-values = 0.238-0.571). For the hindlimbs, we tested SNPs in the hindlimb color candidate genes *mc1r* and *asip* for associations with brightness and green wavelengths in the pedigree (table S9). *Mc1r* was significantly associated with hindlimb brightness (LOD = 1.13, q-value = 0.064), while *asip* was marginally so (LOD = 0.62, q-value = 0.182). See “Linkage analysis in a captive *R. imitator* pedigree” section of the supplement for further details.

Mimetic phenotypes were not facilitated by introgression

In butterfly mimicry systems some species closely match model phenotypes due to alleles acquired through introgression from the models (8, 27). To determine whether this is the case for *R. imitator*, we tested for patterns of excess genome-wide allele sharing between *R. imitator* and their respective model species using ABBA-BABA tests (Fig. 4A, table S10), which provided no evidence for admixture between *R. imitator* and any of the models (D-statistic = -0.002, p-value = 0.6). Similarly, neighbor-joining trees for the specific genes associated with the mimetic phenotypes show clear separation by species (Fig. 4B), corroborating that there are no shared haplotypes between *R. imitator* and the model species.

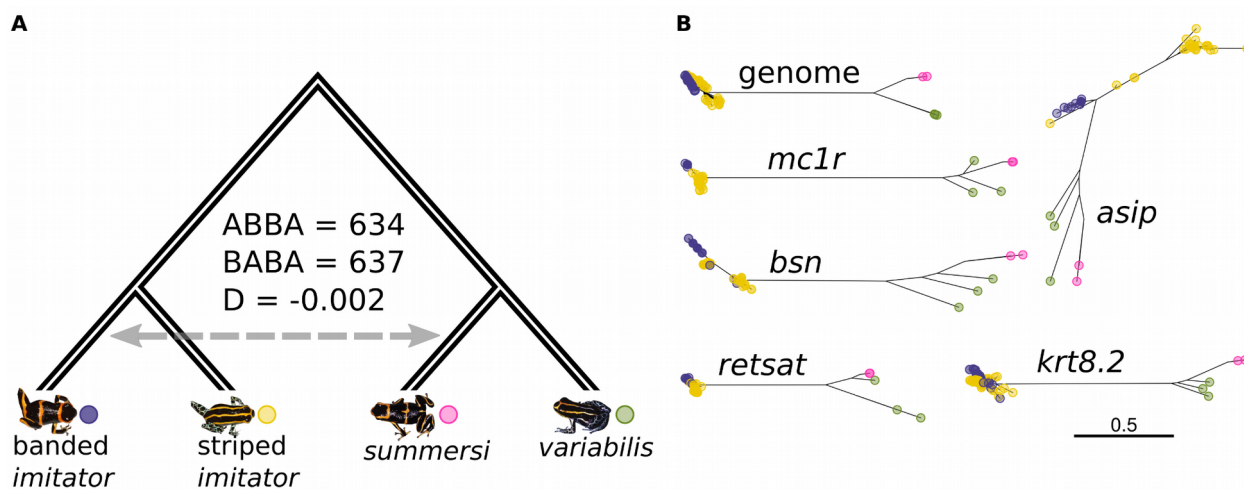


Fig. 4. Genetic comparisons between different *R. imitator* morphs and the species that they mimic. (a) Genome-wide patterns of ABBA-BABA allele topologies among the banded and striped *R. imitator* populations and their respective model species, *R. summersi* (banded) and *R. variabilis* (striped), shows no enrichment of BABA patterns (D = -0.002, p-value = 0.6), suggesting no introgression between *R. imitator* and the models. (b) Unrooted neighbor-joining trees show that all banded (purple dots) and striped (yellow dots) *R. imitator* individuals are clearly more phylogenetically distant to *R. summersi* (pink dots) and *R. variabilis* (green dots)

individuals than they are to each other, both across the entire genome and at each of the candidate coloration genes.

Evolution of mimicry in *R. imitator*

We used Hudson–Kreitman–Aguadé (HKA) tests (28) to determine whether levels of polymorphism within *R. imitator* relative to divergence from the model species was consistent with neutral evolution at the candidate gene exons that contained the strongest phenotype-associated SNPs (table S11). Based on χ^2_1 approximations with genomic control for the goodness-of-fit between observations and neutral expectations, these tests rejected neutrality at the candidate exons of *mc1r*, *bsn*, and *retsat* ($X^2_{imitator} = 5.82 - 18.35$, $q = 1.101e-4 - 0.032$). Furthermore, there was significant evidence ($X^2 = 4.46 - 22.44$, $q = 1.304e-5 - 0.046$) of non-neutral evolution at candidate exons of *mc1r*, *asip*, *bsn*, and *retsat* within at least one of the pure *R. imitator* morphs (table S11). In all cases where exons significantly deviated from neutral expectations, we observed decreased polymorphism within species or populations compatible with selective sweeps (table S11). An alternative approach for obtaining p-values using coalescent simulations, instead of genomic control, also supported non-neutral evolution at the candidate genes (see “Tests of neutral evolution at candidate color genes” of the supplementary text).

Discussion

Coloration is one of the most eco-evolutionarily important traits of dendrobatid frogs, yet little is known regarding its genetic basis (29), except for a few transcriptomic studies that generated extensive lists of candidate genes (12, 30). We used divergence and association mapping to identify candidate genes, which we subsequently validated using linkage analyses in pedigrees. While some of these candidate genes, particularly *mc1r* and *asip*, have been identified to influence color in many previous studies, we also discovered new genes such as *bsn* and *krt8.2*.

Mc1r and its competitive antagonist, *asip*, are directly involved in melanogenesis and have been linked to color phenotypes across taxa (31–36), including the harlequin poison frog, *Oophaga histrionica*, where *mc1r*-mediated melanosome density influences dorsal background color (37). In *R. imitator*, both *mc1r* and *asip* appear to be involved in pattern formation, dorsal body color, and hindlimb color. Another gene that appears to be involved in dorsal pattern and color phenotypes is *bsn*, which encodes a large scaffolding protein (Bassoon) involved in the structural and functional organization of the cytomatrix of the active zone in synapses (38). The iridophore platelets of dendrobatids are arranged in vertical stacks which require a mechanism to govern spacing and thickness, providing one possible role for such a scaffolding protein. Alternatively, it could be anchoring pigment-containing organelles (chromatophores) in specific fields of tissue to create bands, stripes, or other pattern elements. For example, in the retina Bassoon functions in

anchoring the photoreceptor ribbon to the synapse (39). Additional studies are needed to understand the novel functional connection between *bsn* and color patterning.

The last two candidate genes, *retsat* and *krt8.2*, are exclusively associated with colors from different body parts. *RetSAT* is associated with dorsal body color, a pattern found in previous studies on fish. *RetSAT1* (a *retsat* homolog) is differentially expressed at higher levels in orange skin in clownfish (40) and yellow compared to white skin in an East African cichlid fish (41). Although the link between *retsat* and carotenoid coloration remains unclear, *retsat* is known to act on retinol, a carotenoid cleavage product (42), as well as galloxanthin, a colorless apocarotenoid (43). Given its link to yellow/orange coloration in other taxa and the range of reddish/orange to yellow coloration displayed on the dorsum of *R. imitator*, *retsat* is an intriguing color candidate gene whose function warrants further investigation. *Krt8.2*, which encodes a keratin protein found in epithelial cells, was associated with hindlimb color, which varies from orange to blue-green, the latter of which requires co-localization of melanosomes and overlying chromatophores (25). In another poison frog species, *Dendrobates auratus*, differential expression of this gene was associated with variation in skin color (44). So while keratin is not currently known to directly influence amphibian skin color, it is possible that keratinocytes may be playing a role in melanosome transport and organization (45) in poison frogs.

The identification of likely causal genes involved in *R. imitator* mimetic phenotypes provides the first opportunity for elucidating the evolution of mimicry in a vertebrate using molecular methods. Despite evidence that Müllerian mimicry operates in snakes (46), birds (47), catfish (48), bumblebees (49), ants (50), and beetles (51), virtually all of our empirical understanding of its genetic and evolutionary basis (52) stems from research on Neotropical butterflies (53, 54), (see (50) for an exception). This rich body of butterfly work has made it clear that species can evolve to mimic sympatric species through hybridization with the models, thereby obtaining mimetic variants through introgression. In contrast to these findings in butterflies, we find no evidence for introgression with the model species for any of the candidate genes in *R. imitator*. Instead, the large allele frequency differences between *R. imitator* morphs and reduced ratio of segregating to fixed differences with respect to the model species at the candidate genes, suggests rapid turnover of alleles at color and pattern loci driven by partial or full selective sweeps on standing or *de novo* variation. We might expect such a scenario in a dynamic ecological environment where the interaction of *R. imitator* populations with model species changes relatively rapidly on an evolutionary time scale.

The combined contribution of the candidate genes to the phenotypes is relatively low (sum of variance explained < 28%). This may partially be explained by measurement error and incomplete linkage disequilibrium between the most highly associated SNPs and causal variants, but also suggests that the mimicry phenotypes are somewhat polygenic. It should come as no surprise that such complex phenotypes have a complex genetic basis. However, in butterflies,

complex mimicry color patterns have been found to be caused by a supergene controlling multiple phenotypes (55). Our observations of a polygenic basis, absence of introgression and evidence for partial local sweeps, suggests a quite different genetic and evolutionary model for the evolution of mimicry phenotypes in poison frogs and perhaps other vertebrates than that observed in butterflies.

In conclusion, this study provides, for the first time, a link between genetics and coloration mimicry in vertebrates, with candidate genes cross-validated through pedigree analyses. It appears that this mimicry evolved independently in the mimic poison frog (relative to the model species), likely through repeated partial or full sweeps, providing strong evidence for alternative routes to this type of diversification apart from introgression. This study also demonstrates a viable genomics-enabled path towards filling a stark gap in understanding the genetic basis of color phenotypes in amphibians (56) even in the absence of a reference genome.

References and Notes

1. G. D. Ruxton, W. L. Allen, T. N. Sherratt, M. P. Speed, *Avoiding Attack* (Oxford University Press, ed. 2, 2018).
2. M. R. Kronforst, R. Papa, The Functional Basis of Wing Patterning in *Heliconius* Butterflies: The Molecules Behind Mimicry. *Genetics*. **200**, 1–19 (2015).
3. R. M. Merrill, K. K. Dasmahapatra, J. W. Davey, D. D. Dell’Aglia, J. J. Hanly, B. Huber, C. D. Jiggins, M. Joron, K. M. Kozak, V. Llaurens, S. H. Martin, S. H. Montgomery, J. Morris, N. J. Nadeau, A. L. Pinharanda, N. Rosser, M. J. Thompson, S. Vanjari, R. W. R. Wallbank, Q. Yu, The diversification of *Heliconius* butterflies: what have we learned in 150 years? *J. Evol. Biol.* **28**, 1417–1438 (2015).
4. N. J. Nadeau, C. Pardo-Diaz, A. Whibley, M. A. Supple, S. V. Saenko, R. W. R. Wallbank, G. C. Wu, L. Maroja, L. Ferguson, J. J. Hanly, H. Hines, C. Salazar, R. M. Merrill, A. J. Dowling, R. H. French-Constant, V. Llaurens, M. Joron, W. O. McMillan, C. D. Jiggins, The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*. **534**, 106–110 (2016).
5. N. W. VanKuren, D. Massardo, S. Nallu, M. R. Kronforst, Butterfly Mimicry Polymorphisms Highlight Phylogenetic Limits of Gene Reuse in the Evolution of Diverse Adaptations. *Mol. Biol. Evol.* **36**, 2842–2853 (2019).
6. N. B. Edelman, P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey, R. B. Dikow, G. García-Accinelli, S. M. Van Belleghem, N. Patterson, D. E. Neafsey, R. Challis, S. Kumar, G. R. P. Moreira, C. Salazar, M. Chouteau, B. A. Counterman, R. Papa, M. Blaxter, R. D. Reed, K. K. Dasmahapatra, M. Kronforst, M. Joron, C. D. Jiggins, W. O. McMillan, F. Di Palma, A. J. Blumberg, J. Wakeley, D. Jaffe, J. Mallet, Genomic architecture and introgression shape a butterfly radiation. *Science*. **366**, 594–599 (2019).
7. K. K. Dasmahapatra, J. R. Walters, A. D. Briscoe, J. W. Davey, A. Whibley, N. J. Nadeau, A. V. Zimin, D. S. T. Hughes, L. C. Ferguson, S. H. Martin, C. Salazar, J. J. Lewis, S. Adler, S.-J. Ahn, D. A. Baker, S. W. Baxter, N. L. Chamberlain, R. Chauhan, B. A. Counterman, T. Dalmay, L. E. Gilbert, K. Gordon, D. G. Heckel, H. M. Hines, K. J. Hoff, P. W. H. Holland, E. Jacquin-Joly, F. M. Jiggins, R. T. Jones, D. D. Kapan, P. Kersey, G.

- Lamas, D. Lawson, D. Mapleson, L. S. Maroja, A. Martin, S. Moxon, W. J. Palmer, R. Papa, A. Papanicolaou, Y. Pauchet, D. A. Ray, N. Rosser, S. L. Salzberg, M. A. Supple, A. Surridge, A. Tenger-Trolander, H. Vogel, P. A. Wilkinson, D. Wilson, J. A. Yorke, F. Yuan, A. L. Balmuth, C. Eland, K. Gharbi, M. Thomson, R. A. Gibbs, Y. Han, J. C. Jayaseelan, C. Kovar, T. Mathew, D. M. Muzny, F. Onger, L.-L. Pu, J. Qu, R. L. Thornton, K. C. Worley, Y.-Q. Wu, M. Linares, M. L. Blaxter, R. H. ffrench-Constant, M. Joron, M. R. Kronforst, S. P. Mullen, R. D. Reed, S. E. Scherer, S. Richards, J. Mallet, W. O. McMillan, C. D. Jiggins, Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. **487**, 94–98 (2012).
8. W. Zhang, K. K. Dasmahapatra, J. Mallet, G. R. P. Moreira, M. R. Kronforst, Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biol.* **17**, 25 (2016).
 9. R. Symula, R. Schulte, K. Summers, Molecular phylogenetic evidence for a mimetic radiation in Peruvian poison frogs supports a Müllerian mimicry hypothesis. *Proc. R. Soc. Lond. B Biol. Sci.* **268**, 2415–2421 (2001).
 10. E. Twomey, J. S. Vestergaard, P. J. Venegas, K. Summers, Mimetic Divergence and the Speciation Continuum in the Mimic Poison Frog *Ranitomeya imitator*. *Am. Nat.* **187**, 205–224 (2016).
 11. R. L. Rogers, L. Zhou, C. Chu, R. Márquez, A. Corl, T. Linderoth, L. Freeborn, M. D. MacManes, Z. Xiong, J. Zheng, C. Guo, X. Xun, M. R. Kronforst, K. Summers, Y. Wu, H. Yang, C. L. Richards-Zawacki, G. Zhang, R. Nielsen, Genomic takeover by transposable elements in the Strawberry poison frog. *Mol. Biol. Evol.* **35**, 2913–2927 (2014).
 12. A. M. M. Stuckert, M. Chouteau, M. McClure, T. M. LaPolice, T. Linderoth, R. Nielsen, K. Summers, M. D. MacManes, The genomics of mimicry: Gene expression throughout development provides insights into convergent and divergent phenotypes in a Müllerian mimicry system. *Mol. Ecol.* **30**, 4039–4061 (2021).
 13. J. E. Crawford, R. Nielsen, Detecting adaptive trait loci in nonmodel systems: divergence or admixture mapping? *Mol. Ecol.* **22**, 6131–6148 (2013).
 14. M. Liang, R. Nielsen, Understanding admixture fractions. bioRxiv 008078 [Preprint] (2014). <https://doi.org/10.1101/008078>
 15. D. J. Balding, R. A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci. Int.* **64**, 125–140 (1994).
 16. D. J. Balding, R. A. Nichols, A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica.* **96**, 3–12 (1995).
 17. L. Cal, P. Suarez-Bregua, I. Braasch, U. Irion, R. Kelsh, J. M. Cerdá-Reverter, J. Rotllant, Loss-of-function mutations in the melanocortin 1 receptor cause disruption of dorso-ventral countershading in teleost fish. *Pigment Cell Melanoma Res.* **32**, 817–828 (2019).
 18. N. I. Mundy, N. S. Badcock, T. Hart, K. Scribner, K. Janssen, N. J. Nadeau, Conserved Genetic Basis of a Quantitative Plumage Trait Involved in Mate Choice. *Science.* **303**, 1870–1873 (2004).
 19. M. W. Nachman, H. E. Hoekstra, S. L. D’Agostino, The genetic basis of adaptive melanism in pocket mice. *Proc. Natl. Acad. Sci.* **100**, 5268–5273 (2003).
 20. E. B. Rosenblum, H. E. Hoekstra, M. W. Nachman, Adaptive reptile color variation and the evolution of the *MC1R* gene. *Evolution.* **58**, 1794–1808 (2004).

21. B. J. Norris, V. A. Whan, A gene duplication affecting expression of the ovine *ASIP* gene is responsible for white and black sheep. *Genome Res.* **18**, 1282–1293 (2008).
22. B. B. Trigo, A. T. H. Utsunomiya, A. A. A. D. Fortunato, M. Milanesi, R. B. P. Torrecilha, H. Lamb, L. Nguyen, E. M. Ross, B. Hayes, R. C. M. Padula, T. S. Sussai, L. B. Zavarez, R. S. Cipriano, M. M. T. Caminhas, F. L. Lopes, C. Pelle, T. Leeb, D. Bannasch, D. Bickhart, T. P. L. Smith, T. S. Sonstegard, J. F. Garcia, Y. T. Utsunomiya, Variants at the *ASIP* locus contribute to coat color darkening in Nellore cattle. *Genet. Sel. Evol.* **53**, 40 (2021).
23. R. Guillot, R. M. Ceinos, R. Cal, J. Rotllant, J. M. Cerdá-Reverter, Transient Ectopic Overexpression of Agouti-Signalling Protein 1 (*Asip1*) Induces Pigment Anomalies in Flatfish. *PLoS ONE.* **7**, e48526 (2012).
24. R. A. Fisher. *Statistical Methods for Research Workers* (Oliver and Boyd, ed. 4, 1932).
25. J. T. Bagnara, P. J. Fernandez, R. Fujii, On the blue coloration of vertebrates. *Pigment Cell Res.* **20**, 14–26 (2007).
26. M. D. Shawkey, N. I. Morehouse, P. Vukusic, A protean palette: colour materials and mixing in birds and butterflies. *J. R. Soc. Interface.* **6**, (suppl. 2), 221–231 (2009).
27. K. M. Kozak, M. Joron, W. O. McMillan, C. D. Jiggins, Rampant Genome-Wide Admixture across the *Heliconius* Radiation. *Genome Biol. Evol.* **13**, evab099 (2021).
28. R. R. Hudson, M. Kreitman, M. Aguadé, A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics.* **116**, 153–159 (1987).
29. R. Márquez, T. P. Linderoth, D. Mejía-Vargas, R. Nielsen, A. Amézquita, M. R. Kronforst, Divergence, gene flow, and the origin of leapfrog geographic distributions: The history of colour pattern variation in *Phyllobates* poison-dart frogs. *Mol. Ecol.* **29**, 3702–3719 (2020).
30. A. Rodríguez, N. I. Mundy, R. Ibáñez, H. Pröhl, Being red, blue and green: the genetic basis of coloration differences in the strawberry poison frog (*Oophaga pumilio*). *BMC Genomics.* **21**, 301 (2020).
31. F. Almathen, H. Elbir, H. Bahbahani, J. Mwacharo, O. Hanotte, Polymorphisms in *MC1R* and *ASIP* Genes are Associated with Coat Color Variation in the Arabian Camel. *J. Hered.* **109**, 700–706 (2018).
32. L. Cal, P. Suarez-Bregua, J. M. Cerdá-Reverter, I. Braasch, J. Rotllant, Fish pigmentation and the melanocortin system. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **211**, 26–33 (2017).
33. H. Chen, J. Wang, J. Du, Z. Si, H. Yang, X. Xu, C. Wang, *ASIP* disruption via CRISPR/Cas9 system induces black patches dispersion in Oujiang color common carp. *Aquaculture.* **498**, 230–235 (2019).
34. D. Hepp, G. L. Gonçalves, G. R. P. Moreira, T. R. O. de Freitas, Epistatic Interaction of the Melanocortin 1 Receptor and Agouti Signaling Protein Genes Modulates Wool Color in the Brazilian Creole Sheep. *J. Hered.* **107**, 544–552 (2016).
35. M. G. Mills, L. B. Patterson, Not just black and white: Pigment pattern development and evolution in vertebrates. *Semin. Cell Dev. Biol.* **20**, 72–81 (2009).
36. N. J. Nadeau, F. Minvielle, S. Ito, M. Inoue-Murayama, D. Gourichon, S. A. Follett, T. Burke, N. I. Mundy, Characterization of Japanese Quail *yellow* as a Genomic Deletion Upstream of the Avian Homolog of the Mammalian *ASIP* (*agouti*) Gene. *Genetics.* **178**, 777–786 (2008).
37. A. Posso-Terranova, J. Á. Andrés, Diversification and convergence of aposematic phenotypes: truncated receptors and cellular arrangements mediate rapid evolution of

- coloration in harlequin poison frogs. *Evolution*. **71**, 2677–2692 (2017).
38. S. Dieck, L. Sanmartí-Vila, K. Langnaese, K. Richter, S. Kindler, A. Soyke, H. Wex, K.-H. Smalla, U. Kämpf, J.-T. Fränzer, M. Stumm, C. C. Garner, E. D. Gundelfinger, Bassoon, a Novel Zinc-finger CAG/Glutamine-repeat Protein Selectively Localized at the Active Zone of Presynaptic Nerve Terminals. *J. Cell Biol.* **142**, 499–509 (1998).
 39. O. Dick, S. tom Dieck, W. D. Altrock, J. Ammermüller, R. Weiler, C. C. Garner, E. D. Gundelfinger, J. H. Brandstätter, The Presynaptic Active Zone Protein Bassoon Is Essential for Photoreceptor Ribbon Synapse Formation in the Retina. *Neuron*. **37**, 775–786 (2003).
 40. P. Salis, T. Lorin, V. Lewis, C. Rey, A. Marcionetti, M. Escande, N. Roux, L. Besseau, N. Salamin, M. Sémon, D. Parichy, J. Volff, V. Laudet, Developmental and comparative transcriptomic identification of iridophore contribution to white barring in clownfish. *Pigment Cell Melanoma Res.* **32**, 391–402 (2019).
 41. E. P. Ahi, L. A. Lecaudey, A. Ziegelbecker, O. Steiner, R. Glabonjat, W. Goessler, V. Hois, C. Wagner, A. Lass, K. M. Sefc, Comparative transcriptomics reveals candidate carotenoid color genes in an East African cichlid fish. *BMC Genomics*. **21**, 54 (2020).
 42. J. von Lintig, Colors with Functions: Elucidating the Biochemical and Molecular Basis of Carotenoid Metabolism. *Annu. Rev. Nutr.* **30**, 35–56 (2010).
 43. P. Weber, R. E. Flores, M. F. Kiefer, M. Schupp, Retinol Saturase: More than the Name Suggests. *Trends Pharmacol. Sci.* **41**, 418–427 (2020).
 44. A. M. M. Stuckert, E. Moore, K. P. Coyle, I. Davison, M. D. MacManes, R. Roberts, K. Summers, Variation in pigmentation gene expression is associated with distinct aposematic color morphs in the poison frog *Dendrobates auratus*. *BMC Evol. Biol.* **19**, 85 (2019).
 45. H. Ando, Y. Niki, M. Ito, K. Akiyama, M. S. Matsui, D. B. Yarosh, M. Ichihashi, Melanosomes Are Transferred from Melanocytes to Keratinocytes through the Processes of Packaging, Release, Uptake, and Dispersion. *J. Invest. Dermatol.* **132**, 1222–1229 (2012).
 46. K. L. Sanders, A. Malhotra, R. S. Thorpe, Evidence for a Müllerian mimetic radiation in Asian pitvipers. *Proc. R. Soc. B Biol. Sci.* **273**, 1135–1141 (2006).
 47. J. P. Dumbacher, R. C. Fleischer, Phylogenetic evidence for colour pattern convergence in toxic pitohuis: Müllerian mimicry in birds? *Proc. R. Soc. Lond. B Biol. Sci.* **268**, 1971–1976 (2001).
 48. M. A. Alexandrou, C. Oliveira, M. Maillard, R. A. R. McGill, J. Newton, S. Creer, M. I. Taylor, Competition and phylogeny determine community structure in Müllerian co-mimics. *Nature*. **469**, 84–88 (2011).
 49. P. Williams, The distribution of bumblebee colour patterns worldwide: possible significance for thermoregulation, crypsis, and warning mimicry. *Biol. J. Linn. Soc.* **92**, 97–118 (2007).
 50. J. S. Wilson, K. A. Williams, M. L. Forister, C. D. von Dohlen, J. P. Pitts, Repeated evolution in overlapping mimicry rings among North American velvet ants. *Nat. Commun.* **3**, 1272 (2012).
 51. C. P. Muñoz-Ramírez, P.-P. Bitton, S. M. Doucet, L. L. Knowles, Mimics here and there, but not everywhere: Müllerian mimicry in *Ceroglossus* ground beetles? *Biol. Lett.* **12**, 20160429 (2016).
 52. T. G. Aubier, M. Elias, Positive and negative interactions jointly determine the structure of Müllerian mimetic communities. *Oikos*. **129**, 983–997 (2020).
 53. M. Elias, Z. Gompert, C. Jiggins, K. Willmott, Mutualistic Interactions Drive Ecological Niche Convergence in a Diverse Butterfly Community. *PLoS Biol.* **6**, e300 (2008).

54. J. J. Lewis, R. C. Geltman, P. C. Pollak, K. E. Rondem, S. M. Van Belleghem, M. J. Hubisz, P. R. Munn, L. Zhang, C. Benson, A. Mazo-Vargas, C. G. Danko, B. A. Counterman, R. Papa, R. D. Reed, Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. *Proc. Natl. Acad. Sci.* **116**, 24174–24183 (2019).
55. M. Joron, L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, P. A. Wilkinson, C. Salazar, C. Davidson, R. Clark, M. A. Quail, H. Beasley, R. Glithero, C. Lloyd, S. Sims, M. C. Jones, J. Rogers, C. D. Jiggins, R. H. ffrench-Constant, Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*. **477**, 203–206 (2011).
56. J. Elkin, A. Martin, V. Courtier-Orgogozo, M. E. Santos, Analysis of the genetic loci of pigment pattern evolution in vertebrates. *Biol Rev.* doi:10.1111/brv.12952 (2023).
57. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics*. **195**, 693–702 (2013).
58. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*. **9**, 671–675 (2012).
59. R Core Team, R: A language and environment for statistical computing, version 3.6.3 (2020); <https://www.R-project.org/>.
60. S. Kimball, P. Mattis, GIMP Development Team, GNU Image Manipulation Program, version 2.8 (2012); <https://www.gimp.org/>.
61. E. C. Carter, J. D. Schanda, R. Hirschler, S. Jost, M. R. Luo, M. Melgosa, Y. Ohno, M. R. Pointer, D. C. Rich, F. Viénot, L. Whitehead, J. H. Wold, “CIE 015:2018 Colorimetry, 4th Edition” (International Commission on Illumination, 2018); doi:10.25039/TR.015.2018.
62. R. Maia, C. M. Eliason, P.-P. Bitton, S. M. Doucet, M. D. Shawkey, pavo: an R package for the analysis, visualization and organization of spectral data. *Methods Ecol. Evol.* **4**, 906–913 (2013).
63. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
64. S. Singhal, *De novo* transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set. *Mol. Ecol. Resour.* **13**, 403–416 (2013).
65. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics*. **10**, 421 (2009).
66. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2010).
67. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
68. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
69. A. F. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0 (2013); <http://www.repeatmasker.org>.

70. K. Bi, D. Vanderpool, S. Singhal, T. Linderoth, C. Moritz, J. M. Good, Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*. **13**, 403 (2012).
71. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* doi:10.1101/pdb.prot5448 (2010).
72. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).
73. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. **17**, 10–12 (2011).
74. A. Morgulis, E. M. Gertz, A. A. Schäffer, R. Agarwala, A Fast and Symmetric DUST Implementation to Mask Low-Complexity DNA Sequences. *J. Comput. Biol.* **13**, 1028–1040 (2006).
75. J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*. **30**, 614–620 (2014).
76. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, P. A. Pevzner, SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
77. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **28**, 3150–3152 (2012).
78. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. **22**, 1658–1659 (2006).
79. X. Huang, A. Madan, CAP3: A DNA Sequence Assembly Program. *Genome Res.* **9**, 868–877 (1999).
80. Novocraft Technologies Sdn Bhd, Novoalign, version 3.04.06 (2016); <https://www.novocraft.com/products/novoalign>.
81. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
82. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN] (2013).
83. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *GigaScience*. **10**, giab008 (2021).
84. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. **15**, 356 (2014).
85. M. Fumagalli, F. G. Vieira, T. Linderoth, R. Nielsen, ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*. **30**, 1486–1487 (2014).
86. J. Reynolds, B. S. Weir, C. C. Cockerham, Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. *Genetics*. **105**, 767–779 (1983).
87. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Association Testing for Next-Generation Sequencing Data Using Score Statistics. *Genet. Epidemiol.* **36**, 430–437 (2012).
88. Geneious, version 4.8.5 (2010); <https://www.geneious.com>.
89. G. R. Abecasis, S. S. Cherny, W. O. Cookson, L. R. Cardon, Merlin—rapid analysis of

- dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
90. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Royal Stat. Soc. Ser. B.* **57**, 289–300 (1995).
 91. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29** (2001).
 92. F. G. Vieira, F. Lassalle, T. S. Korneliussen, M. Fumagalli, Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biol. J. Linn. Soc.* **117**, 139–149 (2016).
 93. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).
 94. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
 95. S. Soraggi, C. Wiuf, A. Albrechtsen, Powerful Inference with the D-Statistic on Low-Coverage Whole-Genome Data. *G3 Genes|Genomes|Genetics.* **8**, 551–566 (2018).
 96. X. Cheng, M. DeGiorgio, Detection of Shared Balancing Selection in the Absence of Trans-Species Polymorphism. *Mol. Biol. Evol.* **36**, 177–199 (2019).
 97. F. Yates, Contingency Tables Involving Small Numbers and the χ^2 Test. *Suppl. J. R. Stat. Soc.* **1**, 217–235 (1934).
 98. D. Clayton, snpStats: SnpMatrix and XSnpmatrix classes and methods, version 1.46.0 (2022); doi:10.18129/B9.bioc.snpStats.
 99. J. Hey, HKA (2010); <https://bio.cst.temple.edu/~tuf29449/software>.

Acknowledgements: The authors wish to thank Pablo Venegas of CORBIDI for assistance with research, collection and export permits in Peru, and Manuel Panaifo of Chazuta, San Martin Province for assistance with fieldwork in Peru. We would also like to thank Lydia Smith at the Evolutionary Genomics Lab and Shana McDevitt at the QB3 Genomics core at UC Berkeley for their support in generating genomic data. Sarah Fitzpatrick’s lab and Chris Kozakiewicz provided helpful discussion in revising the manuscript. Tissue exports were authorized under Contrato de Acceso Marco a Recursos Geneticos 0009–2013-MINAGRI-DGFFS/DGEFFS, with CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) permits number 003302 and 001718. Permission for the protocols used for research on vertebrate animals in this project was obtained from the ECU Institutional Animal Care and Use Committee (IACUC), under AUPs 281D and 376. Research Permits for collection and export of tissues were obtained from the Peruvian Servicio Forestal y de Fauna Silvestre (SERFOR): Authorizations 050–2006-INRENA-IFFS-DCB, 067–2007-INRENA-IFFS-DCB, 005–2008-INRENA-IFFS-DCB, Resolución Directoral No. 137-2009-AG-DGFFS-DGEFFS, No. 0266-2011-AG-DGFFS-DGEFFS, No. 0402-2013-MINAGRI-DGFFS/DGEFFS and No. 232-2016-SERFOR/DGGSPFFS.

Funding:

National Science Foundation, in the form of a collaborative grant to K. Summers, R. Nielsen and M. MacManes from the Division of Environmental Biology, Evolutionary Processes: DEB1655336 (KS), DEB1655191 (RN), DEB1655585 (MM)

Author Contributions:

Conceptualization: TL, RN, KS

Methodology: TL, RN, KB

Formal Analysis: TL, DA, EW, AK, JLR, KB, MDM

Investigation: TL, EW, ET, AS, AK, NG, JC, MDM, KS

Resources: ET, AS, KS

Writing - original draft: TL, RN, KS, DA, EW

Writing - review and editing: TL, RN, KS, EW, ET, AS, AK, NG, JLR, JC, MDM

Visualization: TL

Supervision: RN, KS, TL

Project administration: RN, KS, TL

Funding acquisition: RN, KS, MDM

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All genetic sequencing and phenotypic data will be deposited to the Sequence Read Archive and Dryad. All code used in this study is available at https://github.com/tplinderoth/poison_frog_color/tree/main/R_imitator_color.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S10

Tables S1 to S12

References (58-99)

Supplementary Materials for

Genetic basis of aposematic coloration in a mimetic radiation of poison frogs

Tyler Linderoth, Diana Aguilar-Gómez, Emily White, Evan Twomey, Adam Stuckert, Ke Bi, Amy Ko, Natalie Graham, Joana Rocha, Jason Chang, Matthew MacManes, Kyle Summers, Rasmus Nielsen

Corresponding author: lindero1@msu.edu, summersk@ecu.edu, rasmus_nielsen@berkeley.edu

The PDF file includes:

- Materials and Methods
- Supplementary Text
- Figs. S1 to S10
- Tables S1 to S12
- References

Materials and Methods

Ranitomeya imitator and model species samples

We sampled *R. imitator*, *R. variabilis* and *R. summersi* from sites throughout the San Martin province of Peru (Fig. 1) throughout the morning and early afternoon in 2010, 2012, 2013, and 2014. We hand-caught frogs as we encountered them out foraging during the day or resting in plant axils. Frogs were placed individually into film canisters or 15 mL Falcon tubes and transported back to a field-based laboratory where we collected phenotypic data and toe clips, which were stored in 96% ethanol. On the following day, frogs were released to the location from which they were caught. We collected samples from each location in one to three consecutive days. The original *R. imitator* parentals used to generate the multigenerational pedigree were obtained from a commercial breeder (Understory Enterprises LLC) that breeds different morphs of *R. imitator* from known localities in Peru.

Phenotyping

We quantified the color and pattern of *R. imitator* individuals from digital photographs taken with a Nikon D7000 camera using a Nikkor 85mm macro lens. Images were converted from RAW to the highest quality JPEG format using Nikon ViewNX2 software. Prior to phenotyping, we measured the anteroposterior length (running from the tip of the pelvis to the rostrum) of each frog as well as its midline orientation to the horizontal axis using imageJ v. 1.53c software (58). These measurements were used in R v. 3.6.3 (59) to standardize the size and

orientation of the frog subjects across all of the photos. The standardized photos were imported into GIMP v. 2.8.22 (60), which we used to manually generate black masks for black portions of the frogs to reduce noise from glare and debris. We also used GIMP to isolate subjects from backgrounds and generate binary masks for restricting analyses to regions of interest (ROI). The preprocessed images were exported in PNG format for input into our phenotyping scripts. All scripts used for image processing and phenotyping are available at https://github.com/tplinderoth/image_phenotype.

We measured the proportion of the dorsal body surface (limbs excluded) that was black, denoted PB. This was calculated simply as the fraction of black pixels among all non-transparent pixels in the ROI. We also measured four aspects of color for the non-black areas in the CIELAB ($L^*a^*b^*$) color space (61): dorsal body color, DC_a , and hindlimb color, LC_a , along the red-green color axis (a^*), and dorsal body color, DC_L , and hindlimb color, LC_L , along the lightness axis (L^*). To quantify color from PNG images we first determined the average red, green, and blue values in the RGB color space over all non-transparent pixels in the ROI. Specifically, for a raster image ROI with d non-transparent pixels having color values $(\psi_{v,1}, \dots, \psi_{v,d})$ for color channels $v \in \{\text{red}, \text{green}, \text{blue}\}$, the average for each RGB channel, $\bar{\psi}_v$, was calculated as

$$\bar{\psi}_v = \left(\frac{1}{d} \sum_{k=1}^d \psi_{v,k}^2 \right)$$

These average RGB values were then converted to $L^*a^*b^*$ color space using CIE Standard Illuminant D65 as the reference white for both color spaces.

Pattern was measured in terms of the degree to which black streaks run parallel to the midline axis along the body, denoted PT_{rot} and referred to as ‘pattern orientation’ throughout the text. To elaborate on how PT_{rot} is calculated from a raster image we impose a binary classification scheme for pixel color, which in our case is black and non-black. Let γ_r and γ_c denote the number of transitions between black and non-black pixels in a straight line along the entire horizontal axis and vertical axis of a raster image, respectively. Note that throughout we use the subscript r to refer to the horizontal axis of a raster image (the rows) and c to refer to the vertical axis (the columns). Pattern track length, τ , is defined as the number of directly adjacent black pixels along a straight line. Let τ_r^{max} denote the maximum τ in a set of tracks for a given raster image row, and similarly, τ_c^{max} denotes the maximum τ among the set of tracks for a given column. The effective segment length, ϕ , is the count of all pixels in a given row or column to be analyzed and so is bounded from above by the image's width or height in pixels, respectively. For a raster image of width, w , and height, h

$$\eta = \sum_{i=1}^h \gamma_{r,i} + \sum_{j=1}^w \gamma_{c,j}$$

$$\xi = \frac{1}{h} \sum_{i=1}^h \frac{\tau_{r,i}^{max}}{\phi_i} - \frac{1}{w} \sum_{j=1}^w \frac{\tau_{c,j}^{max}}{\phi_j}$$

$$PT_{rot} = \frac{1}{\eta} \left(\sum_{j=1}^w \gamma_{c,j} - \sum_{i=1}^h \gamma_{r,i} + \xi \right)$$

As PT_{rot} increases in the positive direction it corresponds to more stripedness, while increasing magnitude in the negative direction indicates more bandedness. A PT_{rot} value of 0 corresponds to a frog that is perfectly intermediate between striped and banded. We checked the effectiveness of our methods for quantifying color and pattern by ensuring that image orderings using the rank values for each of the aforementioned measures were sensible (fig. S1-3).

We used an Ocean Optics USB spectrometer and the associated Ocean Optics software to collect spectral reflectance data for captive-bred frogs. We collected 10 data points (8 dorsal, 2 hindlimb, 3 ventral, see fig. S4) for each frog. We calibrated the spectrometer with a pure white standard between each frog to maintain accuracy of the measurements. We used the `pavo` package (62) in R to visualize, clean, and analyze the resulting spectral reflectance data. We first subdivided the data into groups with reference to body region, as shown in fig. S4. For each region (body and legs) on each frog, we aggregated the data points by taking the mean and set negative values to zero. Visualization showed noise at both ends of the spectral curve, which we controlled for by truncating the spectra to a wavelength range of 450 - 950 nm (previously 400 - 1041 nm). After processing the spectral reflectance data, we visualized them again, this time with the average and standard deviation plotted. We quantified mean brightness (B2), and the relative contribution of green, yellow, and red wavelengths to total brightness (S1G, S1Y, and S1R respectively). The code for processing and analyzing spectral reflectance data is available at <https://github.com/ksummers26/pavo-code>.

Exome capture design

The large size and highly duplicated composition of the *R. imitator* genome (12) poses an obstacle to population genomic investigations into this species. We surmounted this hurdle by capturing and sequencing approximately 28 Mb of the *R. imitator* exome. This allowed us to obtain genomic-scale, short-read data for 124 individuals at coding regions of the genome potentially relevant to color phenotypes that we could reliably assemble and map. At the time of our study, there were no existing dendrobatid genomic resources to guide capture probe design, so we designed probes using *R. imitator* transcriptomes from different tissues and developmental stages.

In order to obtain transcriptomes for capture probe design, we first isolated mRNA from five tissues belonging to two young adult, Tarapoto-morph *R. imitator* (brain plus eyes, skin from the dorsal trunk, nape plus dorsal head, ventral jaw, and rear leg regions) as well as mRNA from entire tadpole skins sampled at weeks one (n = 1), two (n = 3), four (n = 3), and five (n = 3) of development (fig. S5). We prepared cDNA libraries with unique barcodes for each individual and tissue separately using Illumina TruSeq stranded total RNA kits. Various developmental stages were sequenced since we had no *a priori* knowledge regarding when color and pattern-related genes are expressed, though color becomes visible at around week five (fig. S5). Two library pools, one consisting of libraries from the young adults and another from all five developmental series, were each sequenced on two lanes of the Illumina HiSeq 2000 using 100 bp paired-end reads.

We assembled the sequencing reads from each cDNA library using Trinity (63), yielding 273,039 transcripts. Following the procedure of (64), we assigned Ensembl gene identifiers to 12,305 unique transcripts with a maximum BLASTX (65) e-value of 1e-10 and at least 50% sequence similarity to *Xenopus tropicalis* proteins downloaded from Ensembl (transcriptome annotation procedures are implemented in the script available at <https://github.com/CGRL-QB3-UCBerkeley/MarkerDevelopmentPylogenomics/blob/master/5-Annotation>). We also identified transcripts with unusually high differences in mapped read counts among the different young adult tissues and between the different developmental stages using EdgeR (66) that could suggest differential expression. Additionally, we considered any transcripts with TPM values, calculated with Kallisto (67), that were within the 95th percentile across all stages of development to be constitutively expressed genes potentially crucial for development. TransDecoder v 2.1.0 (68) was run on the unannotated transcripts with putative differential or constitutive expression patterns in order to extract those with open reading frames of at least 100 bp, from which we further trimmed away any non-gene-like portions. These trimmed, unannotated transcripts with interesting expression patterns along with all annotated transcripts were used as template sequences for capture probe design.

We applied a suite of filters to the set of transcripts used for probe design in order to optimize capture performance. This filtering entailed first performing a reciprocal blast of the transcripts and retaining one isoform per gene. For the retained transcripts, we trimmed each of the 3' and 5' untranslated regions (UTRs) to 500 bp in order to avoid designing probes on exceptionally long UTRs that could potentially represent assembly artifacts. We then masked out any known vertebrate repetitive elements and low-complexity regions from our target using repeatMasker v. 4.0.5 (69), since such elements could promote off-target capture. We retained only transcript regions consisting of at least 80 contiguous unmasked bases with GC content in the range of 40-70%, outside of which capture is ineffective (70). Lastly, since mitochondrial genes are unlikely to affect color phenotypes, we excluded all mitochondrial transcripts except for cytochrome B, which was kept to ensure some haploid representation for downstream quality control. Our final, filtered capture target included 13,265 genes spanning 28,281,490 bp. Among these genes,

10,904 were annotated with Ensembl IDs and accounted for 24,530,337 bp, while the remaining 2,361 unannotated genes with intriguing expression patterns totalled 3,751,153 bp in length. Custom capture probes were designed from this target sequence and synthesized as a NimbleGen SeqCap EZ Developer Library kit.

Genomic library preparation, exome capture, and sequencing

We extracted DNA from single toe clips for each *R. imitator* sample using either Qiagen DNeasy Blood and Tissue kits following the manufacturer's protocols or a standard salt precipitation. For the salt precipitation method we first incubated each toe clip in 500 μ L cell lysis buffer (1 mM Tris pH 8, 100 mM NaCl, 10 mM EDTA pH 8, 0.5% SDS) and 10 μ L 20 mg/mL proteinase K at 55 $^{\circ}$ C for 24 hours. After the tissues were completely digested we added 3 μ L of 10 mg/mL RNase A to the samples and incubated them for an additional 30 minutes at 37 $^{\circ}$ C. Following this RNA digestion the samples were placed in a freezer at -20 $^{\circ}$ C for five minutes. We then added 200 μ L of 5 M NaCl to each sample followed by vortexing for 20 seconds, and then placed them back in the freezer for another five minutes. Samples were then centrifuged at 11,000 RPM for eight minutes to pellet the proteins. The supernatant containing DNA was then transferred to new tubes, leaving behind the protein pellet, and 600 μ L of 100% isopropanol was added to the tubes containing DNA, followed by 50 gentle inversions of the tubes. The samples were then placed in the freezer for one hour and then centrifuged at 10,000 RPM for 10 minutes to pellet the DNA. After pouring off the supernatant, we washed the DNA with 600 μ L of fresh 70% ethanol. The tubes were inverted to dislodge the pellet in the 70% ethanol, and then centrifuged for eight minutes at 10,000 RPM. After centrifugation, the supernatant was poured off, leaving the DNA pellet behind, which was allowed to dry for at least 12 hours or until completely dry. Lastly, the samples were resuspended in 50 μ L TE buffer.

For samples that had been extracted using the Qiagen kits, we performed an additional post-extraction RNase A digestion to ensure complete removal of RNA. Specifically, three μ L of 10 mg/mL RNase A was added to DNA suspended in Qiagen AE buffer and incubated at 37 $^{\circ}$ C for 30 minutes. The DNA was then recovered by adding 0.1x volume of 3M sodium acetate (pH 5.2) and two volumes of 100% isopropanol to the samples, followed by gently inverting the samples 50 times. The samples were then placed into a -20 $^{\circ}$ C freezer for 15 minutes. After removing the samples from the freezer, we centrifuged them at 10,000 RPM for 12 minutes at room temperature to pellet the DNA. We then discarded the supernatant and washed the DNA pellet by adding 600 μ L of fresh 70% ethanol to each sample followed by inverting the tubes to dislodge the pellet. The samples were then centrifuged at 10,000 RPM for 12 minutes. Lastly, we poured off the supernatant, which we allowed to dry for at least 12 hours. After the samples were completely dry, we eluted them in 50 μ L TE buffer.

We prepared genomic libraries for capture following the protocol of Meyer & Kircher (71). We started the library preparation with 0.92-1.4 μ g DNA per sample as measured by a Nanodrop fluorospectrometer, save for one sample (CH-14-5) for which we used 0.4 μ g DNA. The DNA

for each sample was sheared using a Diagenode Bioruptor to an average fragment size of 250 bp for striped and banded samples and 300 bp for the admixed samples. For striped and banded samples this was achieved by shearing samples at the high Bioruptor setting for a total of seven minutes using cycles defined by 30 seconds of shearing followed by a 30-second pause (each sample was in the Bioruptor for a total of 13 minutes). The admixed samples were sheared at the medium Bioruptor setting for a total of four minutes using the same 30 seconds on / 30 seconds pause cycle scheme (each sample was in the Bioruptor for a total of eight minutes). Following each enzymatic reaction up until the indexing PCR step, all samples were cleaned using 1.6x volume Sera-Mag bead solution. A sample-specific, seven-nt barcode was incorporated into the library sequences for each sample via indexing PCR using 4 μ L of template library DNA and 10 PCR cycles under the reaction conditions specified in (71). We performed three independent indexing PCR reactions for the striped and banded individuals followed by bead purification using 1.3x volume Sera-Mag beads. For the admixed individuals we performed five independent PCR reactions, each of which were purified using 0.8x volume Sera-Mag beads. All of the PCR products for a respective individual were pooled across each of the independent PCR reactions.

Barcoded libraries for striped and banded samples were pooled equimolarly in batches of 22 individuals to generate three multiplexed pools for capture. All 58 admixed individual libraries were equimolarly pooled to form one multiplexed pool for capture. We isolated the targeted sequence from these capture pools using our custom *R. imitator* capture kit following NimbleGen's protocols with slight modifications. Given the large genome size of *R. imitator*, we used approximately 2-3 times the amount of input DNA for hybridization compared to the 1 μ g called for by the NimbleGen protocol in order to preserve the complexity of the captured libraries: The three banded/striped captures used 2.522, 2.388, 2.663 μ g of DNA and the admixed capture used 3 μ g of DNA as measured by a Qubit fluorometer. We also doubled and tripled the amount of barcode blockers and COT-1 DNA used in the banded/striped and admixed hybridization reactions, respectively, relative to the amounts specified in the NimbleGen protocols. For COT-1 blocking DNA we used a cocktail of chicken Hybloc, human COT-1, and mouse COT-1 combined in equal amounts. All libraries were hybridized with the capture probes for 75 hours. The captured libraries were LM-PCR amplified in three separate reactions using 20 μ L of template DNA solution. The number of PCR cycles ranged from 11-14 cycles in order to yield DNA concentrations between 20-25 ng/ μ L per reaction according to Nanodrop measurements. PCR products from the three reactions for each capture were pooled in equal amounts based on Qubit fluorometer measurements. Enrichment efficiency for the four capture experiments was assessed using qPCR with negative (off-target) and positive (targeted) control loci. The three banded/striped captures were pooled equally based on qPCR measurements for sequencing. The banded/striped multiplexed library and the admixed multiplexed library were sequenced on the Illumina HiSeq 4000 using 100bp and 150 bp paired-end reads, respectively.

De novo exome assembly

We applied quality controls to the sequencing reads using the program readCleaner (<https://>

github.com/tplinderoth/ngsQC/tree/master/readCleaner) prior to assembly and mapping in the following order: 1) remove reads derived from PCR duplicates, 2) trim adapter sequences from the ends of reads, 3) trim low quality bases from the ends of reads in a two-step approach; first, trim bases with average Phred-scaled quality in 4 bp sliding windows below 20 from the ends of reads, and secondly, apply the BWA (72) algorithm implemented in cutadapt (73) using a minimum quality threshold of 20, 4) remove low complexity reads with a DUST score (74) above four, 5) merge read pairs that overlap by at least six base pairs and that have an observed expected alignment score (75) p-value less than 0.01, 6) remove reads that map to the human GRCh38 or *Escherichia coli* genome assemblies (NCBI GenBank accessions U00096, AE000111-AE000510), 7) remove reads shorter than 36 bp and/or have at least 50% of their sequence comprised of Ns.

We assembled quality-controlled reads from six *R. imitator* individuals (three banded morph: ET13012, ET13017, ET13018, and three striped morph: ET13029, ET13032, ET13033) separately with SPAdes v 3.10.0 (76) under default parameters. Contigs with at least 80% sequence identity to targeted *R. imitator* transcripts, determined using blastn, were clustered using CD-HIT (77, 78) and merged with CAP3 (79). All resulting contigs belonging to the same gene were concatenated together in syntenic order with interspersed runs of 39 Ns to prevent reads from mapping across the gaps. Scripts used for assembly are available at <https://github.com/CGRL-QB3-UCBerkeley/seqCapture>.

Variant discovery

We mapped *R. imitator* reads with at least 30 high-quality bases (NovoAlign option -l) to the *de novo* exome assembly using NovoAlign v 3.04.06 (80), with the highest acceptable alignment score for the best alignment set to 150. Only reads with a minimum Phred-scaled mapping quality of 20 were retained. The NovoAlign average insert size and standard deviation mapping parameters were set according to values determined from bioanalyzer traces for the captured genomic libraries. Banded and striped individuals were mapped using an average insert size of 220 bp and a standard deviation of 58, while for admixed individuals the insert size was set to 271 bp with a standard deviation of 63. We used SAMtools v 1.11 (81) to merge unpaired and paired alignments from the same individual. We mapped RNAseq reads from the model species, *R. summersi* and *R. variabilis*, to the *R. imitator* exome assembly using *bwa mem* v. 0.7.17 (82) with default parameter settings. For the models, we used SAMtools view to retain only primary alignments for which the reads mapped in proper pairs while excluding all reads that were PCR or optical duplicates or which failed quality checks (-F 1804).

Prior to identifying variants, we produced quality control masks denoting genomic regions where variant calling would be reliable. This was achieved by first generating an all-sites (variable and monomorphic sites) VCF containing all *R. imitator* and model species individuals, which included coverage and map quality information at every site, as well as p-values for tests of various biases between reference and alternate alleles called using bcftools v 1.10.2 (83). Given

the high levels of duplication throughout the *R. imitator* genome we were particularly cautious to mask regions refractory to short reading mapping. We performed this masking using ngsParalog (<https://github.com/tplinderoth/ngsParalog>), which was ran for the banded, striped, and admixed *R. imitator* populations separately. Specifically, the ngsParalog calcLR function calculates a likelihood ratio (LR) statistic for testing whether within-individual read proportions and sample-wide genotype likelihoods at a site are consistent with reads sourced from multiple loci. This enabled identification of chimeric regions in the exome assembly. We annotated the all-sites VCF with these likelihood ratios, and used bcftools to mask out multiallelic sites and indels, as well as sites with an RMS mapping quality score of less than 30, total site coverage greater than the 99.8 exome-wide percentile, p-values for either strand, base quality, map quality, or end-distance bias that were less than 1e-100, p-values for excess heterozygosity from called genotypes less than 1e-3, and ngsParalog LR's greater than 500 (approximately the 98th genome-wide percentile) in any of the *R. imitator* populations.

In addition to the above filters we required at least 15 individuals from each of the striped, banded, and admixed *R. imitator* populations to be covered by at least three reads at a site in order for the site to pass masking. We also imposed three levels of coverage requirements for the model species according to their relevance to particular analyses in order to maximize the amount of quality data while minimizing bias (fig. S6): 1) no coverage requirement for GWAS, which were performed among *R. imitator* only, 2) one individual of either model species with data for comparative population genetic analyses (e.g. PCA) and tests for introgression and selection, 3) all individuals from both model species with data for phylogenetic inference. We called SNPs based on a maximum p-value of 1e-6 for the likelihood ratio test of the minor allele frequency being greater than zero among all *R. imitator* and model species at unmasked sites using ANGSD. Estimation of low frequency variants was particularly sensitive to missing data (fig S7), therefore we masked variable sites with minor allele frequency less than 2% in *R. imitator* (five minor alleles) from all analyses. The distribution of allele frequencies estimated with ANGSD was robust to missing data at frequencies above this cutoff.

Population genetic characterization

We estimated levels of genetic diversity within *R. imitator* populations in terms of the number of segregating sites, S , Watterson's estimator, $\theta_{\text{Watterson}}$, and nucleotide diversity, π , in ANGSD based on the genotype-likelihoods of individuals in our sample. The latter two diversity estimates were estimated using an empirical Bayes procedure that uses each population's respective exome-wide, folded site frequency spectrum (SFS, also calculated with ANGSD) as a prior (fig. S6).

Genomic relationships among *Ranitomeya* individuals were assessed with principal components analysis (PCA). We first calculated the genetic covariance matrix among all individuals with ngsCovar (85) based on individual's genotype posterior probabilities under a Hardy-Weinberg allele frequency prior in ANGSD. We decomposed this matrix using the eigen() function in R v.

3.6.3 to obtain the eigenvectors and their corresponding eigenvalues for the PCA. We also repeated this analysis on *R. imitator* only. We quantified pairwise genetic divergence between the different *R. imitator* populations using Reynold's F_{ST} estimator (86) with ANGSD. As a prior on the distribution of joint allele frequencies for calculating F_{ST} , we used the unfolded joint SFS between the respective pair of populations considering the reference base as ancestral (mispolarization with respect to ancestral and derived allelic state does not matter for this calculation). Lastly, we estimated the proportions of striped and banded *R. imitator* ancestry of individuals sampled from the introgression zone using NGSadmix (57). We ran NGSadmix on genotype likelihoods from ANGSD for all striped, banded, and admixed *R. imitator* individuals, specifying 2-5 ancestral populations (K parameter) (fig. S8).

Divergence mapping

We expect to see increased divergence between the striped and banded morphs of *R. imitator* at loci that contribute to the color and pattern differences between them. However, we expect the two morphs to be somewhat differentiated even at random loci. In order to identify loci that are more differentiated than expected given the genome-wide distribution of allele frequency differences, we used the test statistic $\delta = |\hat{p}_1 - \hat{p}_2|$, where \hat{p}_1 and \hat{p}_2 are the estimated allele frequencies in the banded and striped population, respectively. To assign p-values we simulated data of the joint allele frequencies in the two populations using the Balding-Nichols model (15, 16). Given an ancestral allele frequency, p_0 , the allele frequencies under the Balding-Nichols model in two derived populations with genome-wide divergence expressed in terms of F_{ST} , are two independent draws from the beta distribution $\beta(p_0(1-F_{ST}), (1-p_0)(1-F_{ST})/F_{ST})$. Using two draws from this distribution, (p_1, p_2) , we modeled allele sampling based on these true frequencies with a binomial distribution: $\hat{p}_1 \sim B(2n_1, p_1)$ and $\hat{p}_2 \sim B(2n_2, p_2)$, where n_1 and n_2 are the diploid sample sizes from populations 1 and 2 respectively. The approach for approximating p-values for observed allele frequency differences, δ_{obs} , based on these expected allele frequency differences under the Balding-Nichols model is outlined below:

1) We simulated distributions of δ under the Balding-Nichols model in bins of different ancestral allele frequencies (p_0). We used bin widths for p_0 equal to $1/(2(n_1+n_2)) = 0.007575758$ for our sample sizes. Each distribution of δ was generated by sampling p_0 uniformly in the range of its corresponding bin one million times. Note that we only need to consider allele frequencies up to 0.5 because distributions of δ are symmetric for higher frequencies, resulting in a total of n_1+n_2 bins. We also discretized the distributions of δ into k bins of width $\arg\max(1/2n_1, 1/2n_2) = 0.01515152$ for our sample sizes and defined $P(\delta_r | p_0 \text{ bin } i)$ as the proportion of δ values in the simulations with values of p_0 in p_0 -bin i that fell in δ -bin r , $r \in \{1, \dots, k\}$.

2) We approximated the distribution of p_0 , i.e. the probability of p_0 falling in the i th bin, empirically as

$$\alpha_i = S^{-1} \sum_{j=1}^S I_i(\widehat{p}_{0j}), \quad i=1,2,\dots,n_1 + n_2$$

where $I_i(\widehat{p}_{0j})$ is an indicator of \widehat{p}_{0j} falling in the i th bin, \widehat{p}_{0j} is the value of $(n_1\widehat{p}_1 + n_2\widehat{p}_2)/(n_1 + n_2)$ calculated for the j th SNP, and S is the total number of sites.

3) Denote an observed allele frequency difference falling in δ -bin j by $\delta_{obs,j}$. We then approximated the p-values as

$$P(\delta \geq \delta_{obs,j}) = \sum_{r=j}^k \sum_{i=1}^{\binom{n_1+n_2}{k}} P(\delta_r | p_0 \text{ bin } i) \alpha_i$$

While this simple model may not adequately capture all of the intricacies of a more parametric demographic model, it provides a convenient statistical framework for identifying divergent loci. This procedure for approximating p-values is implemented in <https://github.com/tplinderroth/PopGenomicsTools/blob/master/betaAFOutlier.R>.

Our estimated genome-wide F_{ST} in the Balding-Nichols model was obtained using ANGSD. Observed allele frequency differences between the banded and striped morphs were calculated using allele frequencies estimated for each of the respective populations with ANGSD -doMaf 1 and setting the major allele to the reference allele with -doMajorMinor 4. P-values were approximated for the observed allele frequency differences using the procedure outlined above, which we used to test for significant δ s at a 0.05 level after a Bonferroni-adjustment for the number of tested SNPs.

Association mapping in the admixed population

As a follow-up on the candidate genes identified from the divergence mapping, we performed a second, independent genome-wide scan among admixed individuals of *R. imitator*. We used a general linear modeling (GLM) approach to quantify the effect of genotypes on continuously measured phenotypes while accounting for covariates (87). For each SNP, we estimated the genotypic effect on pattern orientation (PT_{rot}), dorsal body color (DC_a), and hindlimb color (LC_a) (see ‘‘Phenotyping’’ of Materials and Methods) based on the following models with N denoting a normal distribution with variance σ^2

$$PT_{rot,i} \sim N(\alpha_0 + \alpha_1 m_i + \alpha_2 PB_i + \beta g_i, \sigma^2),$$

$$DC_{a,i} \sim N(\alpha_0 + \alpha_1 m_i + \alpha_2 DC_{L,i} + \beta g_i, \sigma^2),$$

$$LC_{a,i} \sim N(\alpha_0 + \alpha_1 m_i + \alpha_2 LC_{L,i} + \beta g_i, \sigma^2)$$

The genotype of individual i , $g_i \in \{0,1,2\}$, can have 0, 1, or 2 copies of the minor allele. The admixture proportion, m_i , is the proportion of banded ancestry for individual i estimated with NGSadmix, and is equal to 1 – striped ancestry proportion. This covariate controls for confounded associations from differences in genome-wide levels of striped and banded ancestry. DC_L and LC_L are lightness values quantified from digital images in CIELAB color space to control for lighting effects. The GLMs for color quantified along the red-green CEILAB axis (a^*) included lightness as a covariate in order to account for non-standardized lighting conditions across frog photos. We compared twice the log likelihood ratio of a genetic association, which is the ratio of the log likelihood at the maximum likelihood estimate of β and the likelihood when $\beta=0$, to a χ_1^2 -distribution to obtain p-values for each SNP.

We combined the p-values for each SNP from the color-morph divergence test, p_d , with the p-values from the GLM-based tests of association in the admixed population, p_a , using Fisher’s method (24). This test statistic is defined as $X_i^2 = -2 \left[\log(p_{d,i}) + \log(p_{a,i}) \right]$ for each SNP, i , in the set of loci common to both tests. While the distribution for these combined p-value statistics, X^2 , would theoretically follow a χ_4^2 -distribution under the null when each independent test aligns with the null (no association), we observed subtle genome-wide inflation (fig. S9). Consequently, we fit the degrees of freedom, $k \in \mathbb{R}^+$, of a χ_k^2 -distribution to our observed X^2 distribution in R v. 3.6.3. and used this fitted distribution as our null for calculating p-values (fig. S9). This assumes that nearly all SNPs in the genome are not associated with the tested phenotype and is otherwise conservative. We tested for significant X^2 values at the 0.05 level for falsely rejecting the null of no genetic associations after correcting for the number of tested SNPs using a Bonferroni adjustment.

Phenotypic variance explained by candidate SNPs

We estimated how much phenotypic variation in pattern orientation (PT_{rot}), dorsal body red-green color (DC_a), or hindlimb red-green color (LC_a) among *R. imitator* from the admixture zone was explained by genotypes at the most highly associated color gene SNPs using linear models in R v. 3.6.3. Specifically, the percentage of variance explained for each of the

phenotypes was quantified as $\frac{RSS_{reduced} - RSS_{full}}{TSS} \times 100$, where TSS stands for the total sum of

squares and RSS_{full} and $RSS_{reduced}$ denote the residual sum of squares for the linear models with and without a candidate SNP predictor variable, respectively. In the full models, for each candidate SNP, the phenotypes of individual i were modeled as

$$PT_{rot,i} = \beta_0 + \beta_1 m_i + \beta_2 PB_i + \beta_3 g_i + \varepsilon_i,$$

$$DC_{a,i} = \beta_0 + \beta_1 m_i + \beta_2 DC_{L,i} + \beta_3 g_i + \varepsilon_i,$$

$$LC_{a,i} = \beta_0 + \beta_1 m_i + \beta_2 LC_{L,i} + \beta_3 g_i + \varepsilon_i$$

where the genome-wide admixture proportion, m_i , is individual i 's banded ancestry proportion (equal to 1 – striped ancestry proportion) and $g_i \in \{0,1,2\}$ is their number of minor alleles at the candidate SNP. PB , DC_L , and LC_L are covariates for the black proportion of the body dorsum and the CIELAB lightness values for the body dorsum and legs, respectively, as defined in the “Phenotyping” section of the Methods. β_0 is an intercept term, $\beta_{1,2,3}$ are effect sizes of the predictor variables, and ε is the model error. The reduced models were exactly the same, but lacked the candidate SNP predictor, G . The genotypes at candidate SNPs were called using ANGSD -doPost 1, -postCutoff 0.95, and -doGeno, which calls a genotype if its posterior probability is above 0.95, otherwise it is considered missing. The β parameters of the linear models were estimated using the `lm()` function and the incremental variance explained by inclusion of G was assessed using `anova()` to compare the full and reduced models in R.

Linkage analyses in lab-reared pedigrees

We genotyped candidate coloration gene SNPs identified from the genome-wide scans (or physically proximal SNPs in the same genes that were strongly linked to candidate SNPs; whichever we could design the most reliable primers for) in the captive pedigree of *R. imitator* in order to conduct linkage analyses. We did not conduct analyses on *krt8.2* in the pedigree due to technical difficulties sequencing this locus stemming from limited sequence for primer design around the candidate SNP of interest. We amplified target genic regions in 10 μ L PCR reactions consisting of final concentrations of 1 μ L DNA template, 1X Invitrogen *Taq* DNA polymerase PCR buffer, 2.5 mM $MgCl_2$, 0.05 mM of each dNTP, 0.2 μ M of each of the forward and reverse primers (table S12), and 0.5 μ L *Taq* DNA polymerase. Thermal conditions for PCR amplification consisted of an initial denaturation for 2 minutes at 94°C, followed by 30 cycles of 94°C for 30 seconds, 56°C for 40 seconds, and 72°C for one minute, and then lastly, a final extension at 72°C for 10 minutes. After DNA amplification, remnant primers and dNTPs were hydrolyzed using ExoSAP-IT by incubating at 37°C for 30 minutes, followed by 80°C for 15 minutes in a thermalcycler. Amplicons were cycle-sequenced in 10 μ L reactions consisting of 2 μ L of purified PCR product and final concentrations of 0.25X BigDye Terminator 3.1 ready reaction mix, 0.75X BigDye Terminator v3.1 sequencing buffer, and 0.3 μ M primer. Sequencing was carried out using an initial denaturation at 96°C, and 30 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for four minutes. Cycle sequencing products were strained through a sephadex matrix using centrifugation at 3000 RPMs for two minutes, and then the purified products were sequenced using an Applied Biosystems 3730 Genetic Analyzer.

Sequence chromatograms were analyzed in Geneious v. 4.8.5 (88). We first manually inspected and trimmed low quality positions with ambiguous peaks from the ends of both the forward and reverse sequences. Any sequences containing ambiguous peaks throughout their entire lengths were discarded. The forward and reverse sequences of each sample were then aligned and

manually checked for genotyping errors, ensuring correct calls for both homozygous and heterozygous sites by retaining only genotype calls where both the forward and reverse sequence matched and were high quality with clear peaks. Sites were translated into exomic coordinates by aligning the forward-reverse consensus sequences for each sample to the respective candidate gene contigs of the exome assembly.

Linkage analyses were performed using the Merlin software (89). We used the default settings, aside from changing the complexity threshold to allow more complex (larger) families. We tested for a genetic association with pattern orientation at candidate SNPs in *mc1r*, *asip*, and *bsn*. Some phenotypic measures of color were correlated (fig. S10), and so we used a hierarchical testing scheme to minimize redundant testing. We first tested for an association between candidate SNPs in *mc1r*, *asip*, *bsn*, and *retsat* with dorsal body and hindlimb brightness. If we failed to find a statistically significant association, then we performed subsequent tests of chroma in order to rigorously test for any association with color. Association analyses were performed separately for each body region and candidate gene SNP. Logarithm of the odds (LOD) scores were calculated, with positive scores indicating that allele sharing among individuals with similar phenotypes is higher than expected in the absence of a true association between genotype and phenotype. We controlled for falsely identified associations from multiple testing using the procedure of Benjamini-Hochberg (90), specifying a false discovery rate threshold of 10%. Though this procedure assumes independent tests and some of the tested phenotypes were correlated (fig. S1), this approach for controlling for false positives is considered to be mostly robust to non-independence (91).

Phylogenetic relationships and tests for introgression

We estimated neighbor-joining trees to visualize the evolutionary relationships between *R. imitator* and the model species across the genome and at candidate color genes in order to discern whether *R. imitator* acquired color alleles from the model species. Specifically, we calculated a pairwise genetic distance matrix among all individuals based on their genotype likelihoods using ngsDist (92) set to all defaults. For each SNP, the likelihoods of all nine possible genotypes for a pair of individuals were considered. We performed bootstrap sampling of single sites to generate 100 bootstrap replicates using ngsDist for node support. The resulting distance matrix served as input to construct a phylogenetic tree with FastME (93) using the BIONJ algorithm and SPR tree topology improvement. The tree was visualized in R using the ‘phytools’ package (94).

We used ANGSD -doAbbababa2 (95) to test for introgression from the model species into *R. imitator*. Analyses were restricted to reads from input BAM files with minimum base quality of 20 and which aligned uniquely with a map quality of at least 20. We ran -doabbbabba2 using all of the reads that passed quality control at each site (-sample 0) up to total depth of 7,100 (-maxDepth 7100) and the default block size (-blocksize 5 Mb) which effectively counts ABBA and BABA patterns for entire contigs (genes in our case) if they are less than 5 Mb long. For the ABBA-BABA test, we specified banded *R. imitator*, striped *R. imitator*, and *R. variabilis* as the H1, H2, and H3 ingroups and *R. summersi* as the outgroup. Given this configuration, excess ABBA patterns would result in a positive D-statistic, which, if significantly greater than zero,

would suggest introgression between the mimics and models. We used the `estAvgError.R` script packaged with ANGSD to calculate the average exome-wide D-statistic and to test for whether it was significant at the 0.05 level for falsely rejecting the null hypothesis of no admixture.

Selection

We used Hudson–Kreitman–Aguadé (HKA) tests (28) to determine whether the levels of polymorphism within *R. imitator* compared to substitutions from the model species were consistent with neutral evolution at exons containing SNPs associated with coloration. These tests were performed for three different sets of ingroups: the banded *R. imitator* morph, the striped *R. imitator* morph, and the banded and striped *R. imitator* morphs combined. The outgroup for this test consisted of the model species *R. variabilis* and *R. summersi* together. We used ANGSD to identify the minor allele from the genotype likelihoods (`-GL 1` and `-doMajorMinor 1`) for all striped and banded *R. imitator*, *R. variabilis*, and *R. summersi* jointly (hereon called the “joint minor allele”) at all high-quality sites across the genome where there was data for at least 15 individuals from each of the striped and banded morphs and one individual from either model species (see “Variant Discovery” of the Materials and Methods). We called a site variable if the likelihood ratio test p-value (outputted with `-SNP_pval`) for whether the joint minor allele frequency (estimated with `-doMaf 1`) was different from zero was at or below 0.01. These SNPs were considered polymorphic within a *R. imitator* morph if the morph-specific minor allele frequency (MAF) was at least 1% and polymorphic for the striped and banded morphs combined if the site was polymorphic within at least one of the morphs or a fixed difference between them. Similarly, a SNP was considered variable in one of the model species if the species-specific MAF was at least 1% and a SNP was variable among the model species (the outgroup for all HKA tests) if the site was variable within at least one of the model species or a fixed difference between them. The HKA procedure used the subset of SNPs that were either polymorphic in the ingroup or substitutions from the outgroup and omitted any sites that were polymorphic in the outgroup. All within-morph and within-species MAFs were estimated using ANGSD `-doMaf 1`. All genotype likelihood and allele frequency calculations in ANGSD were limited to bases with minimum quality scores of 20 and reads that mapped in a proper pair with a minimum map quality of 20. Minor allele frequency estimate cutoffs needed to be used to identify variable sites since hard calling genotypes would diminish accuracy in the context of our medium-depth sequencing data.

We calculated the HKA test statistic, X^2 , as in (96). Specifically, for each set of tests involving a different ingroup we used the fraction of SNPs across the genome that were segregating in the ingroup, and substitutions from the outgroup, as estimates for the neutral rates of ingroup polymorphism, p_s , and substitutions, p_d , respectively. Since we condition on sites that are either segregating in the ingroup or substitutions from the outgroup, $p_d = 1 - p_s$. For a locus consisting of L conditional SNPs, the expected number of segregating sites was calculated as $e_s = Lp_s$ and the expected number of substitutions as $e_d = Lp_d$. The HKA test statistic for a particular locus was

calculated as $X^2 = \frac{(|n_s - e_s| - 0.5)^2}{e_s} + \frac{(|n_d - e_d| - 0.5)^2}{e_d}$, where n_s and n_d are the observed number of

ingroup segregating sites and substitutions from the outgroup at the locus in question, respectively. Subtracting 0.5 from the differences of the observed and expected counts imposes the Yate's correction for continuity such that the distribution of the X^2 statistic can be better approximated by a χ^2 distribution when at least some counts are small (97). We tested for neutral evolution at the exons (and whatever flanking sequence that was captured with them) containing phenotype-associated SNPs of interest.

Factors such as linkage-disequilibrium among SNPs and complex demographic histories can cause the null distribution for the X^2 statistic to deviate from a theoretical χ^2_1 distribution. To account for this we used the `qq.chisq()` function of the `snpStat` R package (98) to calculate a X^2 correction factor, λ , based on the genome-wide distribution of X^2 statistics, which comprised all X^2 statistics for every exon in our exome assembly. λ was calculated as the ratio of the means between the observed genome-wide X^2 and expected χ^2_1 distributions after trimming values above the median as per the `qq.chisq` "trim" argument default. Values of λ were, respectively, 0.46, 0.60, and 0.56 for the tests using the banded morph ingroup, striped morph ingroup, and collective banded and striped morph ingroup. After scaling the candidate exon X^2 statistics by the respective values of λ , we compared them to a χ^2_1 distribution in order to obtain p-values, which we adjusted based on the number of tested exons for a given ingroup using the procedure of (90).

We also tested for neutrality at the candidate exons containing phenotype-associated SNPs using the program HKA (99). Given m loci, the HKA statistic in this case was calculated as

$$X^2 = \sum_{i=1}^m \frac{(n_{s,i} - \hat{E}(S_i))^2}{\hat{V}ar(S_i)} + \frac{(n_{d,i} - \hat{E}(D_i))^2}{\hat{V}ar(D_i)},$$

where $n_{s,i}$ and $n_{d,i}$ are the observed number of ingroup

segregating sites and substitutions from the outgroup, respectively, at locus i . $\hat{E}(S_i)$ and $\hat{V}ar(S_i)$ are estimates of the expectation and variance for the number of ingroup segregating sites at locus i , and $\hat{E}(D_i)$ and $\hat{V}ar(D_i)$ are the expectation and variance for the number of substitutions between the ingroup and outgroup at locus i . These expectations and variances are obtained using estimates of the split time between the ingroup and outgroup, T , and the population-scaled mutation rate, θ_i , for each locus i , under a standard neutral Wright-Fisher infinite sites model assuming equal population sizes for both the ingroup, outgroup, and ancestral population.

We calculated the HKA X^2 statistic using coalescent simulations for the set of candidate gene exons containing phenotype-associated SNPs for ingroup configurations consisting of the banded and striped *R. imitator* morphs separately, and for the two morphs together. For each set of tests (ingroups), a null distribution for the X^2 statistic was obtained using 1000 neutral coalescent simulations under the assumptions of the null HKA model (28) and estimates for T and θ_i . P-

values were calculated as the proportion of the respective null simulations greater or equal to the observed X^2 statistic.

Supplementary Text

Exome assembly and variant discovery

In order to map loci underlying color traits in the complex, 6.8 Gb (12) genome of *Ranitomeya imitator*, we designed custom sequence capture probes for 13,265 genes spanning 28.28 Mb of the exome from a developmental series of *R. imitator* transcriptomes. We used this capture system to obtain moderate-depth sequencing data (mean 12x) for 124 *R. imitator* individuals (33 banded, 33 striped, 58 admixed; Fig. 1). Sequencing data from three banded and three striped individuals were used to assemble 76.4 Mb of the *R. imitator* exome. This assembly, with an N50 of 786 bp, encompasses 13,084 (98.6%) of our targeted genes. The average contig length per gene was 5,825 bp. We mapped the sequence capture reads from all *R. imitator* samples and RNAseq reads from the model species, *R. summersi* and *R. variabilis*, to this exome assembly and jointly called 928,822 high-quality, transspecies SNPs after discarding sites segregating with a MAF below 2% in *R. imitator*. The conditional removal of sites that segregate within *R. imitator* at low frequencies was due to the fact that estimates of allele frequencies below 2% in *R. imitator* were sensitive to differences in quality control thresholds (fig. S7). Among *Ranitomeya*-wide SNPs, 295,279 were polymorphic in *R. imitator*.

Population genetic characterization of *Ranitomeya imitator*

Greater spread among striped compared to banded *R. imitator* individuals across the first two principal components of the PCA based on *R. imitator* genome-wide variation (Fig. 2A, right panel) reflects higher genetic diversity in the striped population: nucleotide diversity, $\pi = 1.56 \times 10^{-3}$, $\theta_{\text{Watterson}} = 1.38 \times 10^{-3}$, 20,751 private SNPs among 33 individuals versus banded $\pi = 9.56 \times 10^{-4}$, $\theta_{\text{Watterson}} = 9.34 \times 10^{-4}$, 8,835 private SNPs among 33 individuals. The genetic diversity of the population in the admixture zone ($\pi = 1.31 \times 10^{-3}$, $\theta_{\text{Watterson}} = 1.30 \times 10^{-3}$, 25,054 private SNPs among 58 individuals) is slightly lower than that of the striped. Individuals from this population are positioned between the striped and banded populations along PC1, consistent with their geographic location and mixed ancestry from these pure-morph populations (10). This agrees with estimates of ancestry proportions based on two ancestral lineages (Fig. 2B).

Phenotypic variance explained by candidate SNPs

We quantified how much phenotypic variance in the wild, admixed population was explained by genotypes at the candidate SNPs with the largest X^2 (Fisher's combined test statistic for the divergence and admixture zone mapping experiments) for each respective phenotype, after accounting for differences in genome-wide ancestry among individuals and technical covariates (table S8). Candidate SNPs in the pattern-associated genes *bsn*, *mc1r*, and *asip* explained 9%, 9%, and 4% of the variation in pattern orientation, respectively. These genes were also associated with body color, with the highest X^2 -ranked SNPs in *bsn*, *mc1r*, and *asip*

explaining 7%, 3%, and 2% of the color variation along the CIELAB (57) red-green axis, while the other body color candidate gene, *retsat*, explained 12%. The highest X^2 -ranked hindlimb color SNPs in *krt8.2*, *mc1r*, and *asip* explained 18%, 8%, and 1% of the hindlimb red-green axis variation.

Linkage analysis in a captive *R. imitator* pedigree

Each of the genotyped SNPs from all of the pattern candidate genes, *mc1r*, *asip*, and *bsn*, showed significant associations with pattern orientation among pedigree frogs (table S9, LOD range 0.93-2.54, q-value range 0.003 - 0.091).

The dorsal body colors in the pedigree frogs ranged from yellow to reddish/orange (Fig. 3A). Brightness (B2) and yellow wavelength chroma (S1Y, 550-625 nm) measured from the body dorsum were uncorrelated (Pearson's $r = 0.18$, $t = 1.6$, two-tailed p-value = 0.1, see also fig. S10), so we tested for associations between both of these color features and SNPs in candidate body color genes (*mc1r*, *asip*, *retsat*, *bsn*). *Asip* was significantly associated with brightness (LOD = 2.47, q-value = 0.003) but not yellow intensity (LOD = 0.17, q-value = 0.432), whereas *retsat* showed a highly significant association with dorsal yellow intensity (LOD = 3.49, q-value = 8.54×10^{-4}), but no association with brightness (LOD = 0.33, q-value = 0.342). Neither *mc1r* or *bsn* were significantly associated with brightness or yellow intensity in the pedigree (LOD scores ranging 0.09-0.48, q-values = 0.238-0.571).

There is pronounced variation in hues on the hindlimbs of the pedigree frogs (Fig. 3A), with colors ranging from yellow/orange (as in the wild banded population) to bluish/greens as in the striped population (Fig. 1). Accordingly, we measured both mean brightness (B2) and green wavelength intensities (S1G, 400-510 nm) from the hindlimbs. The green intensities are significantly correlated with brightness (Pearson's $r = 0.47$, $t = 4.2$, two-tailed p-value = 7.6×10^{-5} , fig. S10), so we only tested for associations with green wavelength if no significant associations with brightness were detected, to minimize redundant testing. We tested SNPs in the hindlimb color candidate genes *mc1r* and *asip* and found that the *mc1r* SNP was significantly associated with brightness (LOD = 1.13, q-value = 0.064). While *asip* was marginally associated with brightness (LOD = 0.62, q-value = 0.182), it showed no evidence of affecting green wavelengths (LOD = 0.0, q-value = 0.950). We did not test for associations between *krt8.2* and hindlimb color in the pedigree due to technical difficulties designing reliable sequencing primers based on the restrictive amount of assembled sequence flanking the candidate *krt8.2* SNP.

Tests of neutral evolution at candidate color genes

We used the Hudson–Kreitman–Aguadé (HKA) test (28) to determine whether levels of polymorphism within *R. imitator* relative to divergence from the model species was consistent with neutral evolution at the candidate gene exons that contained the strongest phenotype-associated SNPs (table S11). These tests rejected neutrality for the second exons of *mc1r* ($X^2_{imitator}$

= 6.25, $q = 0.032$) and *bsn* ($X_{imitator}^2 = 18.35$, $q = 1.101e-4$) and exon 11 of *retsat* ($X_{imitator}^2 = 5.82$, $q = 0.032$). Furthermore, we found evidence that some of these exons evolved non-neutrally within the different *R. imitator* morphs (table S11). For the banded morph, exons containing associated SNPs showed significant departure from neutrality for all candidate genes, except *krt8.2: mc1r* exon 2 ($X_{banded}^2 = 22.44$, $q = 1.304e-5$), *asip* exon 2 ($X_{banded}^2 = 4.67$, $q = 0.046$), *bsn* exon 2 ($X_{banded}^2 = 14.28$, $q = 4.717e-4$), *retsat* exon 11 ($X_{banded}^2 = 9.94$, $q = 0.003$). For the striped morph we found significant departure from neutral patterns at *bsn* exon 2 ($X_{striped}^2 = 15.34$, $q = 5.394e-4$), while *mc1r* exon 2 ($X_{striped}^2 = 4.46$, $q = 0.069$) and *retstat* exon 11 ($X_{striped}^2 = 4.57$, $q = 0.069$) approached the 5% false discovery significance level. We note that, in all cases where exons significantly deviated from neutral expectations, we observed decreased polymorphism within species or populations compatible with selective sweeps (table S11).

In these analyses we used genomic control to adjust for the fact that SNPs in linked sites are correlated and that the nominal p-values using standard tests of homogeneity, therefore, will not be calibrated accurately (28). However, we also used an alternative approach for obtaining p-values based on coalescent simulations under the standard neutral coalescent model. In close agreement with the HKA results based on χ_1^2 approximations, there was significant evidence of non-neutral evolution at candidate gene exons containing associated SNPs (*mc1r* exon 2, *asip* exon 1, *asip* exon 2, *bsn* exon 2, *retsat* exon 11, *krt8.2* exon 3) in the test involving both morphs combined ($X_{imitator}^2 = 10.78$, $p = 0.03$) and in the banded morph ($X_{banded}^2 = 18.70$, $p = 0.01$) while neutral evolution was nearly rejected for the striped morph ($X_{striped}^2 = 9.42$, $p = 0.06$).

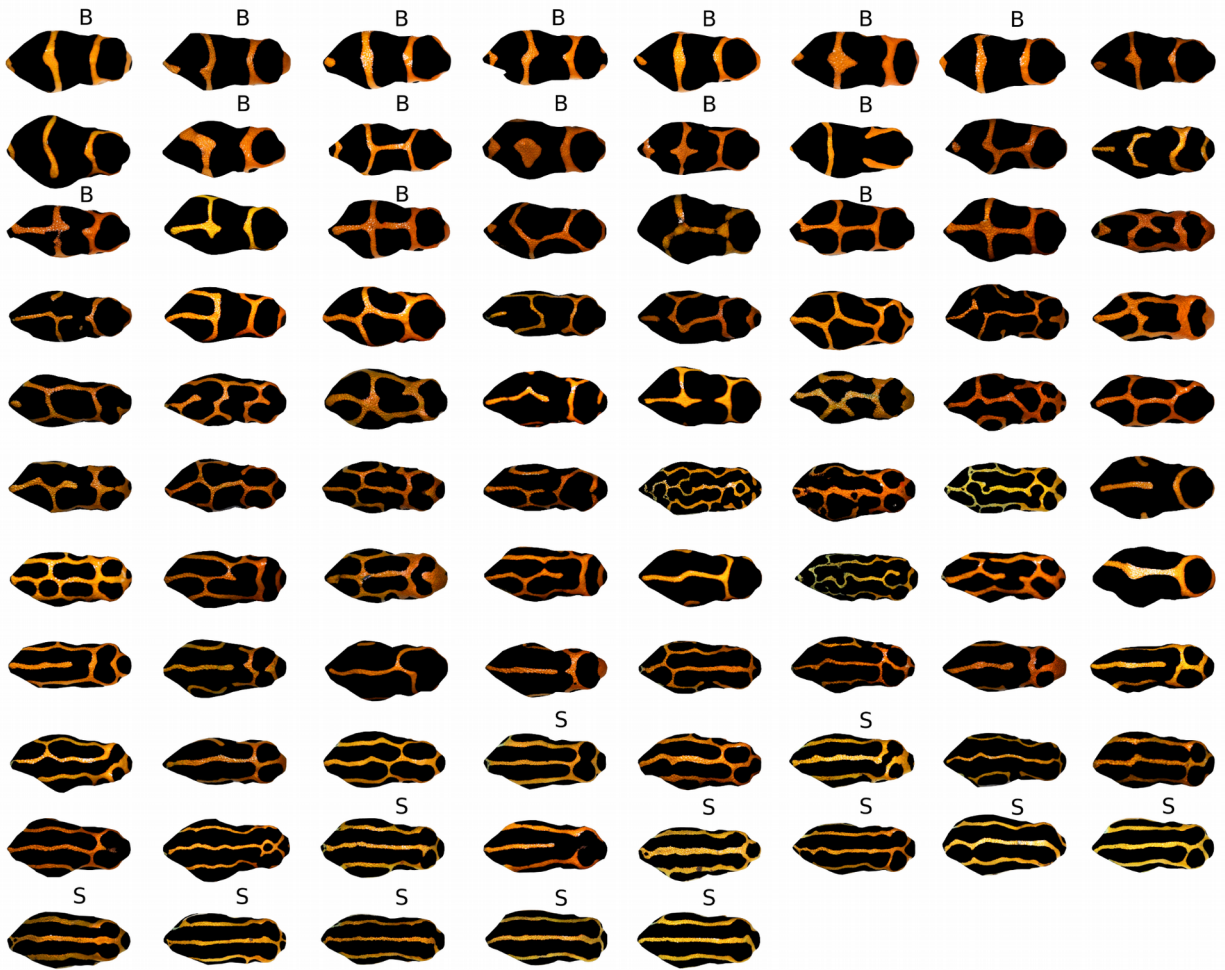


Fig. S1. Images of *Ranitomeya imitator* sorted according to pattern orientation score. Digital images from photographs of the dorsal surface of *R. imitator* are sorted according to the pattern orientation score, with the score increasing from the top to bottom row as well as left to right; the top-left frog has the lowest (most banded) score while the bottom-right frog has the highest (most striped) score. Frogs labeled with a “B” and “S” are from the banded and striped morph populations respectively, and all other frogs were sampled from the admixture zone.

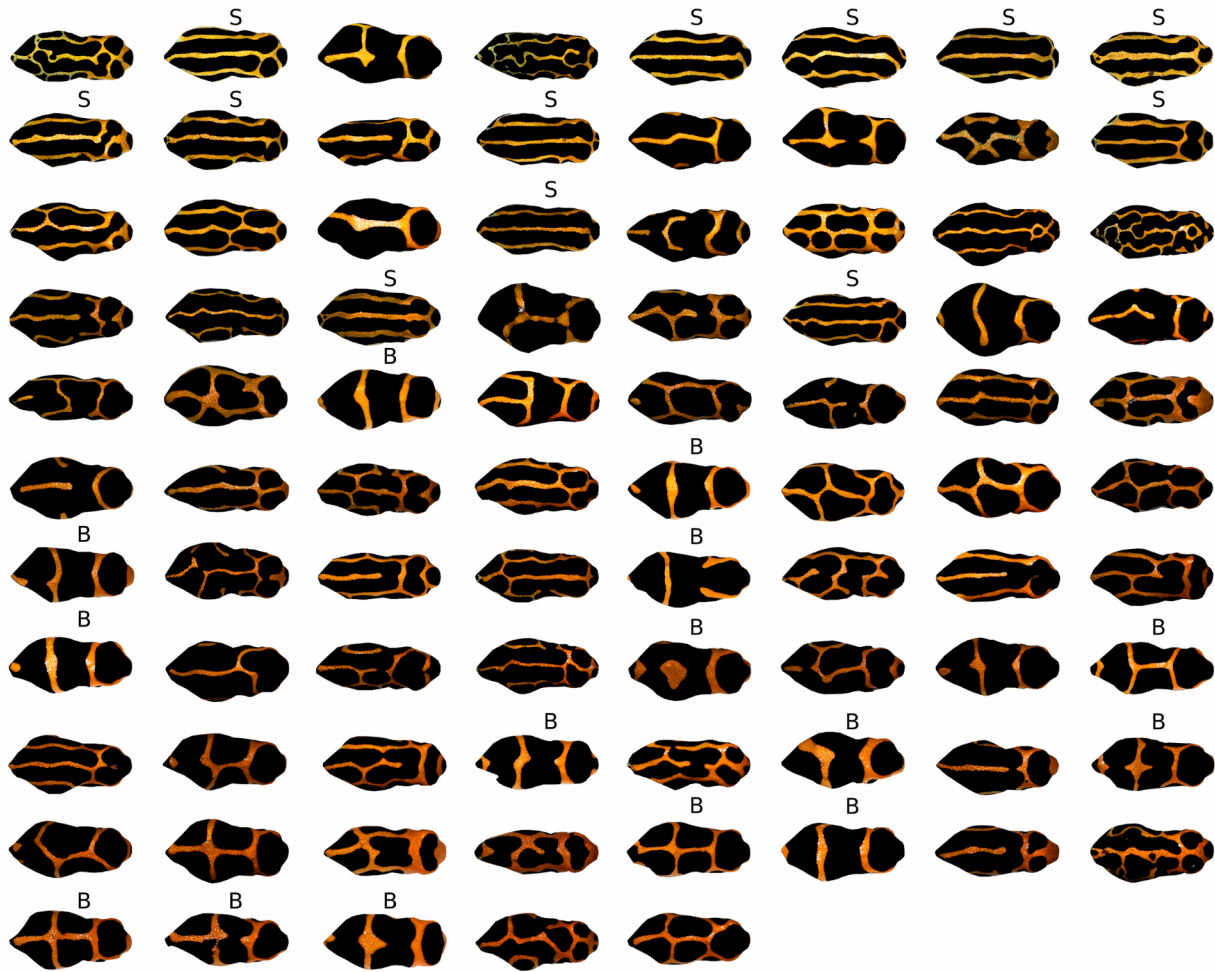


Fig. S2. Images of *Ranitomeya imitator* sorted according to body dorsum a* color value. Digital images from photographs of the dorsal surface of *R. imitator* are sorted according to their CIELAB a* (green-red axis) value, with a* increasing from the top to bottom row as well as left to right; the top-left frog has the lowest (most green) value while the bottom-right frog has the highest (most red) value. Frogs labeled with a “B” and “S” are from the banded and striped morph populations respectively, and all other frogs were sampled from the admixture zone.

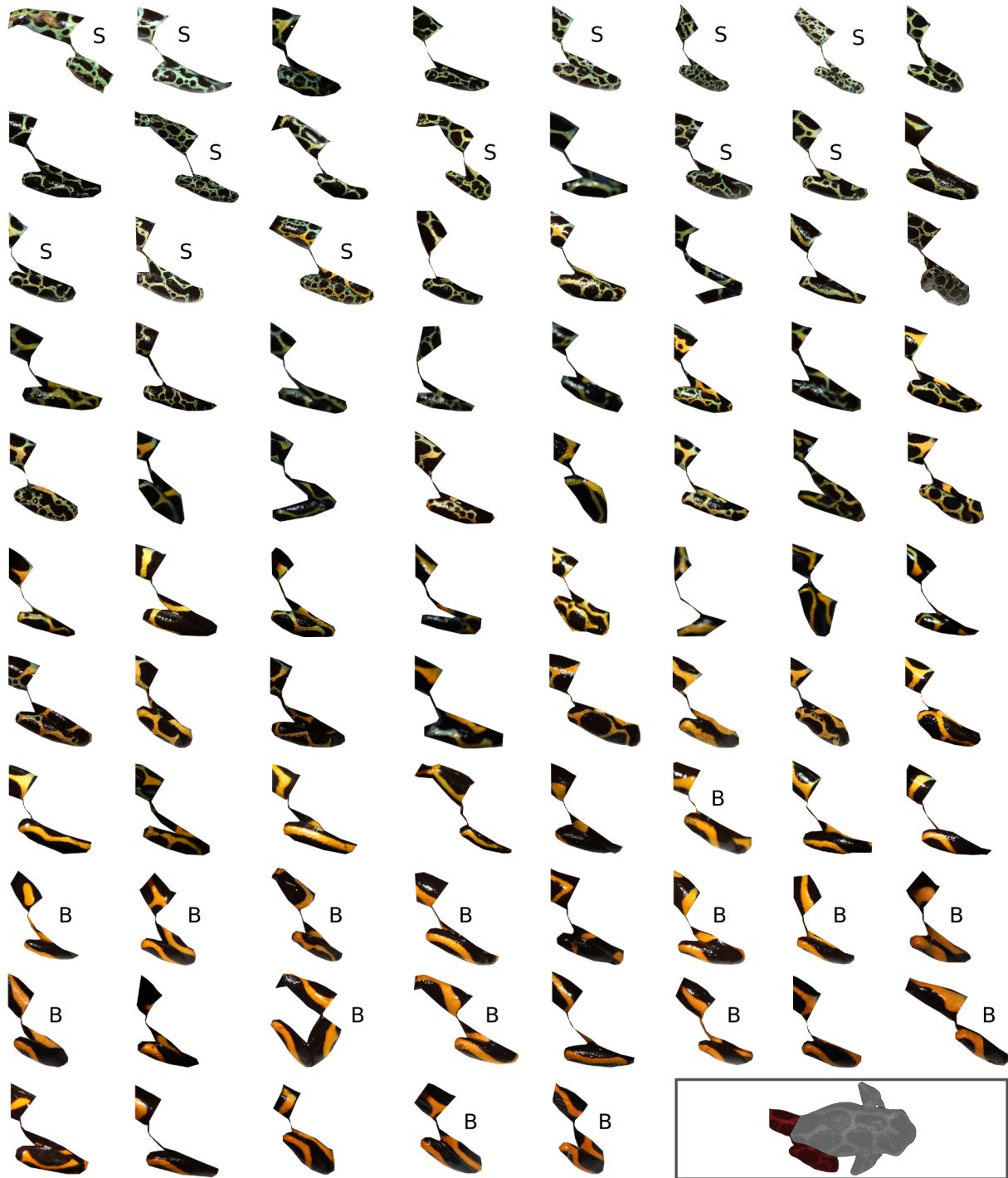


Fig. S3. Images of *Ranitomeya imitator* hindlimbs sorted according to a* color value. Digital images of *R. imitator* hindlimbs corresponding to the red-highlighted region of interest in the inset photo (bottom right). The images have been sorted according to their CIELAB a* (green-red axis) value, with a* increasing from the top to bottom row as well as left to right; the top-left frog's hindlimbs have the lowest (most green) value while the bottom-right frog's hindlimbs

have the highest (most red) value. Images labeled with a “B” and “S” indicate frogs sampled from the banded and striped morph populations respectively, and all other frogs were sampled from the admixture zone.

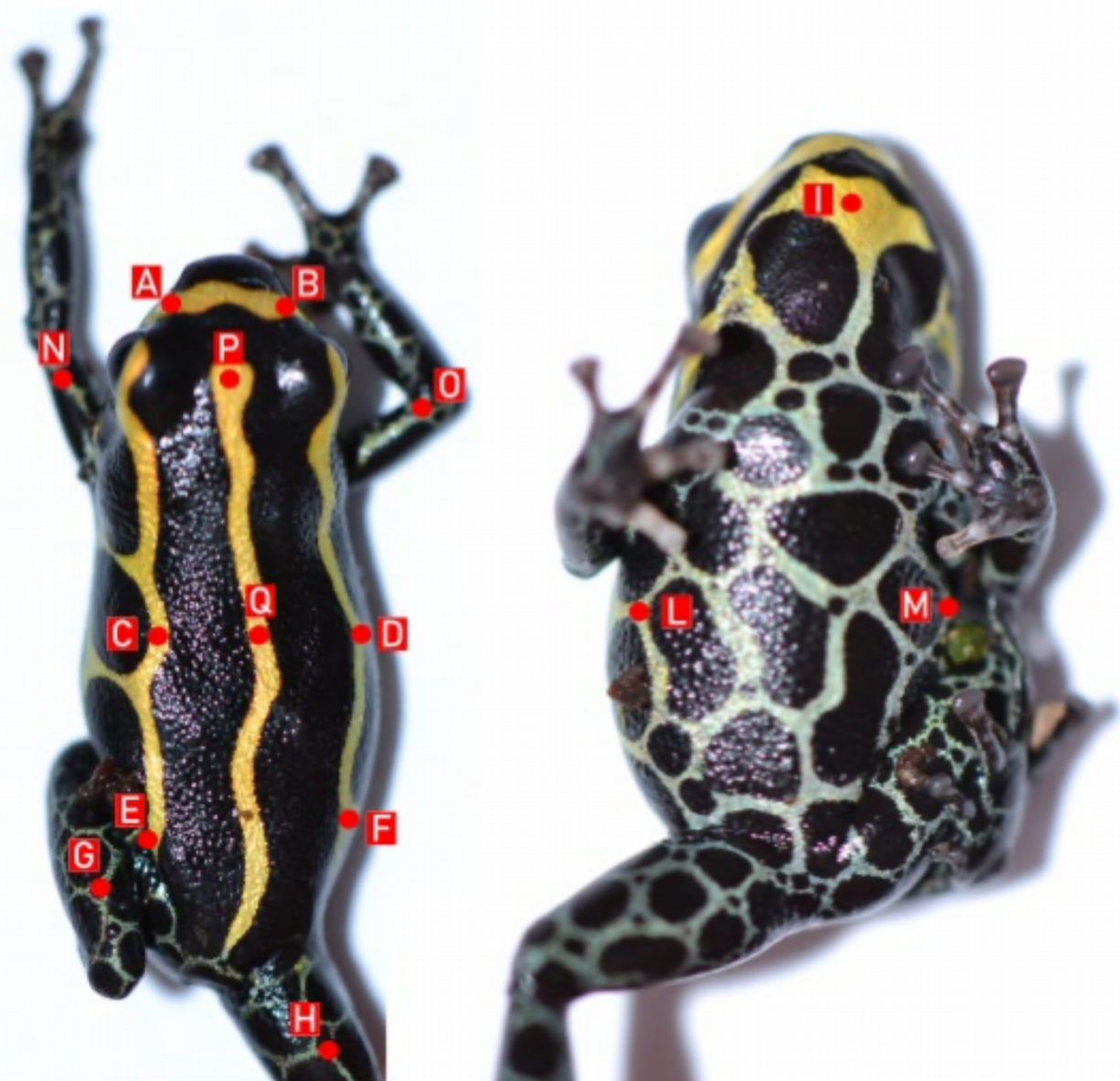


Fig. S4. Positions on the body used for spectral reflectance measurements. Red, alphabetically-labeled points mark the locations for which spectral reflectance was measured on captive-reared frogs used in the genetic linkage mapping. Reflectance was always measured from a non-black point on each frog as near to the displayed points as possible.

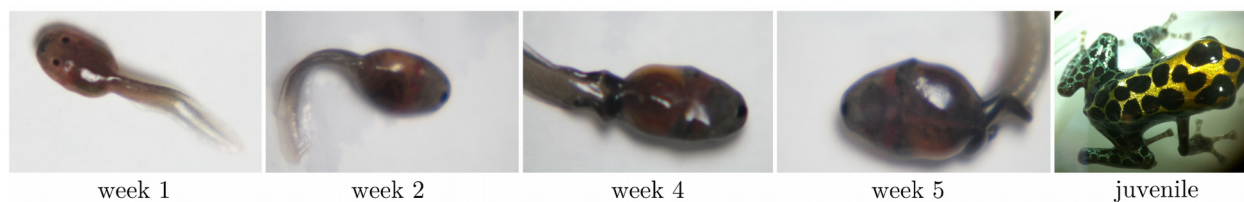


Fig. S5. Stages of development used for *Ranitomeya imitator* transcriptome sequencing.

Photos of five *R. imitator* individuals taken at the different time points throughout development for which we sequenced transcriptomes for exon capture probe design.

increases from the top to bottom panels at the expense of fewer SNPs among *R. imitator*. Higher representation among model species is important for interspecific comparisons while higher SNP density is conducive for intraspecific *R. imitator* analysis such as GWAS.

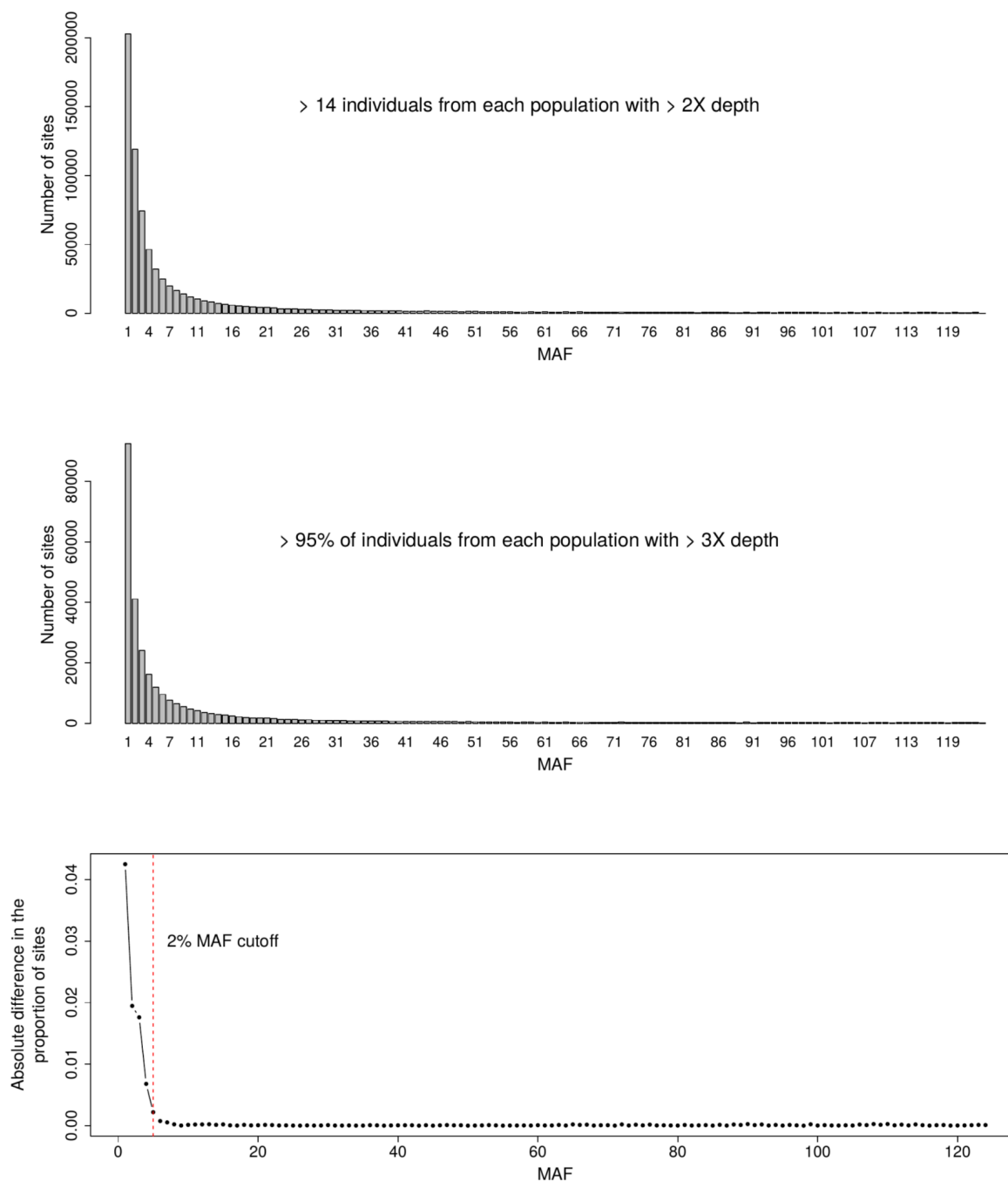


Fig. S7. Effect of depth filters on the *R. imitator* site frequency spectrum. The proportion of sites belonging to different minor allele frequency classes below 2% was sensitive to different depth of coverage criteria for the *R. imitator* sample (bottom panel). The middle panel shows the *R. imitator* SFS estimated using stricter sequencing depth criteria compared to the top panel, leading to a higher proportion of singletons, but lower doubletons and tripletons for example. In order to ensure that inferences were robust to differences in how data were filtered we removed sites with MAF below 2% in *R. imitator* from all analyses.

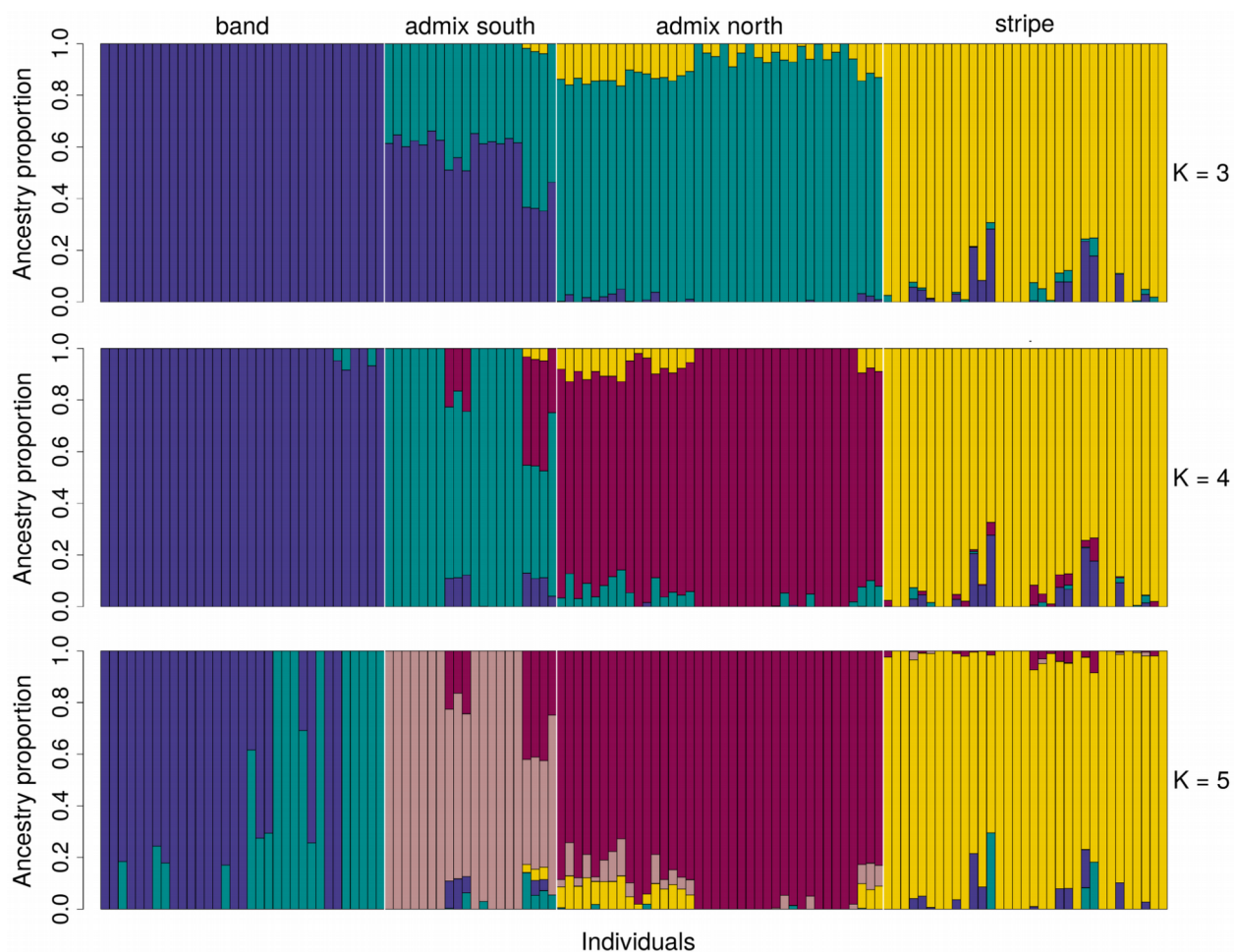


Fig. S8. Admixture analysis based on 3-5 ancestral lineages. Proportions of distinct ancestries (different colors) for the 124 *R. imitator* individuals in our sample (bars) inferred using 3-5 ancestral lineages (K) with NGSadmix. Individuals have been grouped by population, which are delineated with vertical white bars. Individuals from the “admix south” and “admix north” populations were sampled south and north of the Huallaga River, respectively (Fig. 1).

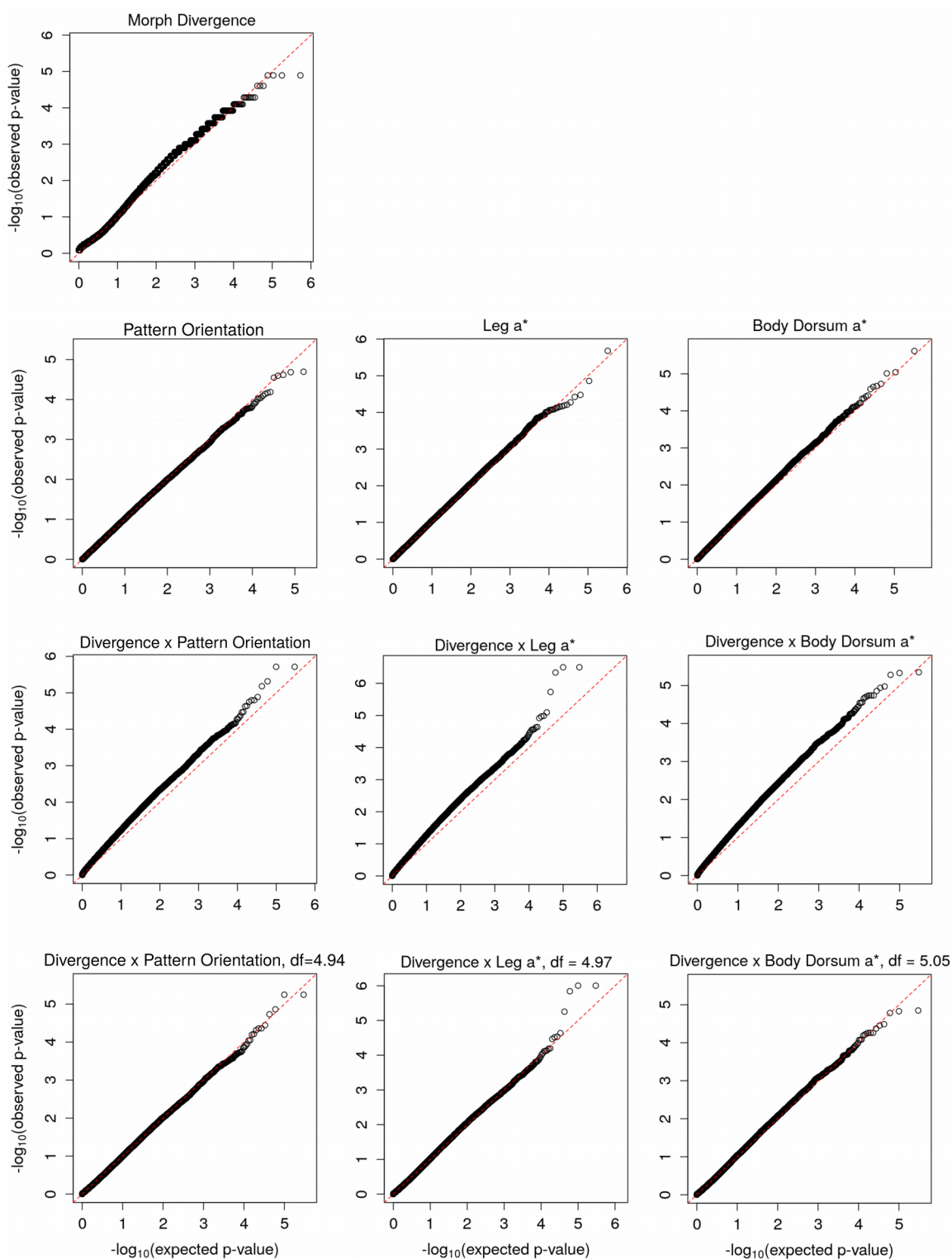


Fig. S9. Comparisons of the observed versus expected distribution of p-values for genome-wide association tests. Quantile-quantile (QQ) plots of the observed versus expected distribution of p-values for the divergence test between banded and striped morphs (top row) and the

association test using general linear models to relate genotypes to pattern orientation, hindlimb a*, and body dorsum a* (second-from-top row). The third-from-top row shows p-value QQ-plots for the X^2 statistic from combining p-values from the two aforementioned tests using Fisher's method and using the theoretical χ^2_4 null distribution. The bottom row shows p-value QQ-plots for the combined p-value X^2 statistic using a χ^2_{df} null distribution where we fit the degrees of freedom (df) to the empirical X^2 distribution.

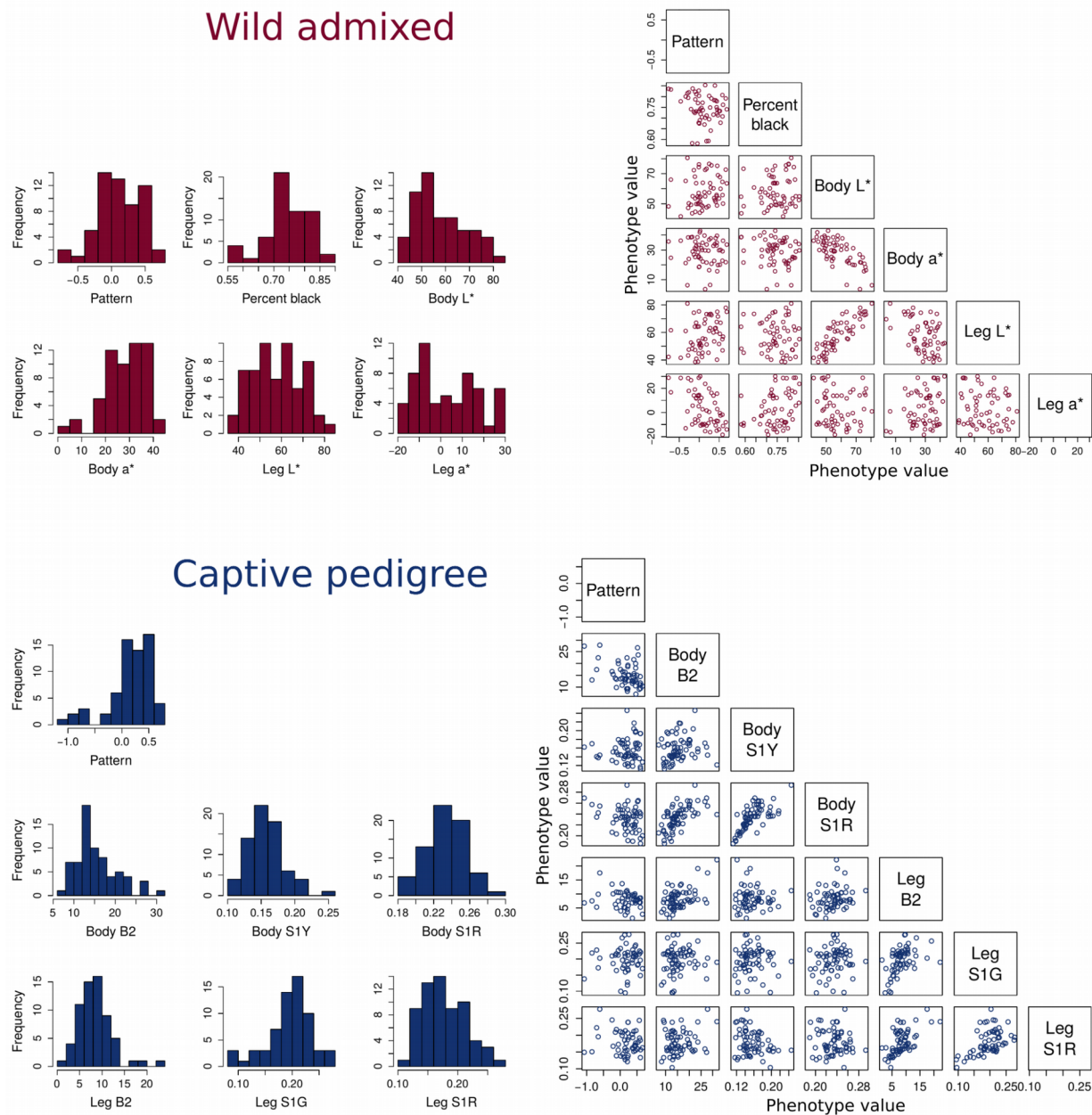


Fig. S10. Distribution of morph-related traits in captive and wild *R. imitator*. Histograms show the distribution of phenotypic values for color morph traits while scatterplots show relationships between them among wild, admixed individuals used for genetic association mapping (top, maroon) and captive pedigree individuals used for linkage mapping (bottom, blue). For the wild population individuals, we measured pattern orientation (orientation of pattern along the body dorsum), percentage of the body dorsum that was black, and the value of the non-black portions of the body and hindlimb (“leg”) dorsal surfaces along the lightness (L*) and red-green (a*) color axes of the CIELAB color space. For the captive pedigree individuals, we measured pattern orientation as well as the brightness and yellow (S1Y), red (S1R), and green (S1G) wavelength reflectance for the non-black portions of the body and/or hindlimb dorsal surfaces.

Table S1. The top 20 most divergent SNPs between the banded and striped *R. imitator* morphs in descending order of absolute allele frequency difference, δ . P-values correspond to the probability, given our sample sizes, of observing allele frequency differences larger than δ by chance under a Balding-Nichols model of pure drift. SNPs in genes deemed as candidates for influencing coloration are highlighted in gray.

SNP	Gene	Banded allele frequency	Striped allele frequency	δ	P-value
contig13293:1813	<i>mc1r</i>	0.000	0.917	0.917	1.282e-05
contig6404:219	<i>asip</i>	0.016	0.930	0.914	1.282e-05
contig2707:11381	<i>diaph1</i>	0.000	0.913	0.913	1.282e-05
contig13293:1662	<i>mc1r</i>	0.000	0.909	0.909	1.282e-05
contig6404:1582	<i>asip</i>	0.000	0.901	0.901	2.497e-05
contig11564:201	<i>ctnnb1</i>	0.038	0.935	0.897	2.497e-05
contig12059:1764	<i>bsn</i>	0.046	0.939	0.893	5.202e-05
contig2561:8521	<i>ptch2</i>	0.032	0.921	0.889	5.202e-05
contig7676:5850	<i>mphisoph10</i>	0.018	0.906	0.888	5.202e-05
contig6404:1500	<i>asip</i>	0.000	0.885	0.885	5.202e-05
contig6404:1542	<i>asip</i>	0.000	0.885	0.885	5.202e-05
contig6404:1303	<i>asip</i>	0.000	0.883	0.883	5.202e-05
contig6288:1749	NA	0.010	0.890	0.880	5.202e-05
contig6404:1031	<i>asip</i>	0.000	0.879	0.879	7.999e-05
contig734:658	<i>six1</i>	0.045	0.923	0.878	7.999e-05
contig5689:165	<i>myadm</i>	0.027	0.904	0.877	7.999e-05
contig734:1033	<i>six1</i>	0.911	0.035	0.876	7.999e-05
contig6404:1020	<i>asip</i>	0.000	0.875	0.875	7.999e-05
contig3484:3551	<i>paqr4</i>	0.039	0.912	0.873	7.999e-05
contig6404:1511	<i>asip</i>	0.000	0.872	0.872	7.999e-05

Table S2. The top 20 SNPs with the highest genotypic association with pattern orientation in the wild, admixed population of *R. imitator*. SNPs are listed in descending order of the association likelihood ratio test statistic, X^2 , along with corresponding p-values.

SNP	Gene	Minor allele frequency	X^2	P-value
contig1270:11795	<i>sorcs3</i>	0.332	18.174	2.016e-05
contig8956:965	NA	0.161	18.118	2.077e-05
contig10062:13999	<i>synj1</i>	0.222	17.827	2.420e-05
contig6234:8403	<i>fgfr2</i>	0.247	17.734	2.541e-05
contig6234:8391	<i>fgfr2</i>	0.248	17.523	2.838e-05
contig6742:866	NA	0.197976	15.948	6.5114e-05
contig5698:4152	<i>htr2c</i>	0.106	15.869	6.787e-05
contig9088:4497	<i>lta4h</i>	0.228048	15.742	7.257e-05
contig10814:1169	<i>sds</i>	0.087	15.549	8.039e-05
contig2164:182	NA	0.086	15.331	9.023e-05
contig1319:1356	NA	0.394	15.261	9.363e-05
contig1550:12192	<i>slit2</i>	0.242	15.158	9.889e-05
contig8772:4829	<i>kdm6b</i>	0.190	14.895	1.137e-04
contig12824:4264	<i>abcc9</i>	0.060	14.708	1.255e-04
contig10154:5719	<i>rf112.1</i>	0.149	14.675	1.278e-04
contig10049:266	<i>cryaa</i>	0.103	14.375	1.498e-04
contig2237:4308	<i>ikbkg</i>	0.202	14.226	1.621e-04
contig1093:2551	<i>stard3nl</i>	0.058	14.222	1.625e-04
contig11418:13299	<i>erbb4</i>	0.283	14.212	1.633e-04
contig5073:6671	<i>anapc1</i>	0.059	14.160	1.679e-04

Table S3. The top 20 SNPs with the highest genotypic association with dorsal body color in the wild, admixed population of *R. imitator*. SNPs are listed in descending order of the association likelihood ratio test statistic, X^2 , along with corresponding p-values.

SNP	Gene	Minor allele frequency	X^2	P-value
contig5757:284	<i>nox1</i>	0.057	22.228	2.421e-06
contig7597:2988	NA	0.085	19.704	9.041e-06
contig3228:5091	NA	0.196	19.567	9.716e-06
contig9908:7739	<i>adcy9</i>	0.184	18.323	1.864e-05
contig5867:49401	NA	0.338	18.062	2.139e-05
contig6133:6440	<i>ccdc180</i>	0.124	17.972	2.242e-05
contig3748:17951	<i>ptprf</i>	0.102	17.715	2.566e-05
contig12280:5221	<i>dcdc2</i>	0.079	16.948	3.843e-05
contig625:992	<i>c1galt1c1</i>	0.455	16.827	4.096e-05
contig1872:5884	<i>EIF3D</i>	0.093	16.595	4.628e-05
contig13104:2811	<i>PiGV</i>	0.128	16.572	4.685e-05
contig10362:2901	<i>CEP104</i>	0.445	16.069	6.107e-05
contig6654:1044	<i>CCDC34</i>	0.192	16.032	6.229e-05
contig524:8638	<i>MLXIP</i>	0.186	15.858	6.830e-05
contig6027:2426	<i>SYNGR2</i>	0.117	15.709	7.388e-05
contig8794:4906	NA	0.410	15.702	7.416e-05
contig1921:3694	<i>SLC24A5</i>	0.101	15.598	7.834e-05
contig6133:5488	<i>CCDC180</i>	0.128	15.587	7.881e-05
contig12498:2838	<i>CACYBP</i>	0.162	15.568	7.957e-05
contig2215:2030	<i>RNF166</i>	0.077	15.411	8.650e-05

Table S4. The top 20 SNPs with the highest genotypic association with hindlimb color in the wild, admixed population of *R. imitator*. SNPs are listed in descending order of the association likelihood ratio test statistic, X^2 , along with corresponding p-values.

SNP	Gene	Minor allele frequency	X^2	P-value
contig11350:2306	<i>krt8.2</i>	0.399	22.504	2.097e-06
contig1784:1317	<i>slc6a16</i>	0.224	18.887	1.387e-05
contig1170:1426	<i>jak1</i>	0.379	17.221	3.327e-05
contig13013:723	<i>sept4</i>	0.205	16.981	3.776e-05
contig538:3196	<i>sqlc</i>	0.100	16.325	5.335e-05
contig13013:2275	<i>sept4</i>	0.405	16.052	6.162e-05
contig5657:890	<i>btk</i>	0.084	15.962	6.4643e-05
contig5757:284	<i>nox1</i>	0.057	15.848	6.864e-05
contig8221:7298	<i>usp5</i>	0.418	15.816	6.983e-05
contig12434:590	<i>ndufa9</i>	0.068	15.700	7.423e-05
contig8477:6051	<i>mgc79783</i>	0.101	15.686	7.479e-05
contig4164:16754	<i>nin</i>	0.457	15.535	8.097e-05
contig6133:5488	<i>ccdc180</i>	0.128	15.478	8.347e-05
contig5574:15630	NA	0.177	15.457	8.442e-05
contig11508:2894	NA	0.167	15.432	8.551e-05
contig5309:1074	<i>kiss1</i>	0.470	15.417	8.621e-05
contig1841:1172	<i>tpi1</i>	0.411	15.257	9.381e-05
contig5773:63	<i>phc1</i>	0.135	15.216	9.586e-05
contig6133:6440	<i>ccdc180</i>	0.124	15.201	9.667e-05
contig2791:5447	<i>chd8</i>	0.095	15.161	9.873e-05

Table S5. The top 20 SNPs with the most evidence for influencing pattern orientation based on combined p-values from the test of excessive allele frequency divergence between the striped and banded morphs of *R. imitator* and the association test for pattern orientation in the admixed population. SNPs are sorted in descending order of Fisher’s combined p-value statistic, $X^2_{combined}$, with corresponding p-values after genomic control. The genome-wide quantiles for the absolute allele frequency difference, δ , between the striped and banded morphs and the likelihood ratio statistic, X^2_{admix} , for the test of an association with pattern orientation in the admixed population are also shown for each SNP. SNPs in candidate pattern genes are highlighted in gray.

SNP	Gene	$X^2_{combined}$	P-value	δ quantile	X^2_{admix} quantile
contig13293:1813	<i>mc1r</i>	31.97478	5.655e-06	1.0000000	0.9913467
contig13293:1662	<i>mc1r</i>	31.97478	5.655e-06	0.9999888	0.9913467
contig12059:1764	<i>bsn</i>	29.36860	1.851e-05	0.9999775	0.9922763
contig11798:9830	<i>clip1</i>	27.89937	3.597e-05	0.9990148	0.9990704
contig1270:11795	<i>sorcs3</i>	27.48134	4.342e-05	0.9352374	1.0000000
contig734:1033	<i>six1</i>	27.46881	4.367e-05	0.9999401	0.9865427
contig4416:930	<i>decr2</i>	26.70407	6.158e-05	0.9995767	0.9943476
contig10062:13999	<i>synj1</i>	26.62754	6.373e-05	0.926715	0.9999875
contig6742:866	NA	25.87821	8.916e-05	0.9515741	0.9999688
contig8326:5212	NA	25.82889	9.115e-05	0.9983217	0.9977602
contig6404:1582	<i>asip</i>	25.36693	1.120e-04	0.9999850	0.8723087
contig7105:677	<i>fam178a</i>	25.15061	1.234e-04	0.9975163	0.9975419
contig6769:405	NA	24.92548	1.364e-04	0.9999063	0.9667218
contig9196:5283	<i>rhobtb1</i>	24.88606	1.388e-04	0.9998614	0.9661041
contig3681:405	<i>aifm1</i>	24.42162	1.707e-04	0.9982431	0.9953208
contig6404:219	<i>asip</i>	24.27235	1.824e-04	0.9999963	0.5774028
contig6931:14551	<i>fancd2</i>	24.22644	1.867e-04	0.9997528	0.9692548
contig12985:4310	NA	24.09677	1.972e-04	0.9994381	0.9846711
contig6989:5080	<i>smarca2</i>	24.07623	1.990e-04	0.9996741	0.9776086
contig2491:447	<i>barx1</i>	24.05234	2.011e-04	0.9658205	0.9998004

Table S6. The top 20 SNPs with the most evidence for influencing dorsal body color based on combined p-values from the test of excessive allele frequency divergence between the striped and banded morphs of *R. imitator* and the association test for dorsal body color in the admixed population. SNPs are sorted in descending order of Fisher’s combined p-value statistic, $X^2_{combined}$, with corresponding p-values after genomic control. The genome-wide quantiles for the absolute allele frequency difference, δ , between the striped and banded morphs and the likelihood ratio statistic, X^2_{admix} , for the test of an association with dorsal body color in the admixed population are also shown for each SNP. SNPs in candidate color genes are highlighted in gray.

SNP	Gene	$X^2_{combined}$	P-value	δ quantile	X^2_{admix} quantile
contig12398:6779	<i>retsat</i>	30.19182	1.418e-05	0.998303	0.9996007
contig6288:1749	NA	30.09715	1.480e-05	0.999955	0.9922018
contig11665:7447	<i>arhgef25</i>	29.86085	1.647e-05	0.9975538	0.9996194
contig5867:49401	NA	28.35482	3.249e-05	0.95553	0.999975
contig11798:9830	<i>clip1</i>	28.17448	3.524e-05	0.9990148	0.9986587
contig4701:1592	<i>rps13</i>	27.76905	4.228e-05	0.996887	0.9992514
contig5757:284	<i>nox1</i>	27.18642	5.490e-05	0.3735503	1.0000000
contig13293:1813	<i>mc1r</i>	27.18381	5.496e-05	1.0000000	0.8837636
contig13293:1662	<i>mc1r</i>	27.18381	5.496e-05	0.9999888	0.8837636
contig3867:5730	<i>tenc1</i>	27.08005	5.758e-05	0.9998464	0.9852833
contig12059:6952	<i>bsn</i>	26.94718	6.111e-05	0.9997865	0.9894444
contig4172:7653	<i>tmem63b</i>	26.80860	6.502e-05	0.9986926	0.9977354
contig3228:5091	NA	26.28704	8.208e-05	0.8004937	0.9999875
contig4576:1777	<i>hnrnpa1</i>	26.21656	8.470e-05	0.9990522	0.9961383
contig6404:219	<i>asip</i>	26.21513	8.475e-05	0.9999963	0.8175103
contig7841:6525	<i>hap1</i>	25.77871	1.030e-04	0.9799808	0.9997442
contig13242:5961	<i>pik3c2a</i>	25.63536	1.097e-04	0.9916087	0.9992701
contig10130:6305	<i>arhgef28</i>	25.42282	1.206e-04	0.9988574	0.9945226
contig4457:3964	<i>abcb7</i>	25.37998	1.229e-04	0.9994343	0.9890451
contig2923:348	<i>vps37b</i>	25.12245	1.378e-04	0.999483	0.9875229

Table S7. The top 20 SNPs with the most evidence for influencing hindlimb color based on combined p-values from the test of excessive allele frequency divergence between the striped and banded morphs of *R. imitator* and the association test for hindlimb color in the admixed population. SNPs are sorted in descending order of the combined p-value statistic, $X^2_{combined}$, with corresponding p-values after genomic control. The genome-wide quantiles for the absolute allele frequency difference, δ , between the striped and banded morphs and the likelihood ratio statistic, X^2_{admix} , for the test of an association with hindlimb color in the admixed population are also shown for each SNP. SNPs in candidate color genes are highlighted in gray.

SNP	Gene	$X^2_{combined}$	P-value	δ quantile	X^2_{admix} quantile
contig13293:1813	<i>mc1r</i>	35.80877	9.985e-07	1.0000000	0.9984466
contig13293:1662	<i>mc1r</i>	35.80877	9.985e-07	0.9999888	0.9984466
contig11350:2306	<i>krt8.2</i>	35.00820	1.443e-06	0.9818389	1.0000000
contig1170:1426	<i>jak1</i>	32.06000	5.575e-06	0.9938601	0.9999875
contig10732:3181	NA	28.92956	2.312e-05	0.9938526	0.9997817
contig9566:3553	<i>kdm2a</i>	28.30499	3.066e-05	0.9973777	0.9995072
contig1841:1172	<i>tpi1</i>	28.05951	3.424e-05	0.9856899	0.9999002
contig1841:1172	<i>tpi1</i>	26.68330	6.358e-05	0.9852179	0.9997629
contig6404:219	<i>asip</i>	26.35300	7.372e-05	0.9999963	0.842296
contig5866:263	<i>unc119b</i>	26.29384	7.570e-05	0.9963738	0.998902
contig8221:7298	<i>usp5</i>	26.25306	7.709e-05	0.9623404	0.9999501
contig4164:16754	<i>nin</i>	25.95690	8.801e-05	0.9633931	0.9999314
contig1784:1317	<i>slc6a16</i>	25.75418	9.636e-05	0.8273983	0.9999938
contig648:8901	<i>cep83</i>	25.40405	1.127e-04	0.9695367	0.9998253
contig1696:759	<i>trappc6a</i>	25.18829	1.240e-04	0.9918822	0.9992327
contig10671:3715	<i>pgs1</i>	24.95289	1.377e-04	0.9987413	0.9942481
contig1764:30396	NA	24.86375	1.433e-04	0.9978610	0.9962194
contig8410:2730	NA	24.68980	1.548e-04	0.9966173	0.9975607
contig1841:1204	<i>tpi1</i>	24.68583	1.551e-04	0.9894322	0.9992139
contig1841:1216	<i>tpi1</i>	24.65579	1.572e-04	0.9903163	0.9991952

Table S8. Phenotypic variance explained by SNPs in candidate genes after accounting for genome-wide ancestry and technical covariates.

Phenotype	SNP	Variance explained (%)
pattern orientation	<i>mc1r</i> :1813	9.0
pattern orientation	<i>asip</i> :1582	3.8
pattern orientation	<i>bsn</i> :1764	9.0
body color	<i>mc1r</i> :1813	2.5
body color	<i>asip</i> :219	1.8
body color	<i>bsn</i> :6952	6.5
body color	<i>retsat</i> :6779	12.0
hindlimb color	<i>mc1r</i> :1813	8.3
hindlimb color	<i>asip</i> :219	1.2
hindlimb color	<i>krt8.2</i> :2306	17.6

Table S9. Logarithm of the odds (LOD) for an association between candidate gene SNPs and color phenotypes in the *R. imitator* pedigree. Significant associations with FDR below 10% are shaded gray.

Gene	Phenotype	LOD score	q-value
<i>mc1r</i>	pattern orientation	2.52	0.003
<i>mc1r</i>	body brightness (B2)	0.28	0.350
<i>mc1r</i>	body chroma (S1Y)	0.21	0.420
<i>mc1r</i>	hindlimb brightness (B2)	1.13	0.064
<i>asip</i>	pattern orientation	2.54	0.003
<i>asip</i>	body brightness (B2)	2.47	0.003
<i>asip</i>	body chroma (S1Y)	0.17	0.432
<i>asip</i>	hindlimb brightness (B2)	0.62	0.182
<i>asip</i>	hindlimb chroma (S1G)	0.00	0.950
<i>retsat</i>	body brightness (B2)	0.33	0.342
<i>retsat</i>	body chroma (S1Y)	3.49	0.001
<i>bsn</i>	pattern orientation	0.93	0.091
<i>bsn</i>	body brightness (B2)	0.09	0.571
<i>bsn</i>	body chroma (S1Y)	0.48	0.238

Table S10. D-statistic test for genome-wide introgression between the model species and *R. imitator*. No Significant differences in the number of ABBA and BABA allele configurations among banded *R. imitator* (H1), striped *R. imitator* (H2), *R. variabilis* (H3), and *R. summersi* (outgroup) were identified using Z-tests on the genome-wide, mean D-statistic and D-statistic corrected for bias (JK-D) from jackknife resampling of “nBlocks” genomic blocks. V(JK-D) is the estimate of the variance for D across blocks used to calculate the Z-statistic.

D	JK-D	V(JK-D)	Z	P-value	ABBA	BABA	nBlocks
0.00246	0.00246	0.00002	0.58527	0.55837	637.16901	634.0413	11247

Table S11. Observed and expected counts of segregating sites within *R. imitator* and fixed differences from the model species, *R. summersi* and *R. variabilis*, at candidate gene exons containing color-associated SNPs. The ratio of the observed number of *R. imitator* segregating sites (S_o) to the observed number of fixed differences from the model species (D_o) was compared to the ratio of the expected number of segregating sites (S_E) to the expected number of fixed differences (D_E) under neutrality at the candidate exons using Hudson–Kreitman–Aguadé (HKA) tests. Exons for which the HKA test significantly rejects a model of neutral evolution at a 5% false discovery rate are shaded gray. Notice that a correction for genome-wide deviation from the theoretical χ^2_1 distribution has been applied to the X^2 values.

Morph	Locus	S_o	S_E	D_o	D_E	X^2	q-value
<i>R. imitator</i> (banded and striped combined)	<i>mc1r</i> exon 2	7	13.19	37	30.81	6.25	0.032
	<i>asip</i> exon 1	2	1.50	3	3.50	4.03e-6	0.998
	<i>asip</i> exon 2	7	5.09	10	11.91	0.99	0.481
	<i>bsn</i> exon 2	15	30.27	86	70.73	18.35	1.101e-4
	<i>retsat</i> exon 11	2	6.29	19	14.71	5.82	0.032
	<i>krt8.2</i> exon 3	3	2.10	4	4.90	0.20	0.789
Banded <i>R. imitator</i>	<i>mc1r</i> exon 2	0	9.02	42	32.98	22.44	1.304e-5
	<i>asip</i> exon 1	2	1.07	3	3.93	0.47	0.591
	<i>asip</i> exon 2	0	2.58	12	9.42	4.67	0.046
	<i>bsn</i> exon 2	10	20.83	87	76.17	14.28	4.717e-4
	<i>retsat</i> exon 11	0	4.51	21	16.49	9.94	0.003
	<i>krt8.2</i> exon 3	2	1.29	4	4.71	0.10	0.756
Striped <i>R.</i>	<i>mc1r</i>	7	12.37	37	31.63	4.46	0.069

<i>imitator</i>	exon 2						
	<i>asip</i> exon 1	2	1.41	3	3.59	0.01	0.903
	<i>asip</i> exon 2	7	4.78	10	12.22	1.44	0.345
	<i>bsn</i> exon 2	14	28.11	86	71.87	15.34	5.394e-4
	<i>retsat</i> exon 11	2	5.90	19	15.10	4.57	0.069
	<i>krt8.2</i> exon 3	2	1.69	4	4.31	0.05	0.903

Table S12. Sequencing primers for genotyping candidate gene SNPs in the *R. imitator* captive pedigree.

Target SNP (gene:position)	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
<i>mc1r</i> :1813	GAAGAATCACAACCTTCACT	GAACCAGAATTCAGGACCAT
<i>asip</i> :1231	ACCCTGTTTGCAGAACATAAG	TACTGTTTACAAGCATGCCG
<i>bsn</i> :1764	TGTTCTTATTCTTCTGCAGG	GCTATTCAGCTCAATGGTCT
<i>retsat</i> :6779	CTTCGAACGAGCTCTTGATCCC	TGGTACTTTTGCGTAGCGCC