

UC Santa Cruz

UC Santa Cruz Previously Published Works

Title

The genetics, evolution, and maintenance of a biological rock-paper-scissors game

Permalink

<https://escholarship.org/uc/item/1qd8t0w3>

Journal

Science, 391(6780)

ISSN

0036-8075

Authors

Corl, Ammon

Guzman, Alex

Bi, Ke

et al.

Publication Date

2026

DOI

10.1126/science.adw8265

Peer reviewed

This is the author's version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in *Science* on January 1, 2026 at [DOI: 10.1126/science.adw8265](https://doi.org/10.1126/science.adw8265).

Title: The genetics, evolution, and maintenance of a biological rock-paper-scissors game

Authors: Ammon Corl^{1,2*}, Alex Guzman^{1,2}, Ke Bi^{1,2}, Juan Manual Vazquez², Lydia L. Smith^{1,2}, Pauline Blaimont^{3,4}, Regina Spranger^{3,5}, Robert D. Cooper⁶, Donald Miles⁷, Amy Goldberg⁸, Jian Gao⁹, Xueyan Xiang^{10,11}, Yang Zhou¹¹, Qiye Li^{9,10,11}, Guojie Zhang¹², Peter H. Sudmant², Rauri C. K. Bowie^{1,2}, Jimmy A. McGuire^{1,2}, Barry Sinervo^{3†}, Rasmus Nielsen^{2,13,14}

Affiliations:

¹ Museum of Vertebrate Zoology, University of California, Berkeley, 3101 Valley Life Sciences Building, Berkeley, California, 94720.

² Department of Integrative Biology, University of California, Berkeley, Valley Life Sciences Building, Berkeley, California, 94720.

³ Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, California 95064

⁴ University of North Carolina Chapel Hill, Department of Biology, Chapel Hill, North Carolina, 27599

⁵ Imperiled Species Program, Peninsula Humane Society & SPCA, 24103 Congress Springs Rd, Saratoga, California, 95070

⁶ Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, VA 24060, USA

⁷ Department of Biological Sciences, Ohio Center for Ecological and Evolutionary Studies, Ohio University, Athens, Ohio 45701.

⁸ Department of Human Genetics, University of California, Los Angeles, CA

⁹ College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

¹⁰ BGI Research, Wuhan 430074, China

¹¹ State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Shenzhen 518083, China

¹² Center of Evolutionary & Organismal Biology, Zhejiang University School of Medicine, Hangzhou, 310058, China

¹³ Department of Statistics, University of California, Berkeley, Evans Hall, Berkeley, California, 94720.

¹⁴ Globe Institute, University of Copenhagen, Oester Voldgade 5-7, 1350 Copenhagen K, Denmark

† Deceased

* Corresponding author. Email: ammoncorl@berkeley.edu

Abstract: Side-blotched lizards (*Uta stansburiana*) play a biological rock-paper-scissors game in which three differently colored male morphs utilize alternative mating strategies. We identify the genetic basis of this polymorphism, which was previously posited to arise from three alleles at one locus. Orange usurper and blue mate-guarder morphs are associated with two divergent haplotypes in the regulatory region of the sepiapterin reductase gene, but yellow sneaker morphs appear to arise via phenotypic plasticity from the same genetic background as blue-morphs. Our simulations show that rock-paper-scissors dynamics can better maintain a polymorphism with a genetic system of two-alleles plus plasticity than with a three-allele system. This form of balancing selection that combines genetic determination with phenotypic plasticity expands the possibilities for how stable polymorphisms arise in nature.

Main Text:

Introduction

Phenotypic differences within species allow adaption to novel environments and the formation of new species so it is important to understand the factors that generate and maintain this variability (1-5). Species can exhibit distinct morphs that differ across multiple traits such as coloration and behavior (2, 6-11). Understanding how morphs coexist within a population can reveal the selective forces that maintain variation within species (6, 12, 13). Balancing selection on morphs has been investigated for decades in a population of side-blotched lizards (*Uta stansburiana*) where three color-morphs are observed, each associated with a distinct male mating strategy (Fig. 1A, Fig. S1). The competitive dynamics amongst these morphs were the first known example of a biological rock-paper-scissors game (6). Orange-throated males can dominate blue males and control large territories with many females because of their greater testosterone levels and endurance (14, 15). Blue-throated males guard females to reduce extrapair paternity, giving them an advantage over yellow-throated males that obtain most of their reproductive success via extrapair matings (15). Yellow-throated males do not defend a territory but rather sneak onto other males' territories to obtain mates, which works well in the large territories of orange males (6, 15). Each morph has high fitness when one competitor morph is common, low fitness when the third competitor is common, and morph frequencies cycle over time (6, 16, 17). Research on side-blotched lizards has been influential for understanding how polymorphic strategies are maintained (6, 17-19) and how morphs can differ across a suite of traits including coloration, behavior, hormone levels, and brain morphology (6, 14, 15, 20-22).

Genetic studies of polymorphic species have revealed much about the development of their complex phenotypes and how multiple traits such as coloration and behavior can vary together (7-11, 23). Morphs often arise from a "supergene" in which genetic differences across multiple genes are linked as a result of suppressed recombination, typically via a chromosomal inversion (7, 9, 10, 23). However, other genetic mechanisms could give rise to complex morph phenotypes, such as a single pleiotropic gene that affects multiple traits (11, 24) or two unlinked loci that affect different morph phenotypes (8). In addition, phenotypic plasticity can generate differences in morphology and behavior similar to those observed amongst genetically determined morphs (1, 4, 25-28) and can affect how morphs are maintained in a population (1, 28). Therefore, the role and interaction of genetic and environmental factors needs to be determined to understand how polymorphic phenotypes develop.

Here we investigate the genetic basis of the color-morphs observed in side-blotched lizards to understand how they develop and how the rock-paper-scissors system evolved. Determining the genetic basis of the polymorphism is also critical for clarifying how it is maintained, because the evolutionary dynamics of the morphs have typically been modeled with their phenotypes arising from a single locus with three alleles (29). This three-allele model has been the foundation for decades of field and theoretical studies on *U. stansburiana* (6, 16, 17, 30, 31). Although some heritability of coloration has been observed (30, 31), the genetic basis of the color polymorphism has remained unresolved (29). Below we determine the genetic system underlying the color-morphs and update the models describing their evolutionary dynamics.

Results

The three morphs exhibited distinct throat colors (Fig. 1A, S1). The orange and yellow colors likely arise from pigments, whereas the blue is likely structural coloration (8, 20, 32, 33). We first investigated the genetics of the orange and blue-morphs, because they had the most distinctive color phenotypes (Fig. S1E-F) and potentially different mechanisms generating their coloration.

We assembled two genomes, one for an orange-morph and one for a blue-morph. We then performed a genome-wide association study with 44 orange and 45 blue-morphs chosen from a single population for their distinctive phenotypes. No genome-wide differentiation was detected between the morphs (Fig. S2). Instead, a single genetic region was highly associated with the orange versus blue-morph phenotype for both reference genomes (Fig. 1B, S3A-B, S4A-B). The strongest associations ($P_{min} = 7 \times 10^{-19}$) were located upstream of the gene for sepiapterin reductase (*SPR*). Only weak associations were observed in the *SPR* coding region ($P_{min} = 0.003$, (29)). The *SPR* gene is involved in two pathways relevant to the morph phenotypes (Fig. S5). First, *SPR* helps synthesize tetrahydrobiopterin, which is an essential cofactor for the enzymes that synthesize dopamine, adrenaline, noradrenaline, melatonin, and serotonin (34, 35). Second, *SPR* is part of the pathway for synthesizing pteridine pigments (36) and an orange drosopterin pteridine pigment has been found in throat skin of orange-morphs, but not in blue-morphs (20). Also, structural variation upstream of *SPR* was associated with orange and white-morphs in the distantly related European common wall lizard (*Podarcis muralis*) (8). The dual role of *SPR* could therefore allow a color signal to be coupled with differences in hormones or neurotransmitter levels that affect behavior (Fig. S5, (29)).

High levels of sequence divergence were observed between the two genomes in the associated region upstream of *SPR* (Fig. 1C). Each genome had large regions of sequence not found in the other genome, with most of the genome-specific DNA resulting from the insertion of LINE transposable elements, primarily L2 and RTE-BovB elements (Fig. 1C). The orange and blue-morph genome sequences were colinear at both a broad (Fig. S6A) and local (Fig. 1D) level around the region of top association, which showed that the morph phenotypes were not associated with a genetic inversion. The *U. stansburiana* contig with *SPR* was broadly colinear with chromosome 5 of the eastern fence lizard (*Sceloporus undulatus*) (Fig. S6B), but little homology was observed to the upstream region in the orange and blue genomes, indicating that both associated-regions were highly divergent (Fig. S6C-D).

We further investigated the genetic basis of the color-morphs by genotyping 44 orange, 45 blue, and 20 yellow-morphs for the two divergent haplotypes observed at the *SPR* associated region, which we refer to as the orange (O) and blue (B) haplotypes. The genotypes were based upon haplotype data obtained from PacBio sequencing of long-range PCRs of the *SPR* region and from Illumina short-read data that could establish the presence/absence of the divergent haplotypes. All orange-morphs (44/44) had two O-haplotypes (OO) whereas blue-morphs typically (42/45) had at least one B-haplotype (OB or BB, Fig. 1E). The overall pattern suggests a simple genetic basis in which orange-morphs are homozygous for the O-haplotype and blue-morphs have a B-haplotype that is dominant to the O-haplotype. There were three OO blue-morphs, which could indicate a more complex genetic basis, but all blue-morphs with photos had a B-haplotype so the anomalous samples could reflect a field categorization error (29). Yellow-morphs have been categorized as

having two different phenotypes (31, 37), with some having all-yellow throats and others having yellow over most of their throats except for a light blue patch at the base of the throat (“blue-yellow” phenotype, Fig. S1C). All blue-yellow males had at least one B-haplotype (10/10, Fig. 1E). The majority of all-yellow males had at least one B-haplotype (6/10), but some individuals had two O-haplotypes (4/10). These OO yellow-morphs were smaller than other males (Fig. S7), so their yellow coloration could be a transient phase of their color development (29). Most yellow-morphs have blue-yellow phenotypes (proportion over 23 years of field data = 77% blue-yellow) so if the genotype proportions that we observed are representative (i.e. 100% blue-yellow + 60% all-yellow morphs), this would mean that roughly 91% of yellow-morphs have at least one B-haplotype. Therefore, the vast majority of yellow-morphs have a B-haplotype and are genetically similar to blue-morphs.

Long-term balancing selection is expected to maintain high levels of genetic diversity. To study this, we first characterized variation among 48 B-haplotypes and 108 O-haplotypes. There were large regions of haplotype-specific DNA (Fig. 2A) and many additional alignable nucleotides were fixed between haplotypes (Fig. 2B, S8A). Most of these fixed differences were in a region that had high linkage disequilibrium (Fig. S8C) and lacked shared polymorphisms between the haplotypes (Fig. 2B), suggesting an absence of recombination between the haplotypes. This non-recombining region contained the variants with the highest association to morph phenotypes and also TE insertions that were fixed in each haplotype (i.e. all TEs in the blue-unique regions and the intervening orange-unique region (Fig. 1C, 2A). Levels of nucleotide diversity measured across all haplotypes peaked within and directly adjacent to the non-recombining region (Fig. 2C), consistent with it being under balancing selection. B-haplotypes had less genetic variation than O-haplotypes, suggestive of a recent selective sweep on the B-haplotype. We next compared the nucleotide diversity of the *SPR*-region (excluding haplotype-specific DNA) to estimates of nucleotide diversity from across the rest of the genome. The *SPR*-region had substantially elevated nucleotide diversity ($\pi = 0.0490$, Fig. 2D) compared to other genomic regions (mean = 0.0046, min = 0.0002, max = 0.0175).

The strong association between morph phenotypes and the upstream-region of *SPR* suggests that regulatory changes likely generate morph phenotypes. Therefore, we investigated whether the morphs differed in expression of the *SPR* gene (Fig. 2E, S9, (29)). The morphs had significant differences in *SPR* expression in colored throat skin (Fig. 2E, $R^2 = 0.30$, $F_{2,21} = 6.0$, $P = 0.0087$). Blue-morphs had higher expression than orange-morphs (Fig. 2E, $P = 0.0062$), but no significant differences were observed in the other two morph comparisons unless a single outlying yellow-morph data point was excluded, in which case the difference between blue and yellow-morphs was also significant ($P = 0.0221$). *SPR* expression levels were even better explained by a model that took *SPR*-region genotypes into account (Fig. S9A, $R^2 = 0.36$, $F_{2,21} = 7.5$, $P = 0.0035$), because the genotypes helped explain variation among yellow-morphs, with individuals with a B-haplotype having higher expression than OO individuals (e.g. the outlying yellow-morph was BB). Compared to orange-morphs, blue-morphs and individuals with a B-haplotype had increased expression of *SPR* in colored skin relative to noncolored skin (Fig. S9C, D), so heightened levels of *SPR* accompany the formation of blue coloration (Fig. S5, (29)). B-haplotypes were expressed at higher levels than O-haplotypes in individuals heterozygous for both haplotypes (Fig. 2F, S9E-F, K). This haplotype-specific expression supports a *cis*-regulatory mechanism (38), which is strong evidence that the O and B-haplotypes differentially regulate the expression of *SPR*.

We inferred a phylogeny for multiple O and B-haplotypes to reveal the evolutionary history of the *SPR*-region and to examine whether there were any subtypes of either haplotype that were associated with different phenotypes. The two haplotypes were deeply divergent from each other (Fig. 2G, S10), as expected under long-term balancing selection (5). There was no evidence that genetic differences among O-haplotypes result in different phenotypes, because the O-haplotypes found in heterozygous blue-morph individuals and the three anomalous OO blue-morphs did not form distinct phylogenetic clusters from O-haplotypes in orange-morphs. Instead, the blue and orange-morphs were only distinguished by the presence/absence of the B-haplotype, which is dominant to any O-haplotype in generating blue-morph phenotypes. Importantly, we also found that the haplotypes from yellow-morphs did not form a distinct genetic cluster within either the O or B-haplotypes. This observation rejects the hypothesized three-allele model, which requires a third genetic group that differentiates yellow-morphs from the other two morphs.

A new model for the genetics of the polymorphism and the development of the yellow-morph phenotype was needed. We hypothesized that the yellow-morph phenotype arises from developmental phenotypic plasticity wherein lizards with a B-haplotype that are unable to acquire a territory express yellow coloration and adopt a sneaker strategy, but express blue coloration if they acquire and defend a territory. This two-allele model with plasticity was supported in several ways (29). First, prior field observations and experiments have noted that some yellow-morphs can irreversibly transform to a blue-morph phenotype late in the breeding season, typically coincident with the opening of a territory (14, 21). Second, blue and yellow-morphs have similar *SPR* genotypes (Fig. 1E) and are not distinguished by a distinct type of O or B-haplotype (Fig. 2G), so both morphs can arise from the same genetic background at *SPR*. Third, yellow-morphs could potentially arise from variation at other genes than *SPR*, but genome-wide association tests with yellow-morphs failed to reveal any additional candidate genes (Fig. S3C, D), including the *BCO2* gene (Fig. S4C, D) that gives rise to yellow-morphs in *P. muralis* (8), which makes a multilocus model unlikely. Fourth, we compared the predictions of the two and three-allele models to 14 years of field surveys of the polymorphism. The two-allele model predicts a positive correlation in the frequency of the blue and yellow-morphs within a year because they share the same genetic basis, but a highly negative correlation between these morphs and orange-morphs (Fig. 3A, S11). In contrast, the three-allele model predicts a weak negative correlation among all morphs, because an increase in the frequency of one allele necessitates a decrease in the frequency of the other alleles. The field data clearly matched the predictions of the two-allele model with plasticity, because it showed a positive correlation between the frequencies of the blue and yellow-morphs (Fig. 3B, $F_{1,12} = 5.92$, $P = 0.032$) and highly negative correlations between the frequencies of these morphs and the frequency of orange-morphs (Fig. 3B, $F_{1,12} = 18.10$, $P = 0.001$; Fig. S11).

Our new understanding of the genetics and development of the morphs necessitated a reevaluation of how the mating strategy polymorphism is maintained, because prior studies (6, 17, 19) used an incorrect genetic model. Our goal was to determine whether rock-paper-scissors fitness dynamics can maintain diversity in a two-allele system, with one allele giving rise to two different phenotypes via irreversible plasticity. We first tested a series of deterministic models with orange-morphs arising from homozygotes for the O-haplotype, yellow-morphs arising from a proportion of all individuals carrying a B-haplotype, and blue-morphs arising from the remaining individuals with a B-haplotype. The parameters of

frequency-dependent selection on the morph-phenotypes were determined by prior field studies on male territorial interactions and paternity (6, 15, 19).

The deterministic models showed that two alleles can be maintained by frequency-dependent selection in the form of a rock-paper-scissors game with the plastic generation of two phenotypic strategies (Fig. 3C-E, S12-S14). The two alleles were maintained across a wide range of possible proportions of the yellow sneaker morph, but the system was dominated by the orange allele when the yellow-morph proportion was at low values (Fig. 3D, S12), which shows that a minimum proportion of yellow-morphs is necessary for the polymorphism to exist. The population genetic models (Fig. S12-S13) and a game theoretical analysis (Fig. S15, (29)) both showed that there are many yellow-morph proportions in which a rare orange or rare blue + yellow allele can invade a population of the opposite type and lead to a stably maintained polymorphism. A rock-paper-scissors game based on asexual, fixed strategies exhibits frequency oscillations (39, 40) and morph frequency cycles have been observed in side-blotched lizards (6, 19). Some versions of our two-allele model did exhibit frequency cycles (Fig. 3E, S13H, S14B,D). The six-generation cycles observed in one version of the two-allele model (Fig. 3E) are more similar to the five-year cycles observed in nature (6, 19) than the 44-generation cycles of the three-allele model (Fig. S14J). Additional factors such as density-dependent selection on the color-morphs in females (41) or climatic effects on morph frequencies (17) could also play a role in generating cycles.

Prior theory suggests that plasticity could affect the stability of a rock-paper-scissors game (39, 42, 43). We added genetic drift to the models to compare the stability of the two and three-allele models (Fig. 3F, S16). The two-allele model was more stable than the three-allele model with yellow-morph proportions of 30-60% and with yellow-morph proportions of 70-80% for small population sizes (Fig. 3F). A two-allele model that optimized the blue and yellow-morph proportions relative to the orange-morph frequency had the highest stability. In some cases, there were dramatic differences in stability. At a population size of 100, the three-allele model lost an allele in 91% of the trials, the two-allele model with 40% yellow-morphs lost an allele in 3% of the trials, and the optimized two-allele model always maintained the polymorphism. These results show that for a broad range of conditions the two-allele model with plasticity is superior to the three-allele model at maintaining a genetic polymorphism and that plastic adjustments of morph frequencies within a generation can further promote stability.

Discussion

Our study provides important insights into the genetic basis and evolution of polymorphic phenotypes. First, complex morph phenotypes need not arise from a supergene. Side-blotched lizard morphs do not arise from a chromosomal inversion that links coloration and behavioral genes, but instead from variation at a single gene that could couple differences in hormones or neurotransmitters with distinct throat-colors via pleiotropy. Although supergenes are a common mechanism for the origin of morphs (7, 9, 10, 23), our results and similar observations in other polymorphic species (11, 24) suggest that pleiotropic effects of key regulatory genes are sufficient for generating distinct morph phenotypes. Second, convergent genetic evolution among distantly related species can generate color-morphs, as demonstrated by *SPR* being associated with polymorphism in both *U. stansburiana* and *P.*

muralis (8), which diverged ~179 million years ago (44). Studies of birds suggest that morphs often arise from different genes (7, 10, 11, 23), but our results and studies of butterflies (45, 46) suggest that certain genes may be repeatedly coopted to generate color-morphs.

Our results also broaden our understanding of how multiple behavioral strategies are maintained within a population. Our models demonstrate that the maintenance of genetic diversity by a rock-paper-scissors game does not require that all the strategies are genetically determined and that the stability of the system is enhanced by having plastic strategies. Phylogeographic study of the side-blotched lizard morphs showed that although the polymorphism has been lost eight times, the color-morphs are geographically widespread and have been maintained for millions of years (5), which is consistent with the system being highly stable. Examples of rock-paper-scissors dynamics in other species remain rare (47-50) and theory suggests that three genetically-determined strategies will often exhibit unstable dynamics (51). However, our findings expand the potential for the rock-paper-scissors game to maintain biodiversity by recognizing that plastic strategies can be part of the game and can promote its stability. There are many species with three male mating strategies that involve a degree of plasticity (25-28) and the side-blotched lizard polymorphism was initially contrasted to two such species (52). We revealed that side-blotched lizard morphs have more in common with species that have plastic strategies than was formerly recognized. Therefore, the maintenance of polymorphism via a rock-paper-scissors game may apply to a much broader array of species than was previously envisioned.

Plasticity could not only help maintain morphs, but it could also facilitate the evolution of biological rock-paper-scissors games. A genetically determined rock-paper-scissors game requires at least two mutations to generate three strategies, but sequential evolution of these mutations would be difficult since only two strategies in a population would be unstable (e.g. the evolution of paper in a population of rock would lead to the loss of rock). This difficulty is removed if two strategies arise from the same genetic background because of plasticity, which might easily occur if behavioral differences accompany morphological changes during development or if there are conditional strategies (1, 28). Plastic strategies can be maintained even with unequal fitness if they are conditionally expressed and increase the fitness of individuals in that condition (1, 28) (e.g. sneaker males may have lower fitness than territorial males, but if a male lacks a territory, sneaking is better than zero fitness). Rock-paper-scissors games could thus evolve either in a plastically dimorphic population if a new mutation generates a third fixed strategy (Fig. S12F, L; S15), or in a monomorphic population if a new mutation generates two additional strategies via plasticity (Fig. S12D, J; S15). Thus, plasticity facilitates the simultaneous generation of the three strategies needed for a rock-paper-scissors game, so plastic phenotypes should be integral to many additional biological examples of this game.

References and Notes

1. M. J. West-Eberhard, *Developmental plasticity and evolution* (Oxford University Press, New York, 2003).
2. M. J. West-Eberhard, Alternative adaptations, speciation, and phylogeny (a review). *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1388–1392 (1986).

3. S. Lamichhaney *et al.*, A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science*. **352**, 470–474 (2016).
4. A. Corl *et al.*, The genetic basis of adaptation following plastic changes in coloration in a novel environment. *Current Biology*. **28**, 2970–2977 (2018).
5. A. Corl, A. R. Davis, S. R. Kuchta, B. Sinervo, Selective loss of polymorphic mating types is associated with rapid phenotypic evolution during morphic speciation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4254–4259 (2010).
6. B. Sinervo, C. M. Lively, The rock-paper-scissors game and the evolution of alternative male strategies. *Nature*. **380**, 240–243 (1996).
7. C. Küpper *et al.*, A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2015).
8. P. Andrade *et al.*, Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1–10 (2019).
9. M. Joron *et al.*, Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*. **477**, 203–206 (2011).
10. E. M. Tuttle *et al.*, Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*. **26**, 344–350 (2016).
11. K.-W. Kim *et al.*, Genetics and evidence for balancing selection of a sex-linked colour polymorphism in a songbird. *Nat. Commun.*, 1–11 (2019).
12. P. Nosil *et al.*, Evolution repeats itself in replicate long-term studies in the wild. *Science Advances*. **10**, 1–20 (2024).
13. E. I. Svensson, J. Abbott, R. Hardling, Female polymorphism, frequency dependence, and rapid evolutionary dynamics in natural populations. *The American Naturalist*. **165**, 567–576 (2005).
14. B. Sinervo, D. B. Miles, W. A. Frankino, M. Klukowski, D. F. DeNardo, Testosterone, endurance, and Darwinian fitness: natural and sexual selection on the physiological bases of alternative male behaviors in side-blotched lizards. *Horm. Behav.* **38**, 222–233 (2000).
15. K. R. Zamudio, B. Sinervo, Polygyny, mate-guarding, and posthumous fertilization as alternative male mating strategies. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14427–14432 (2000).
16. C. Bleay, T. Comendant, B. Sinervo, An experimental test of frequency-dependent selection on male mating strategy in the field. *Proc. Biol. Sci.* **274**, 2019–2025 (2007).
17. D. Friedman, J. Magnani, D. Paranjpe, B. Sinervo, Evolutionary games, climate and the generation of diversity. *PLoS ONE*. **12**, e0184052 (2017).

18. R. Calsbeek, S. H. Alonzo, K. Zamudio, B. Sinervo, Sexual selection and alternative mating behaviours generate demographic stochasticity in small populations. *Proc. Biol. Sci.* **269**, 157–164 (2002).
19. B. Sinervo, Runaway social games, genetic cycles driven by alternative male and female strategies, and the origin of morphs. *Genetica*. **112**, 417–434 (2001).
20. D. C. Haisten, D. Paranjpe, S. Loveridge, B. Sinervo, The cellular basis of polymorphic coloration in common side-blotched lizards, *Uta stansburiana*. *Herpetologica*. **71**, 125–135 (2015).
21. S. C. Mills *et al.*, Gonadotropin hormone modulation of testosterone, immune function, performance, and behavioral trade-offs among male morphs of the lizard *Uta stansburiana*. *The American Naturalist*. **171**, 339–357 (2008).
22. L. D. LaDage, B. J. Riggs, B. Sinervo, V. V. Pravosudov, Dorsal cortex volume in male side-blotched lizards, *Uta stansburiana*, is associated with different space use strategies. *Anim. Behav.* **78**, 91–96 (2009).
23. I. Sanchez-Donoso *et al.*, Massive genome inversion drives coexistence of divergent morphs in common quails. *Curr Biol.* **32**, 462–469.e6 (2022).
24. A. L. Ducrest, L. Keller, A. Roulin, Pleiotropy in the melanocortin system, coloration and behavioural syndromes. *Trends Ecol. Evol.* **23**, 502–510 (2008).
25. S. H. Alonzo, M. Taborsky, P. Wirtz, Male alternative reproductive behaviours in a Mediterranean wrasse, *Symphodus ocellatus*: evidence from otoliths for multiple life-history pathways. *Evolutionary Ecology Research*. **2**, 997–1007 (2000).
26. J. M. Rowland, D. J. Emlen, Two thresholds, three male forms result in facultative male trimorphism in beetles. *Science*. **323**, 773–776 (2009).
27. M. Wikelski, S. S. Steiger, B. Gall, K. N. Nelson, Sex, drugs and mating role: testosterone-induced phenotype-switching in Galapagos marine iguanas. *Behav. Ecol.* **16**, 260–268 (2005).
28. M. R. Gross, Alternative reproductive strategies and tactics: diversity within sexes. *Trends Ecol. Evol.* **11**, A92–A98 (2003).
29. See the Supplementary Text in the Supplementary Materials online.
30. E. I. Svensson, A. G. McAdam, B. Sinervo, Intralocus sexual conflict over immune defense, gender load, and sex-specific signaling in a natural lizard population. *Evolution*. **63**, 3124–3135 (2009).
31. B. Sinervo, C. Bleay, C. Adamopoulou, Social causes of correlational selection and the resolution of a heritable throat color polymorphism in a lizard. *Evolution*. **55**, 2040–2052 (2001).

32. G. Zhang *et al.*, Lizards exploit the changing optics of developing chromatophore cells to switch defensive colors during ontogeny. *Proc. Natl. Acad. Sci. U.S.A.* **120**, 1–9 (2023).
33. J. T. Bagnara, P. J. Fernandez, R. Fujii, On the blue coloration of vertebrates. *Pigment Cell Res.* **20**, 14–26 (2007).
34. K. Ikemoto *et al.*, Localization of sepiapterin reductase in the human brain. *Brain Res.* **954**, 237–246 (2002).
35. E. R. Werner, N. Blau, B. Thöny, Tetrahydrobiopterin: biochemistry and pathophysiology. *Biochem. J.* **438**, 397–414 (2011).
36. I. Braasch, M. Scharl, J.-N. Volff, Evolution of pigment synthesis pathways by gene and genome duplication in fish. *BMC Evol. Biol.* **7**, 74 (2007).
37. B. Sinervo *et al.*, Self-recognition, color signals, and cycles of greenbeard mutualism and altruism. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7372–7377 (2006).
38. B. Ge *et al.*, Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat. Genet.* **41**, 1216–1222 (2009).
39. J. M. Smith, *Evolution and the Theory of Games* (Cambridge University Press, New York, 1982).
40. A. Szolnoki *et al.*, Cyclic dominance in evolutionary games: a review. *Journal of The Royal Society Interface.* **11**, 1–20 (2014).
41. B. Sinervo, E. Svensson, T. Comendant, Density cycles and an offspring quantity and quality game driven by natural selection. *Nature.* **406**, 985–988 (2000).
42. M. Kleshnina, S. S. Streipert, J. A. Filar, K. Chatterjee, Mistakes can stabilise the dynamics of rock-paper-scissors games. *PLoS Comput Biol.* **17**, 1–16 (2021).
43. S. Moulherat *et al.*, The roles of plasticity versus dominance in maintaining polymorphism in mating strategies. *Sci. Rep.*, 1–10 (2017).
44. F. T. Burbrink *et al.*, Interrogating genomic-scale data for Squamata (lizards, snakes, and amphisbaenians) shows no support for key traditional morphological relationships. *Syst. Biol.* **69**, 502–520 (2020).
45. R. D. Reed *et al.*, *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science.* **333**, 1137–1141 (2011).
46. N. J. Nadeau *et al.*, The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature.* **534**, 106–110 (2016).
47. B. Kerr, M. A. Riley, M. W. Feldman, B. J. M. Bohannan, Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature.* **418**, 171–174 (2002).

48. B. Sinervo *et al.*, Models of density-dependent genic selection and a new rock-paper-scissors social system. *The American Naturalist*. **170**, 1–8 (2007).
49. B. Sinervo, R. Calsbeek, The developmental, physiological, neural, and genetical causes and consequences of frequency-dependent selection in the wild. *Annu. Rev. Ecol. Evol. Syst.* **37**, 581–610 (2006).
50. D. Semmann, H.-J. Krambeck, M. Milinski, Volunteering leads to rock-paper-scissors dynamics in a public goods game. *Nature*. **425**, 390–393 (2003).
51. S. H. Alonzo, R. Calsbeek, The unstable dynamics of multiple alternative reproductive tactics. *J. Evol. Biol.* **23**, 2614–2624 (2010).
52. J. M. Smith, The games lizards play. *Nature*. **380**, 198–199 (1996).
53. A. Corl *et al.*, Data from the paper: The genetics, evolution, and maintenance of a biological rock-paper-scissors game. *Dryad* (2025), doi:10.5061/dryad.4qrfj6qq7.
54. R. Maia, H. Gruson, J. A. Endler, T. E. White, pavo2: new tools for the spectral and spatial analysis of colour in R. *Methods Ecol. Evol.* **10**, 1097–1107 (2019).
55. J. Fox, S. Weisberg, *An {R} companion to applied regression* (Sage, Thousand Oaks, CA, Third. 2011).
56. S. A. Miller, D. D. Dykes, H. F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
57. R. Kajitani *et al.*, Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).
58. R. Luo *et al.*, SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. **1**, 18 (2012).
59. S. B. Sim, R. L. Corpuz, T. J. Simmonds, S. M. Geib, HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*. **23**, 157 (2022).
60. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods*. **18**, 170–175 (2021).
61. R. C. Harris, Improved pairwise alignment of genomic DNA. *Ph.D. Thesis, The Pennsylvania State University* (2007).
62. A. K. Westfall *et al.*, A chromosome-level genome assembly for the eastern fence lizard (*Sceloporus undulatus*), a reptile model for physiological and evolutionary ecology. *Gigascience*. **10**, 1–17 (2021).

63. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
64. E. W. Sayers *et al.*, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **40**, D13–25 (2012).
65. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0. 2013-2015
 <<http://www.repeatmasker.org>> (2013).
66. P. Joviet, J. W. Foley, Solutions for purifying nucleic acids by solid-phase reversible immobilization (SPRI). Ludmer Centre for Neuroinformatics and Mental Health, Montreal, QC [dx.doi.org/10.17504/protocols.io.bnz4mf8w](https://doi.org/10.17504/protocols.io.bnz4mf8w) (2015).
67. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **25**, 1754–1760 (2009).
68. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience.* **10**, 1–4 (2021).
69. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
70. S. Andrews, FastQC: A quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
71. K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics.* **32**, 292–294 (2016).
72. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* **32**, 3047–3048 (2016).
73. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics.* **15**, 356 (2014).
74. H. Li, Improving SNP discovery by base alignment quality. *Bioinformatics.* **27**, 1157–1158 (2011).
75. K. Hanghøj, I. Moltke, P. A. Andersen, A. Manica, T. S. Korneliussen, Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience.* **8**, 1–9 (2019).
76. T. S. Korneliussen, I. Moltke, NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics.* **31**, 4009–4011 (2015).
77. P. W. Hedrick, R. C. Lacy, Measuring relatedness between inbred individuals. *J. Hered.* **106**, 20–25 (2015).

78. M. Fumagalli, F. G. Vieira, T. Linderöth, R. Nielsen, ngsTools: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*. **30**, 1486–1487 (2014).
79. S. Y. Kim *et al.*, Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinformatics*. **12**, 231 (2011).
80. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org/> (2021).
81. R. R. Sokal, F. J. Rohlf, *Biometry: the principles and practice of statistics in biological research* (W. H. Freeman and Company, New York, ed. 3, 1995).
82. A. Untergasser *et al.*, Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
83. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
84. K. Katoh, K. Misawa, K.-I. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
85. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. **35**, 526–528 (2019).
86. E. Paradis, pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*. **26**, 419–420 (2010).
87. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).
88. L. J. Revell, Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
89. J. Lemon, Plotrix: a package in the red light district of R. *R-News*. **6**, 8–12 (2006).
90. A. J. Page *et al.*, SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*. **2**, 1–5 (2016).
91. C. Theodoris, T. M. Low, P. Pavlidis, N. Alachiotis, quickLD: An efficient software for linkage disequilibrium analyses. *Mol Ecol Resour.* **21**, 2580–2587 (2021).
92. J.-H. Shin, S. Blay, B. McNeney, J. Graham, LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Soft.* **16**, 1–10 (2006).
93. T. Hothorn, F. Bretz, P. Westfall, Simultaneous inference in general parametric models. *Biometrical Journal*. **50**, 346–363 (2008).

94. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).
95. B. Sinervo, K. R. Zamudio, The evolution of alternative reproductive strategies: fitness differential, heritability, and genetic correlation between the sexes. *J. Hered.* **92**, 198–205 (2001).
96. P. W. Hedrick, *Genetics of Populations* (Jones and Bartlett Publishers, Inc., Boston, Mass., Third. 2005).
97. H. Wickham, R. François, L. Henry, K. Müller, D. Vaughan, dplyr: A grammar of data manipulation. R package version 1.1.4, <https://dplyr.tidyverse.org> (2025).
98. H. Wickham, T. Pedersen, D. Seidel, scales: Scale functions for visualization. R package version 1.3.0, <https://scales.r-lib.org>. (2024).
99. H. Wickham, *ggplot2: Elegant graphics for data analysis* (Springer-Verlag, New York, 2016; <https://ggplot2.tidyverse.org>).
100. J. B. Arnold, ggthemes: Extra themes, scales and geoms for “ggplot2.” R package version 4.2.0. <https://CRAN.R-project.org/package=ggthemes> (2019).
101. B. Auguie, gridExtra: Miscellaneous Functions for “Grid” Graphics. R package version 2.3. <https://CRAN.R-project.org/package=gridExtra> (2017).
102. A. Corl, thesis, University of California, Santa Cruz, Santa Cruz (2007).
103. M. Frean, E. R. Abraham, Rock-scissors-paper and the survival of the weakest. *Proc. R. Soc., Ser. B: Biol. Sci.* **268**, 1323–1327 (2001).
104. J. Yu *et al.*, WLS retrograde transport to the endoplasmic reticulum during Wnt secretion. *Developmental Cell*. **29**, 277–291 (2014).
105. N. Farias de Farias *et al.*, Chromosome mapping of retrotransposon AviRTE in a neotropical bird species: *Trogon surrucura* (Trogoniformes; Trogonidae). *Genome*. **67**, 307–315 (2024).
106. I. Ziegler, The pteridine pathway in zebrafish: Regulation and specification during the determination of neural crest cell-fate. *Pigment Cell Res.* **16**, 172–182 (2003).
107. P. Andrade, M. Carneiro, Pterin-based pigmentation in animals. *Biol. Lett.* **17**, 1–12 (2021).
108. D. H. Edwards, E. A. Kravitz, Serotonin, social status and aggression. *Curr. Opin. Neurobiol.* **7**, 812–819 (1997).
109. M. Zerani, A. Gobbetti, NO sexual behaviour in newts. *Nature*. **382**, 31–31 (1996).
110. R. J. Nelson *et al.*, Behavioral abnormalities in male mice lacking neuronal nitric oxide synthase. *Nature*. **378**, 383–386 (1995).

111. B. Sinervo, D. B. Miles, in *Hormones and Reproduction of Vertebrates, Volume - Reptiles*, D. O. Norris, K. H. Lopez, Eds. (Journal of Arid Environments, 2011), vol. 3, pp. 215–246.
112. H. Kim, K. Kim, J. Yim, Biosynthesis of drospterins, the red eye pigments of *Drosophila melanogaster*. *IUBMB Life*. **65**, 334–340 (2013).
113. B. Thony, G. Auerbach, N. Blau, Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochem. J.* **347**, 1–16 (2000).
114. B. Dasgupta *et al.*, in *Gliomas*, W. Debinski, Ed. (2021), pp. 121–137.
115. M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
116. Kanehisa Laboratories, KEGG purine metabolism - reference pathway (2023).

Acknowledgments:

Thanks to the Arbelbide family for permission to work on their land. Thanks to Pacific Biosciences, Noëlle Bittner, Ben Karin, Oanh Nguyen, Bryan Bach, and Xochitl Valencia for help with PacBio sequencing. Thanks to two anonymous reviewers for helpful comments that improved the paper.

Funding:

National Science Foundation grant EF1241848 (BS, DBM)

National Science Foundation grant DEB1950636 (DBM)

Author contributions:

A.C., P.B., R.S., R.C., D.M., J.A.M, R.C.K.B, and B.S. contributed to field work. A.C, J.M.V, P.H.S., J.G., X.X., Y.Z., Q.L., and G.Z. performed lab work and analyses related to genome assembly. A.C., A.Guzman, L.S., and K.B. performed other lab work. A.C., K.B., A.Goldberg, and R.N. analyzed the data. A.C. and R.N. performed the modeling. A.C. and R.N. wrote the first draft of the paper, which was edited by the other authors.

Competing interests:

Authors declare that they have no competing interests.

Data and materials availability:

The genetic data from this study has been deposited in Genbank (accession numbers: PX436996 - PX437172 and PX512685 - PX512686) and in the Sequence Read Archive in BioProject PRJNA1328718 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA1328718>). Additional data and the R-code used to analyze the data is available on Dryad (53). The specimens underlying the genetic work have been deposited in the Museum of Vertebrate Zoology at the University of California, Berkeley. The museum catalog numbers and the genetic accession numbers for all samples are available on Dryad in the files:

608 Uta_DNAseq_samples_with_SRA_accession_numbers.xlsx,
609 Uta_RNAseq_samples_with_SRA_accession_numbers.xlsx, and
610 Uta_PacBio_amplicon_metadata_and_Genbank_ID.xlsx.

611

612 **Supplementary Materials**

613 Materials and Methods

614 Supplementary Text

615 Figs. S1 to S16

616 References (53–115)

617

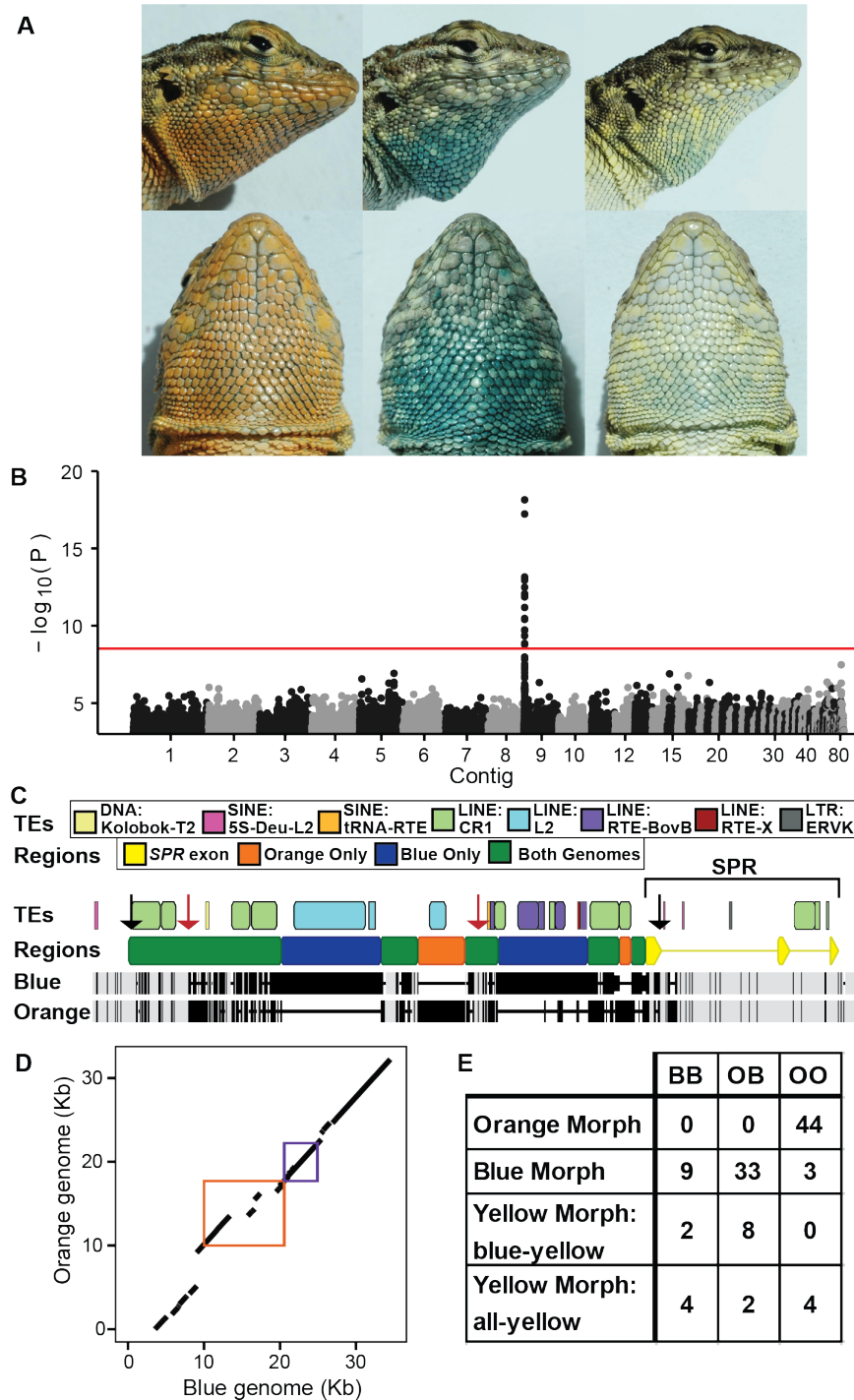


Fig. 1: Morph phenotypes and associated genetic region. (A) Color phenotypes of the orange, blue, and yellow-morphs. (B) Genetic associations with orange versus blue-morph phenotypes. Red line = multiple-testing correction threshold. (C) Alignment of the blue and orange-morph genomes in the associated region (bottom two rows) showing identical positions (gray), differences (black), and gaps (horizontal lines). The regions above show the *SPR* coding-region (yellow), alignable sequences between genomes (green), and sequences restricted to the blue-genome (blue) and orange-genome (orange). The top row shows transposable elements (TEs) colored by the repetitive element family, with the class of TE given before the family name. Arrows: top associated positions (red) and primer locations for long-range PCR (black). (D) Alignment of the orange and blue-morph genomes at the *SPR* coding-region (purple) and divergent-region upstream of *SPR* (orange). (E) Morph genotypes for the O and B-haplotype.

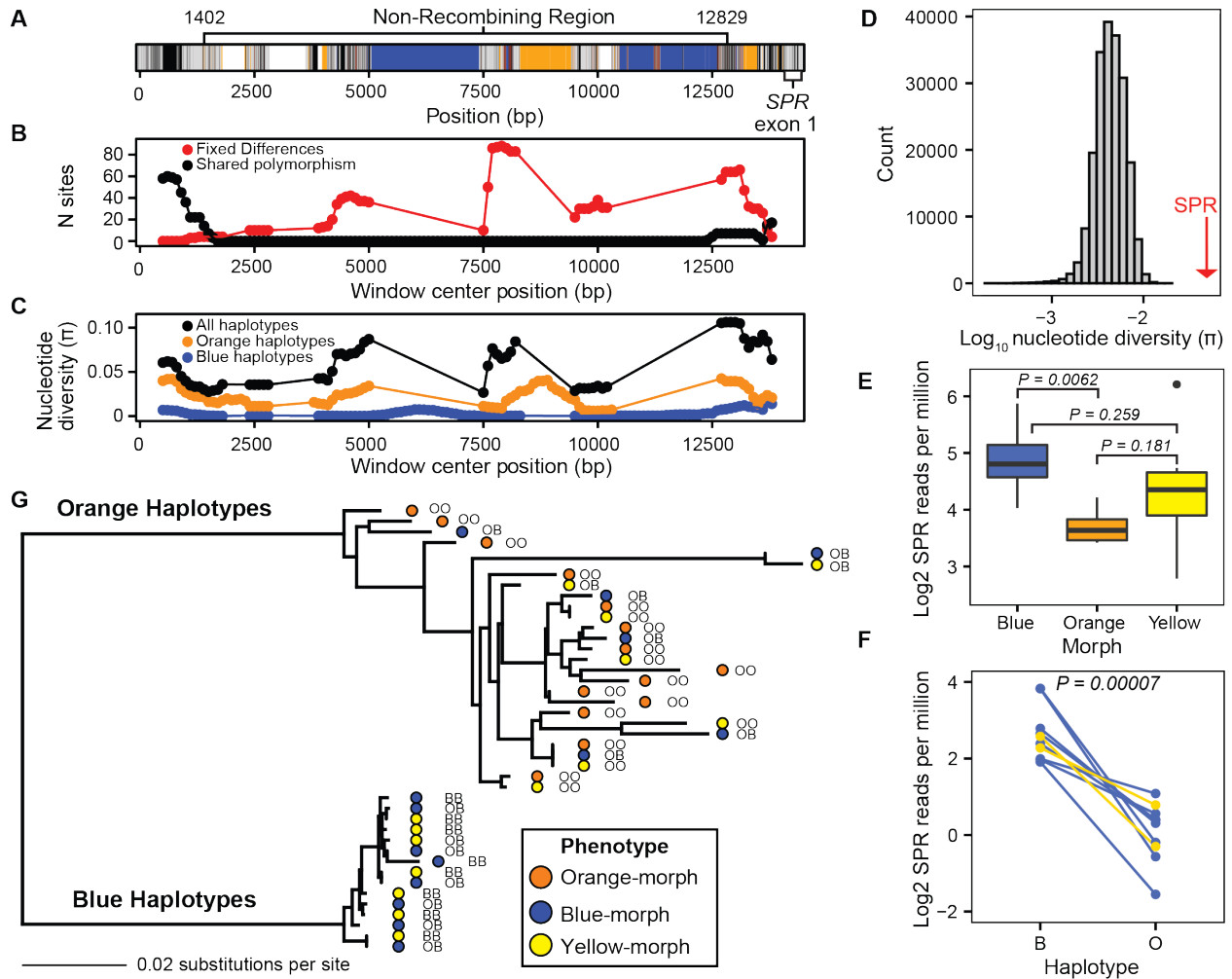


Fig. 2: Genetic diversity, expression, and evolution of the *SPR* region. (A) Top: Overview of the alignment of 48 blue and 108 orange haplotypes showing positions that were: invariant (gray), variable (black), primarily gaps (white), unique to blue/orange haplotypes (blue/orange), and fixed differences (red). (B) The number of fixed differences (red) and shared polymorphic sites (black) between the blue and orange haplotypes. (C) Nucleotide diversity (π) for 1000 bp windows along blue, orange, and both (black) haplotypes, with gaps as lines. (D) Genetic diversity of the *SPR* region (red arrow) relative to other genomic regions. (E) *SPR* expression in colored-throat skin. Boxes show the 25th-75th percentiles with a line for the median, minimum/maximum with whiskers, and a point for an outlier. P -values = post-hoc comparisons. (F) Haplotype-specific expression in colored skin within individuals heterozygous for the O and B-haplotypes. Lines connect the data from an individual and the colors show blue versus yellow-morphs. (G) Phylogeny of *SPR*-region haplotypes with associated morph phenotypes (circles) and the *SPR*-region genotype of the individual.

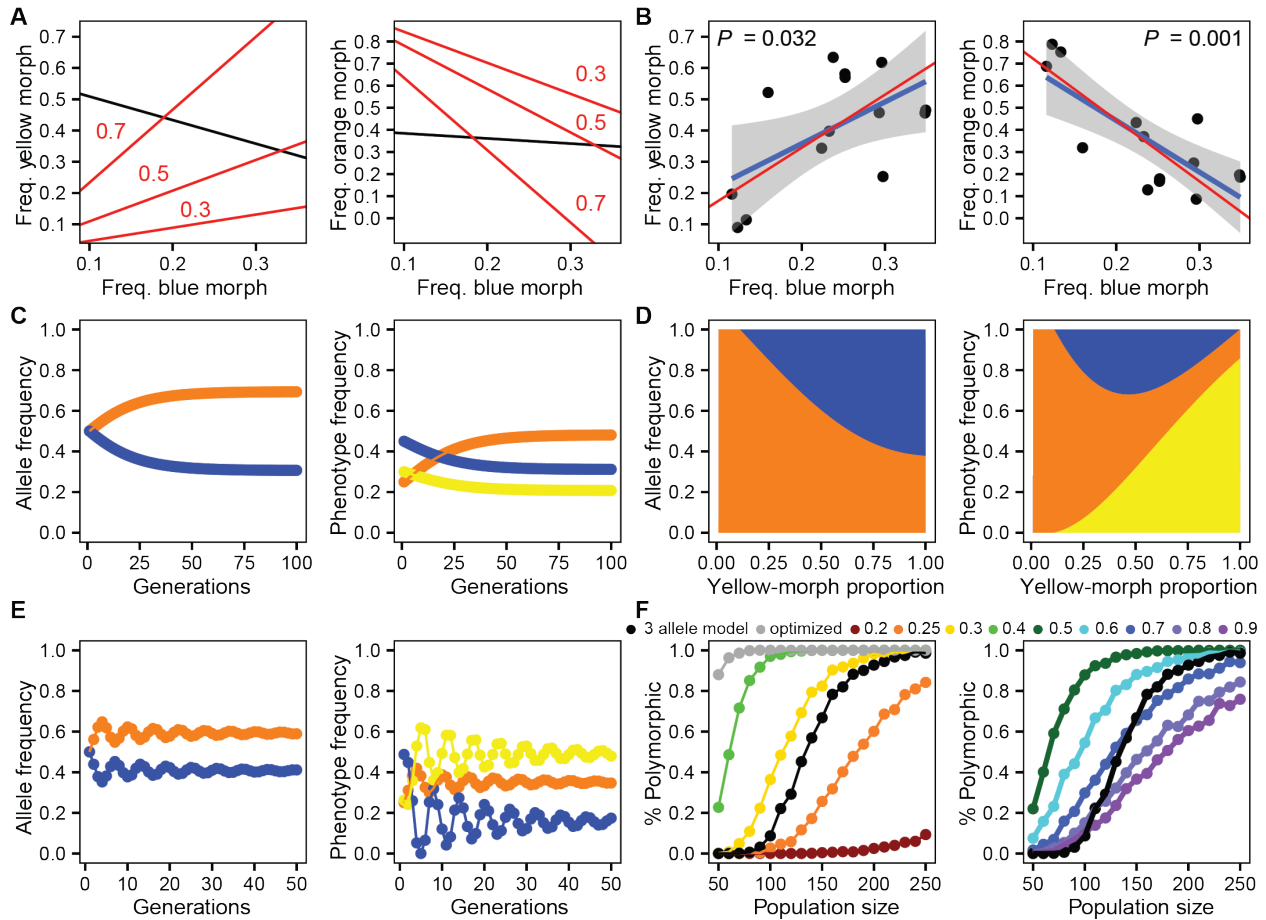


Fig. 3: Morph frequencies, maintenance, and stability. (A) Predicted relationships between the frequencies of different morphs from the 3-allele model (black) and 2-allele model with different proportions of yellow-morphs arising from B-haplotype individuals (red lines). (B) Correlations in morph frequencies observed in the field data (blue line). The red line is the two-allele model prediction with 64% of B-haplotype individuals becoming yellow-morphs as estimated from the field data. (C) Simulation using a two-allele model with 40% of B-haplotype individuals becoming yellow-morphs in each generation. (D) The allele and phenotype frequencies observed after 10,000 generations in simulations starting with equal allele frequencies and varying the fixed proportion of yellow-morphs in the two-allele model. (E) Simulation with a two-allele model in which the proportion of blue/yellow-morphs was optimized in each generation but then offset so that 30% more individuals become blue-morphs. Morph fitness was estimated based on territory-overlap data, whereas other panels used paternity data. (F) Stability at different population sizes of the two-allele model with fixed proportions of yellow-morphs across generations (colored lines) or optimized within a generation (gray) compared to the three-allele model (black). The Y-axis is the proportion of 1000 trials in which the polymorphism was maintained for 1000 generations.



Supplementary Materials for

The genetics, evolution, and maintenance of a biological rock-paper-scissors game

Ammon Corl, Alex Guzman, Ke Bi, Juan Manual Vazquez, Lydia L. Smith, Pauline Blaimont, Regina Spranger, Robert Cooper, Donald Miles, Amy Goldberg, Jian Gao, Xueyan Xiang, Yang Zhou, Qiye Li, Guojie Zhang, Peter H. Sudmant, Rauri C. K. Bowie, Jimmy A. McGuire, Barry Sinervo, Rasmus Nielsen

Corresponding author: ammoncorl@berkeley.edu

The PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S16
References

Other Supplementary Materials for this manuscript include the following:

None.

Materials and Methods

Field sampling

Samples of *U. stansburiana* were collected from our focal study population near Los Banos, CA (36.98761798 N, -121.0476778 W). Animals were collected under California Department of Fish and Game permits to B. Sinervo (CDFW 3879395-4, 033138010, SC-3574) and Pauline Blaimont (CDFW SC-12852) following U.C. Santa Cruz IACUC protocols (Sineb1203, Sineb1404, Sineb1503, Sineb1703, and Sineb1802). Tissue samples collected from one blue and one orange-morph (Fig. S1) were used to assemble a reference genome for each morph. To generate RNA-Seq data, we collected eight males of each of the three morphs during a single field season, each displaying prominent orange, blue, or yellow throat coloration (Fig. S1A). Throat skin and brain samples were collected from these males, which were separately preserved in RNAlater at 4°C overnight and at -80°C thereafter. Additional tissue samples were obtained from tail tips preserved in ethanol that were collected as part of the mark-recapture study of morph dynamics at the focal study population. We generated genome resequencing data for samples from 45 blue, 45 orange, and 10 yellow-morphs that were chosen to maximize their phenotypic differences. Blue-morph individuals had dark blue covering most or all of their throats (i.e. scored as “bb” according to the phenotype codes used by the Sinervo lab, see Supplementary Text), orange-morphs had throats covered by orange (i.e. scored as “oo” according to the phenotype codes used by the Sinervo lab), and yellow-morphs had light yellow and white covering their throats, sometimes with a tiny amount of light blue at the base of their throats (all-yellow phenotype; i.e. scored as “yy” according to the phenotype codes used by the Sinervo lab). Orange and blue-morph samples were chosen so that equal numbers of each morph were sampled from a particular year and approximately equal numbers were sampled from particular rock outcrops within each year to avoid any temporal or spatial genetic structure potentially being correlated with the morph phenotypes. Eight of the yellow-morph samples were from a single year because they were the same individuals used for obtaining RNA-Seq data, with the remaining two samples collected from different years. We also used ten samples of yellow-morphs with the blue-yellow phenotype (i.e. yellow over most of their throats except for a light blue patch at the base of the throat; scored as “by” according to the phenotype codes used by the Sinervo lab, Fig. S1C) to generate PCR amplicons used in PacBio sequencing.

Color phenotype data and analyses

Color measurements were taken with an Ocean Optics spectrometer (either a USB2G1740 or a JAZ model). We measured the coloration of the 24 lizards (8 of each morph) used to generate RNA-Seq data, as well as an additional 1 yellow, 4 orange, and 4 blue-morphs that were part of our genome wide association study. Color measurements taken from four regions of the throat (Fig. S1D) were averaged for subsequent analyses. We used PAVO (54) to plot the colors of the lizards in Commission Internationale de l'Éclairage (CIE) xyz color space along the x and y axes (these two axes are sufficient for capturing the color since $z = 1 - x - y$). For the x and y axes of coloration, we calculated the average Euclidean distance of each individual yellow-morph to all the orange-morph individuals and separately to all the blue-morph individuals. We then tested whether yellow-morphs with or without B-haplotypes differed in their distances to orange and blue-morphs in color space using an ANOVA in the car R-package (55).

Reference Genomes

Liver and tail tissue from a blue-morph was extracted with a modified salt extraction protocol (56) that included an overnight digestion along with 50 µg of RNaseA to allow for accurate DNA quantification. Illumina libraries were made following the standard manufacturer's protocol with insert sizes of 250, 500, 800, 2000, 5000, 10,000, and 20,000 bp. Libraries were sequenced with 150 bp PE reads for insert sizes 800 bp or less and with 50 bp PE reads for the larger insert sizes. Sequencing was done on an Illumina HiSeq 2000. The number of usable bases obtained was 123.6 Gb for the 250 bp insert size, 26.8 Gb for the 500 bp insert, 33.4 Gb for the 800 bp insert, 36.22 Gb for the 2000 bp insert, 14.15 Gb for the 5000 bp insert, 9.0 Gb for the 10,000 bp insert, and 9.62 Gb for the 20,000 bp insert. The genome was assembled with Platanus (57) with the default parameters. Kgf (v1.16) (58) and GapCloser (v1.10.1) (58) were used with default parameters to close intra-scaffold gaps using paired-end reads from the libraries that had insert sizes ≤ 800 bp.

Tail and heart tissue from an orange-morph was extracted with a Circulomics Big DNA Tissue Kit, which yielded DNA with an average size of 33 kb. A HiFi library prep of the sample was done at the University of California, Davis Genome Center (supported by NIH Shared Instrumentation Grant 1S10OD010786-01) where it was sequenced on one PacBio Revio 25M SMRT cell, which yielded 6,325,409 reads with a mean length of 11,424 bp and mean quality of 35. Adapters were removed with HiFiAdapterFilt (59) and the genome was assembled with hifiasm (60).

We searched for structural genetic changes (e.g. inversions) by aligning the contig with the *SPR* gene from the orange-morph genome to the scaffold with the *SPR* gene from the blue-morph genome with Lastz 1.04.22 with parameters gextend, chain, gapped, and strand=plus (61). We took the reverse complement of the orange-morph contig to have it match the overall orientation of the blue-morph scaffold before making a dotplot of the alignment. We performed a second Lastz alignment for the local area around the *SPR* region of interest (i.e. Fig. 1D), which started 10 kb before the morph-associated upstream region and ended 10 kb after the end of the coding region of *SPR*. Before the local area alignment, we filled in two gaps in the blue-morph genome sequence with PacBio amplicon data obtained for the *SPR* region for the blue-morph genome lizard (see below).

We searched for structural changes and homology between the *U. stansburiana* genomes and the divergent *Sceloporus undulatus* genome (62). We made Lastz alignments of the *SPR* regions in the orange and blue-morph genomes to the region with *SPR* in the *S. undulatus* genome using the same parameters for Lastz as described above. The *SPR* gene is at positions 192,627,655-192,632,06 on chromosome 5 of the *S. undulatus* genome. We downloaded from Genbank a one megabase (Mb) region of *S. undulatus* chromosome 5, which had the *SPR* gene at positions 627,655-632,064 within that region so that we could relate the alignment of the 1 Mb region back to the position of *SPR* in the full chromosome. Preliminary inspection of the alignments to *S. undulatus* revealed a small number of misalignments at the very beginning at very end of the alignments (for the orange-morph genome alignment there were 16 regions at the beginning and 15 at the end, and for the blue-morph genome alignment there were 26 regions at the beginning). These misalignments (likely due to alignments to transposable elements) were removed before visualizing the remaining regions. We visualized the alignment of the orange-morph genome to the 1 Mb *S. undulatus* region to assess large scale homology and any structural changes between the genomes. We then did a visualization of the alignment in a local region around the *SPR* region to assess whether there was homology between the *SPR* upstream and coding regions in the orange and blue-morph genomes to *S. undulatus* and whether either genome showed higher similarity to *S. undulatus*. To do the local visualization, we paired down

the alignment to the *S. undulatus* region to only include a region going from 15,000 bp upstream of the *SPR* gene (to account for the ~10,000 bp associated region upstream) to 5000 bp downstream of the *SPR* gene.

We visualized and annotated the local region around *SPR* by aligning regions in the orange and blue-morph genomes around *SPR* with Geneious Prime® 2024.0.7 (<https://www.geneious.com>) with a 65% similarity cost matrix in a global alignment, a gap open penalty of 30, and a gap extension penalty of 0. The location of the *SPR* gene was determined by alignment of RNA-Seq data (see below) and BLAST (63) searches of the genetic region in Genbank (64). We used RepeatMasker (65) to identify transposable elements in the orange and blue genome sequences using an rmbast search of vertebrates.

Genome resequencing data and analyses

DNA was extracted with the salt extraction protocol described above in the Reference Genomes section. Samples were standardized to 16 ng/μL DNA in 100 μL low-TE and sonicated for 12 minutes with a Bioruptor UCD-200 (Diagenode) machine set on the low setting for 30 seconds on and 30 seconds off. Sonicated DNA was run on an agarose gel to check if the average size was ~350 bp with the majority below 600 bp. Additional sonication was performed if necessary. Then 96 μL of the sonicated DNA underwent a double-sided size-selection with lab-made solid-phase reversible immobilization (SPRI) beads (66) with 0.5x beads used for right-side selection and 0.65x for left-side selection, which resulted in the maintained DNA being predominantly between 300-600 bp. Libraries were constructed with a KAPA Hyper Prep kit with quarter reaction sizes and Illumina unique dual indexes. Samples typically went through a single cycle of PCR, but three libraries went through two additional cycles of amplification. A pool was made with 10 ng of DNA from each sample, which was sequenced on an Illumina MiSeq Nano 150 bp PE run to check the pooling. The samples were repooled, checked again on a Nano run, and then sequenced on two 150 bp PE Illumina NovaSeq 6000 S4 lanes. Sequencing was done at QB3 Genomics at University of California, Berkeley (Berkeley, CA, RRID:SCR_022170).

We used HTStream v 1.3.0 (<https://github.com/s4hts/HTStream>) to: 1) calculate summary statistics about the reads, 2) exclude phiX sequences, 3) remove PCR duplicates, 4) remove adapter sequences and merge any overlapping reads, 5) filter out any remaining contamination using a file containing the Illumina adapter sequences and the indexes used for barcoding, 6) remove low quality bases with a window size of 10 bp and a quality threshold of 20, 7) trim off sequences containing N characters, and 8) remove any reads that were smaller than 20 bp. Cleaned reads were mapped to our reference genomes using BWA-MEM 0.7.17-r1188 (67). Samtools was used to generate, merge, sort, and index bam files of the mapped reads (68). Read groups were added with Picard v. 2.23.4 (<http://broadinstitute.github.io/picard/>) and reads surrounding indels were realigned with GATK v3.6 (69). Quality control was assessed with FastQC v 0.11.8 (70) on the unmapped reads, Qualimap v.2.2.1 (71) on the mapped reads, and MultiQC version 1.1 (72) was used to summarize the results of these programs. We observed an average coverage of 5.2x with a minimum of 3.5x and a maximum of 7.6x.

We use ANGSD v. 0.933-79-gda26ba4 (73) for a variety of population genetic analyses as described below. ANGSD was typically used with the following filters to focus on sites in the genome with high quality data. The filtering only included sites with a minimum mapping quality of 20, with a minimum base quality of 20, estimated to be polymorphic with a *P*-value of 0.000001, that had a minimum depth of 1 and maximum depth of 100 per individual, estimated to be in Hardy-Weinberg equilibrium with a *P*-value cutoff of 0.0001, and that were not triallelic.

The filtering also: 1) only considered pairs of reads where both reads mapped correctly, 2) only considered reads that mapped uniquely, 3) included the "remove bads" filter, 4) did base alignment quality (BAQ) calculations to consider whether there were alignment issues with the read (74), and 5) the map quality was adjusted to account for excessive mismatches by setting the -C flag to 50.

We used NgsRelate (75, 76) to determine the pairwise relatedness (77) among the 100 samples that we sequenced. We generated the input file for NgsRelate with ANGSD with our typical quality filters (as described above). In addition, for the NgsRelate analysis, we only considered sites with data from all 100 individuals and that had a minor allele frequency greater than or equal to 0.4, which was done to reduce the number of sites and thereby reduce the memory requirements for NgsRelate. The NgsRelate analysis revealed that two samples had a pairwise relatedness of 1, which indicated that they were likely from the same individual that was caught in two consecutive years. We only used one of these two samples in subsequent analyses, which led to a decrease in the orange-morph sample size from 45 to 44 individuals.

We performed a Principal Component Analysis (PCA) on the 99 samples of the morphs (*i.e.* 44 orange, 45 blue, and 10 yellow-morphs) with the ngsCovar function of ngsTools (78). The input genotype file for ngsCovar was generated in ANGSD with our typical quality filters plus additional filters to only consider sites with data from a minimum of 50 individuals, sites that had a minor allele frequency greater than or equal to 0.05, and genotypes were only called with a posterior probability above 0.95. We observed that individuals with pairwise relatedness values above 0.1 (as determined by NgsRelate) grouped together as outliers in plots of the first 12 principal components (Fig. S2). We therefore used the first 12 principal components as covariates in our genome-wide association tests (see below) to control for relatedness among samples.

We tested for an association between the morph phenotypes and single nucleotide polymorphisms (SNPs) in the genetic data with a method that takes uncertainty in genotype calls into account (79). This test was implemented in ANGSD with our typical quality filters plus additional filters to only consider sites that had a minor allele frequency greater than or equal to 0.05, genotypes were only called with a posterior above 0.95, and the sites needed to have data from at least 45 individuals for the test between orange versus blue-morphs and 28 individuals for tests of blue versus yellow and orange versus yellow-morphs. We converted the likelihood ratio statistic to *P*-values with the pchisq function in R (80). When making Manhattan plots of the full genome data, the sites were filtered to only include associations with a *P*-value < 0.001 for ease of plotting, but all sites were included for localized plots of the association (*i.e.* Fig. S4) and quantile-quantile plots (*i.e.* qqplots) of the distribution of the *P*-values (Fig. S3). We used the Dunn-Šidák method (81) to calculate the threshold *P*-value of significance based upon the total number of sites in each association test. We ordered the contigs by their size in the Manhattan plot that used the orange-morph genome. We visualized localized plots of the associations for: 1) the orange versus blue-morph phenotypes around the *SPR* region and 2) orange/blue versus yellow-morph phenotypes on Contig 12 where the beta-carotene oxygenase 2 (*BCO2*) gene is located, which is associated with yellow-morphs in *Podarcis muralis* (8). We determined the location of the *BCO2* gene by a BLAST search of the mRNA of *BCO2* from the *Sceloporus undulatus* genome (Genbank locus ID: XM_042474882) against the *U. stansburiana* orange genome.

We used ANGSD to obtain estimates of genetic diversity along the genome. We first generated per site allele frequency likelihoods with most of our typical quality filters, except that no polymorphism filter was done (to preserve low frequency variants) and BAQ calculations were omitted (to speed up the computational time). We then obtained a maximum likelihood

estimate of the site frequency spectrum from 200 million sites in the genome. The site frequency spectrum was used to calculate the nucleotide diversity (the average number of pairwise differences, π), which was calculated in 10 kb windows with a step size of 10 kb. We filtered out windows that had fewer than 8000 sites with data within the window.

PacBio amplicon data and analyses

We designed primers to amplify the *SPR* haplotypes via long-range PCR (forward primer: /5AmMC6/gcagtcgaacatgtagctgactcaggtcacGCATGTAGAAGAGCAAGCCCCAAT; reverse primer:

/5AmMC6/tggatcactgtgcaagcatcacatcgtagCGGCGTTGTTGACCAGGAGGAG).

The forward primer was designed to a highly conserved (i.e. only 1 bp different) region of DNA between the lizard *Sceloporus undulatus* (62) and *U. stansburiana* as revealed in a Lastz (61) alignment of the scaffolds containing the *SPR* gene in each of these species. The reverse primer was placed at the end of the first exon of the *SPR* gene. Both primers included additions for PacBio amplicon sequencing (indicated in the primers above by the lower-case text and text between the backslashes). Primer characteristics such as melting temperature were assessed with Primer3 (82) in Geneious. PCR reactions were 10 μ l in total volume and contained 1X PrimeSTAR GXL Premix (Takahara Bio), 0.3 μ M of each primer, and 20 ng of genomic DNA. A touchdown protocol was used that began with an initial denature at 95°C for 5 minutes, then had 20 cycles of 98°C for 10 seconds, an annealing temperature that began at 68°C for 5 seconds, but that decreased by 0.5 °C per cycle, and an elongation at 72°C for 9 minutes. Following the touchdown phase were 20 additional cycles where the annealing temperature was kept constant at 58°C. The PCR ended with a final hold at 72°C for 20 minutes. PCR products were assessed on an agarose gel and reamplified if there was poor amplification or only a single band was observed in the initial PCR. The PCR products were then cleaned with SPRI beads and dual-indexed by a second 20 μ l PCR reaction with similar PCR conditions except that 5 μ M of each indexing primer was used and the PCR reaction was done for just 20 amplification cycles. Sets of 96 PCR reactions were pooled together by combining equal masses of DNA from each PCR reaction. Then equal masses of DNA from each pool were combined into a final pool. DNA fragments less than 5000 bp were removed by a 3.1X diluted Ampure-bead size-selection. The SMRTBell 3.0 Prep Protocol was used to make a library that was sequenced on one PacBio Revio lane, which gave 5.5 million HiFi reads and yielded 43 Gb of HiFi data.

We demultiplexed the PacBio amplicon data with lima 2.1.0 (available from <https://github.com/PacificBiosciences/barcoding>). We then used pbAA 0.1.3 (<https://github.com/PacificBiosciences/pbAA>) to cluster the data for each individual and generate consensus sequences, with two divergent O-haplotypes and one B-haplotype used as reference sequence guides for clustering. All consensus sequences passing the pbAA filters were aligned to the orange or blue reference sequence and to each other as necessary to assess their quality. Consensus sequences were filtered out if they were virtually identical to another consensus sequence (typically only differing in 1-2 positions), with the consensus sequences with the highest read count kept in the data set.

We could determine the two diploid haplotypes for most of the lizards with PacBio amplicon data after vetting the PacBio data along with the Illumina short-read data. To do this, we mapped the Illumina data to two different versions of the blue-morph genome where we used either the O-haplotype or the B-haplotype in the *SPR* region. This allowed us to examine coverage across the *SPR* region separately for each haplotype for each individual lizard. We checked our PacBio data against the Illumina resequencing data when there were too many or too few haplotypes after clustering the PacBio data. In the rare case of too many haplotypes, the

short read data could help filter out contaminant sequences or chimeras, which were haplotypes that did not match the sequence variants observed in the Illumina data. The most common issue was that only a single haplotype passed clustering, in which case the short read data could help determine if a haplotype had been erroneously placed in the “failed cluster sequences” file. This issue likely occurred because PCR reactions preferentially amplify shorter haplotypes. We observed that longer haplotypes often wound up with fewer reads than shorter haplotypes, which sometimes led to them having fewer reads than necessary to pass the clustering threshold used by the pbaa program to call a sequence. For some individuals, we could only confidently score a single haplotype, which could be the consequence of that individual having two identical haplotypes at the SPR region, insufficient data to reconstruct both haplotypes, or other issues like preferential PCR of shorter haplotypes.

We determined the SPR region genotypes for the lizards used in our genome-wide association study plus the individual lizard used to generate the blue-morph genome. The genotypes were determined for each individual lizard by looking at both their PacBio haplotype(s) for the SPR region as well as the coverage patterns of the Illumina data separately mapped to the O- and B-haplotypes. With the Illumina data, individuals homozygous for the O-haplotype had gaps in coverage in regions of the B-haplotype that are not found in the O-haplotype. Individuals homozygous for the B-haplotype could be assessed in a similar way. Heterozygous individuals had coverage across both the O- and B-haplotypes. We were able to determine diploid genotypes for 99 out of these 100 lizards, with the one exception being a blue-morph that had at least one blue haplotype, but for which we were unable to confidently resolve its diploid state. Of these 100 lizards, we had diploid PacBio haplotypes for 69 of them and genotyped the remaining 31 lizards from both the PacBio haplotype and the Illumina data. We also determined diploid genotypes for ten yellow sneaker morphs with the blue-yellow phenotype, for which we only had PacBio amplicon data.

We obtained 156 haplotypes of the *SPR* region (48 B-haplotypes and 108 O-haplotypes) from the lizards in the genome-wide association study (GWAS), which we aligned with MAFFT v7.490 (83, 84) with a gap open penalty of 3 and an offset of 0.123. The alignment of the two haplotypes was quite difficult due to the high level of divergence between the haplotypes, so the number and position of gaps was dependent on the alignment settings. We visualized the patterns of variation across the haplotype alignment by classifying each position in the alignment into the following categories: 1) an invariant position, 2) a gap position if one haplotype group had only a gap and the other haplotype group had 50% or more individuals with a gap, 3) a position unique to either the O or B-haplotypes if one haplotype group had a nucleotide and the other group had only a gap, 4) a fixed difference if each haplotype group had a single invariant nucleotide, but these nucleotides differed from one another, 5) a shared polymorphism if both sets of haplotypes had the same biallelic nucleotide polymorphism, and 6) a complex variable position for the remaining positions, such as when there was variation in both haplotypes, but the variants differed between the haplotype groups. In general, we plotted both the shared polymorphism and complex variable positions together as a single “variable positions” category, unless we were specifically plotting just the shared polymorphic sites.

We used the R-packages ape version 5.6 (85) and pegas version 1.1 (86) to calculate nucleotide diversity (π) in 1000 bp windows with a step size of 100 bp amongst: 1) just the 48 B-haplotypes, 2) just the 108 O-haplotypes, and 3) for all 156 haplotypes combined. We excluded windows in which there less than 500 bp of nucleotide data in the alignment that was being analyzed (e.g. for the alignment of B-haplotypes, a window was excluded if the number of gap positions plus the number of O-haplotype unique positions was 500 or greater). We also calculated the number of shared polymorphic positions and fixed nucleotide differences between

the two haplotypes within each window. We conservatively did not score O and B-haplotype unique positions as fixed differences, because many of them are likely not independent mutational events because they result from the insertion of transposable elements. When plotting the number of fixed differences, we excluded windows in which there less than 500 bp of nucleotide data in the alignment. In addition to plotting the number of fixed differences itself, we divided the number of fixed differences by the estimate of π to highlight regions that had high levels of sequence divergence between the haplotypes relative to the levels of genetic variation observed in the region. We plotted number of shared polymorphic positions as a way to assess recombination, because the presence of shared polymorphism between the O and B-haplotypes indicates that there has been recombination between the haplotypes. After observing that a large region of the alignment lacked any shared polymorphism, we designated the non-recombining region as starting from the first fixed difference following where shared polymorphism stopped being observed and ending with the last fixed difference before where shared polymorphism was again observed.

We inferred a phylogeny for the 156 haplotypes of the *SPR* region. We first removed large regions in the alignments that were only found in one of the haplotypes and areas that were difficult to align between the highly divergent O and B-haplotypes. From this paired down alignment, we also created a second dataset to generate a smaller exemplar phylogeny for the main paper (i.e. Fig. 2G), which consisted of a sample of 14 haplotypes each from orange, blue, and yellow-morph individuals. In the second dataset, we primarily took the haplotypes from individuals with transcriptome data, but we also used the haplotypes from the blue-morph individual used to generate a genome and two yellow-morphs that were used in the GWAS, but that did not have transcriptome data. We inferred a maximum likelihood phylogeny from each of these two datasets using RAxML version 8.2.12 (87) with the GTRCAT model, 100 alternative runs on different starting trees, and rapid bootstrapping. We mapped the morph phenotype associated with each haplotype on to the phylogeny using the dotTree function in phytools (88), along with functions from ape (85) and plotrix (89).

We used the same 156 haplotype dataset that was paired down to alignable regions between the O and B-haplotypes to calculate genetic diversity (π) for the entire *SPR* region with the R-package pegas (86). This estimate of π for the entire *SPR* region from the haplotypes was plotted relative to the levels of π estimated from the Illumina data (see above). The PacBio haplotypes were necessary to obtain an accurate estimate of genetic diversity of the *SPR* region because the structural genetic changes near *SPR* meant that many of the short Illumina reads could not be aligned to the reference genome. Note that structural differences elsewhere in the genome are likely to face similar mapping issues. Therefore, estimates of π from the Illumina data could be lacking accurate estimates of π from other regions with structural genetic differences, which could potentially affect the outlier status of the *SPR* region.

We evaluated levels of linkage disequilibrium among SNPs in our alignment of the 156 haplotypes of the *SPR* region. The alignment was transformed into a vcf file using SNP-sites (90). The r^2 measure of linkage disequilibrium was calculated in quickLD with ploidy set to haploid as appropriate for the haplotype data (91). Linkage disequilibrium was only measured for biallelic SNPs and we filtered the data to include only SNPs with a minimum frequency of 0.2. Levels of linkage disequilibrium among SNPs were plotted with LDheatmap (92).

We conducted association tests between the orange and blue-morph phenotypes and three structural genetic differences (indels) found in the first exon of the *SPR* gene. These tests were conducted on 27 orange and 32 blue-morphs that each had data for two PacBio haplotypes, which allowed us to score the diploid genotypes of the structural variants for these individuals. We binned the data so that individuals heterozygous for the indel were grouped with one of the

homozygotes in a way that would maximize differences between the morphs if one variant exhibited dominance to the other. We similarly scored the genotypes at the SNP in the upstream region to *SPR* with the strongest association to phenotype for the same 27 orange and 32 blue-morph individuals. We used Fisher's exact tests to evaluate an association between morph phenotype and the binned genotypes.

We assessed whether there were differences between genotypes in the propensity to develop into particular morphs, which could indicate differences in plasticity. Specifically, we tested whether blue and yellow-morphs differed in their frequencies of BB and OB *SPR* genotypes. We did not consider OO genotypes because the three anomalous blue-morphs with OO genotypes may be erroneous (see Supplemental Text). For yellow-morphs, we summed the numbers of BB and OB genotypes for both the all-yellow and blue-yellow phenotypes. We used a Fisher's exact test to assess differences in frequencies between the groups.

We assessed whether the three morphs and individuals with different *SPR* genotypes differed in size. We specifically tested whether there were differences in body size ($\log_{10}$ snout-vent length in millimeters (SVL)) and standardized weight ($\log_{10}$ weight in grams divided by $\log_{10}$ SVL) among the morphs and *SPR* genotypes. The two size metrics were log transformed to have the data fit a normal distribution. We were particularly interested in assessing whether the three anomalous OO blue-morphs and the four OO yellow-morphs differed in size from other morphs and genotypes to determine if they were comparable to other lizards with their same morph/genotype. We analyzed the two different phenotypes of yellow-morph (i.e. blue-yellow and all-yellow phenotypes) as separate groups. We conducted these test with type-II ANOVAs performed by the function Anova in the car package (55) with post-hoc Tukey tests performed by the multcomp package (93).

RNA-Seq data and analyses

We quantified gene expression of the *SPR* gene in throat skin and brain using RNA-Seq. For the throat skin, our primary sample was in the main color patch on the throat where there was intense orange, blue, or yellow coloration (regions 1-4, Fig S1D). However, we also sampled skin just outside of the colored region, which was typically white in color, but which sometimes had a very faint amount of the color found in the main color patch. This second skin sample was collected to determine if differential expression of the *SPR* gene was primarily localized to the skin expressing coloration. We extracted RNA from the entire brain, which therefore provided a general measure of expression levels across the whole brain.

RNA was extracted using a New England Biolabs Monarch Total RNA Miniprep kit. Samples were added to a tube with a 4 mm steel bead and homogenized in a PowerLyzer homogenizer set at 3500 rpm for 2 cycles of 45 seconds on, with 30 seconds off between cycles. We doubled the amount of Proteinase K as recommended to increase yield. All samples were filtered through a DNA exclusion column and treated with DNase I. RNA quality was checked for all samples on a Bioanalyzer and skin samples had an average RIN score of 9.0 (range 7.3-9.8) and brain samples had an average score of 8.5 (range 7.8-9.1).

Libraries were prepared with a KAPA (Roche) stranded mRNA-Seq Kit using half-size reactions. Typically, 1 μ g of RNA was used for brain samples and 500 ng of RNA for skin samples. However, one brain sample and five skin samples had lower extraction yields so less input RNA was added (brain sample = 486 ng, skin samples range = 310-478 ng). During the library preparation, samples were fragmented by incubating them at 94°C for four minutes. Unique dual indexes were added with 12 cycles of PCR for brain libraries and 15 cycles for skin libraries. Libraries were normalized for pooling by a MiSeq Nano test run and then run on one NovaSeq X 25B 150 bp PE lane.

We used HTStream v 1.3.3 (<https://github.com/s4hts/HTStream>) to: 1) calculate summary statistics about the reads, 2) exclude phiX sequences, 3) remove PCR duplicates, 4) trim adapters, 5) remove any remaining sequences containing the Illumina adapter sequences, 6) remove polyA/polyT tails, 7) trim off sequences containing N characters, 8) remove low quality bases with a window size of 10 bp and a quality threshold of 20, and 9) remove any reads that were smaller than 50 bp. Cleaned reads were mapped to our orange-morph reference genome with STAR version 2.7.11b (94). We took the number of reads mapping to *SPR* on the reverse strand as quantified by STAR and divided the *SPR* read counts by the total number of uniquely mapped reads across the genome (in millions) to get a standardized metric of the number of fragments mapped to *SPR* per million mapped reads. This metric was $\log_2$ transformed. For the skin data, we calculated an additional metric, which was the number of fragments mapped to *SPR* per million mapped reads in colored skin divided by the number of fragments mapped to *SPR* per million mapped reads in noncolored skin. This metric (i.e. the ratio of expression in colored to noncolored skin) shows whether the expression of *SPR* is elevated in the colored skin. We separately tested whether *SPR* expression levels differed among morphs or differed among genotypes of the *SPR* region haplotypes with type-II ANOVAs performed by the car package (55) with post-hoc Tukey tests performed by the multcomp package (93).

We quantified the levels of expression of the O and B-haplotypes in individuals that were heterozygous for the two haplotypes (i.e. 8 blue-morphs and 2 yellow-morphs). This is similar to quantifying allele-specific expression, which is used to uncover *cis*-acting regulation of genes (38). Allele-specific expression was not an option because there were not any SNPs in the coding region of the *SPR* gene that were perfectly correlated with the two divergent haplotypes upstream of the gene. However, we could quantify the number of reads that mapped to the orange and blue haplotypes as follows. We first deleted the sequence for the *SPR* coding sequence, untranslated regions (UTRs), and the divergent region upstream of the *SPR* gene from the orange-morph reference genome. Then we added in an O and B-haplotype from our PacBio amplicon data, which included the divergent upstream region, the 5' UTR, and most of the first exon of the *SPR* gene. For seven out of the ten heterozygous individuals we were able to add in the exact haplotypes found in each individual lizard, because we had obtained diploid PacBio data for these individuals. This was the ideal situation for mapping the RNAseq data, because there would not be any discrepancies between the RNAseq reads and the two reference haplotypes. For the remaining three individuals, we used haplotype reference sequences of the same size from the orange and blue-morph reference genomes. We used STAR to map the RNAseq data to these reference genomes plus individualized *SPR* haplotypes and quantified expression on each haplotype in a 700 bp window that included the 5' UTR and most of *SPR* exon 1. In the reference haplotype from the orange-morph genome, the coding sequence was 349 bp and the 5' UTR was 269 bp. However, the length of the expressed region could vary among individuals due the presence of indel variants, so we used a 700 bp window to account for any length differences. We then used a paired t-test to evaluate if there was a difference in the $\log_2$ transformed number of fragments mapped to each haplotype per million mapped reads. We also used a paired t-test to evaluate whether the expression of each haplotype (O or B-haplotype) differed between colored and noncolored skin. We wanted to qualitatively evaluate if sequence variation among O and B-haplotypes might result in differences in gene expression (e.g. we wanted to discern whether particular clades of O-haplotypes generally had lower expression of *SPR* than other clades). To do this, haplotype-specific expression of *SPR* in colored skin was plotted on to a phylogeny of the haplotypes using the plotTree.barplot function in phytools (88). Expression data was only plotted for the seven heterozygous individuals that had both of their

haplotypes sequenced so that expression data could be associated with an exact haplotype in the phylogeny.

Field data on the polymorphism

We utilized previously reported survey data on the throat color polymorphism at the Los Banos field site (17). The numbers of each morph were based on the prior groupings of phenotypes (according to the phenotype codes used by the Sinervo lab, see Supplemental text), with orange-morphs = *oo*, *ob*, and *oy* phenotypes, blue-morphs = *bb* phenotypes, and yellow-morphs = *yy* and *by* phenotypes (31, 37, 95). We calculated the proportion of blue-yellow males that make up the yellow sneaker morph category by dividing the *by*-phenotyped individuals by the number of (*by* + *yy*) phenotyped individuals. Each phenotype was totaled across all years (i.e. 1990-2012) and locations of sampling.

To determine the morph frequencies within each year, we utilized the field data from 1999-2012 (17). Prior to 1999, the field surveys were only done on lizards from a single rock outcrop, but beginning in 1999 more extensive surveys were done, which included data from four different rock outcrops at the field site. We therefore utilized the data from 1999-2012 to focus on the years with the most data and with comparable degrees of sampling within a year. We summed the number of individuals of each morph across all four outcrops before calculating the frequency of each morph within a year. We used a linear model in R (80) to test for a correlation in the frequencies of the morphs. We used the *acf* function to assess autocorrelation in the residuals, but did not observe any substantial autocorrelation (i.e. all *acf* values were less than 0.5). We also used the field data from 1999-2012 to estimate the proportion of sneaker morphs generated from B-haplotypes, assuming the one-locus two-allele model with plasticity. To calculate the sneaker morph proportion, we divided the total number of yellow-morph individuals by the total number of blue-morph + yellow-morph individuals observed across all years and outcrops.

Models of the maintenance and stability of the polymorphism

We explored whether and how rock-paper-scissors dynamics could maintain a two-allele genetic polymorphism with two of the three players arising via irreversible phenotype plasticity and the third player arising from a genetically determined difference. Our models had discrete time for reproduction, nonoverlapping generations, and Hardy-Weinberg equilibrium for the initial frequencies of the genotypes. We made the simplifying assumption that plasticity is only observed in individuals with a B-haplotype (i.e. OB and BB genotypes), because a B-haplotype was observed in the vast majority of individuals scored as either a yellow or a blue-morph. The yellow-morph sneaker phenotype was determined by a proportion (*p*) of all individuals carrying a B-haplotype and the blue-morph mate-guarder phenotype by a proportion ($1 - p$) of all individuals carrying a B-haplotype. The orange-morph phenotype was based upon the frequency of homozygotes for the O-haplotype (OO genotypes). We first made the simplifying assumption that the proportion of yellow and blue-morphs was fixed within and between generations (fixed plasticity models). However, in the field some yellow-morph individuals can irreversibly transform into the blue-morph phenotype, likely in response to the availability of a territory (14, 21), which indicates that there could be some environmental or conditional dependence to the plasticity. Therefore, we also explored a model in which the proportion of yellow and blue-morphs was set to a level that optimized the fitness of individuals with the B-haplotype, given the orange-morph frequency in a particular generation (optimal plasticity models, see below).

Selection in the models arose from fitness differences amongst males that were dependent on the frequency of other male morph phenotypes. The measures of frequency-dependent fitness of the orange territory usurpers (W_O), blue mate guarders (W_B), and yellow sneaker morphs (W_Y) were taken from prior field studies that estimated male fitness from either overlap with female territories (6, 19) or from genetic paternity estimates of siring success (15, 19). For estimates of male fitness from overlap with female territories, we took the values of the payoff matrix estimated by Sinervo and Lively (1996) that were reported in (96). For estimates of male fitness from genetic paternity data, we estimated the values of the payoff matrix from Fig. 2B of Sinervo (2001) (19). We did not use the payoff matrices estimated by Friedman et al. (2017), which were estimated under the assumption that there were three genetically determined alleles underlying the morphs (i.e. replicators) (17). Below are the payoff matrices.

Sinervo & Lively 1996 Fitness of	Competing with		
	Orange Usurper	Blue Guarder	Yellow Sneaker
Orange Usurper	1	4.31	0.4
Blue Guarder	0.2	1	1.8
Yellow Sneaker	1.76	1	1

Sinervo 2001 Fitness of	Competing with		
	Orange Usurper	Blue Guarder	Yellow Sneaker
Orange Usurper	1	1.75	1.2
Blue Guarder	0.85	1	1.95
Yellow Sneaker	2.2	0.75	1

A rock-paper-scissors game is mathematically represented by a payoff matrix in which each strategy has high fitness against one other strategy, but low fitness against a third strategy. The two payoff matrices for side-blotched lizards each have one small deviation from this expectation. With the payoff matrix estimated by Sinervo and Lively (1996), a rare yellow-morph in an environment where blue-morphs are common has a fitness of 1, rather than a fitness value lower than 1 as expected for a rock-paper-scissors game. With the payoff matrix estimated by Sinervo (2001), a rare orange-morph in an environment where yellow-morphs are common has a fitness of 1.2, rather than a fitness value lower than 1 as expected for a rock-paper-scissors game. A variety of explanations are possible for these deviations. One possibility is that the deviations represent real aspects of the biological interactions among the morphs (e.g. perhaps the fitness of 1 for a rare yellow-morph competing against a common blue-morph is due to them sharing a common genetic basis). Another possibility is the system follows a true rock-paper-scissors game but there is estimation error for the exact values in the payoff matrices, because each matrix was estimated using field data. In line with this, if the two matrixes are averaged, the fitness values would conform to the full mathematical expectations of a rock-paper-scissors game. We ran simulations using both matrixes to assess how our results changed when there were different fitness payoffs among the morphs. However, we primarily used the paternity data matrix for the main paper, because the paternity data provides a more direct measure of male fitness than territory overlap with females.

The fitness of each phenotype/strategy was calculated from the frequencies of each morph (F) times the payoff of the focal morph-phenotype competing against the other morphs as follows (using the Sinervo 2001 paternity data matrix in the example):

$$\begin{aligned} W_O &= F_O + 1.75F_B + 1.2F_Y \\ W_B &= 0.85F_O + F_B + 1.95F_Y \\ W_Y &= 2.2F_O + 0.75F_B + F_Y \end{aligned}$$

Where the frequencies of the three morph phenotypes were determined by the frequencies of the genotypes (f) and the proportion of yellow-morphs (p) as:

$$\begin{aligned} F_O &= f_{oo} \\ F_B &= (1 - p)(f_{bb} + f_{ob}) \\ F_Y &= p(f_{bb} + f_{ob}) \end{aligned}$$

The expected genotype frequencies after selection (f') in males were calculated as the frequency of the genotype before selection multiplied by the fitness of the mating strategies that arise from that genotype (W_{strategy}), divided by the mean fitness of all the genotypes ($\bar{W}$):

$$\begin{aligned} f'_{oo} &= \frac{f_{oo} W_O}{\bar{W}} \\ f'_{ob} &= \frac{(1 - p) f_{ob} W_B + p f_{ob} W_Y}{\bar{W}} \\ f'_{bb} &= \frac{(1 - p) f_{bb} W_B + p f_{bb} W_Y}{\bar{W}} \end{aligned}$$

where

$$\bar{W} = f_{oo} W_O + (1 - p) f_{ob} W_B + p f_{ob} W_Y + (1 - p) f_{bb} W_B + p f_{bb} W_Y$$

The expected allele and genotype frequencies in the offspring generation were calculated assuming random mating between males with genotype frequencies f' and females with genotype frequencies f :

$$\begin{aligned} f_{oo.\text{new}} &= (f'_{oo} + f'_{ob}/2)(f_{oo} + f_{ob}/2) \\ f_{ob.\text{new}} &= (f'_{oo} + f'_{ob}/2)(f_{bb} + f_{ob}/2) + (f'_{bb} + f'_{ob}/2)(f_{oo} + f_{ob}/2) \\ f_{bb.\text{new}} &= (f'_{bb} + f'_{ob}/2)(f_{bb} + f_{ob}/2) \end{aligned}$$

We first explored whether both alleles could be maintained at appreciable frequencies across varying proportions of yellow-morphs ranging from 0.01 to 1, in increments of 0.01, using deterministic iteration of the equations given above. We ran the deterministic simulations for both 100 and 10,000 generations and for three different starting conditions for the allele frequencies: 1%, 50%, and 99%. We plotted the allele and phenotype frequencies against the proportion of the yellow-morph that was used for each of the different starting conditions and number of generations.

We also explored a model in which the proportion of yellow and blue-morphs in each generation was set to an optimal level given the frequency of the orange-morphs. The optimal proportion of yellow and blue-morphs is achieved when the fitness of blue-morphs (W_B) is equal to the fitness of yellow-morphs (W_Y). With the Sinervo 2001 paternity data matrix, this means:

$$0.85F_O + F_B + 1.95F_Y = 2.2F_O + 0.75F_B + F_Y$$

After some rearranging and substituting in the genotype frequencies this becomes:

$$1.35f_{oo} = 0.95(1 - f_{oo})p + 0.25(1 - f_{oo})(1 - p)$$

This leads to the solution:

$$p_{\text{optimal}} = -0.25 + \left(\frac{10}{7}\right)\left(\frac{1.35f_{oo}}{1 - f_{oo}}\right)$$

Similarly, the solution to the Sinervo and Lively 1996 matrix is:

$$p_{\text{optimal}} = \frac{1.95f_{oo}}{1 - f_{oo}}$$

Any values of $p_{\text{optimal}} < 0$ were set to zero and any values of $p_{\text{optimal}} > 1$ were set to one. We ran the deterministic simulations using p_{optimal} for three different starting conditions for the allele frequencies: 1%, 50%, and 99%.

The use of p_{optimal} in the model provides some environmental dependence to the proportion of sneakers akin to what has been observed in the lizards (14, 21). The plasticity of the lizards in the field is likely to be somewhere between the two forms of plasticity used in our models (i.e. fixed versus optimal proportions of blue and yellow-morphs). A potential reason for the lizards to resemble the fixed plasticity model would be if a relatively stable proportion of lizards become yellow-morphs each year due to condition-dependent differences among the lizards and a limited number of territories in the environment (e.g. if the average population size across years is 100 males and there are 50 territories, these territories would be filled by the 50 males in the best condition that become blue-morphs, leaving 50 males as yellow morphs). However, the number of blue versus yellow-morphs may also dynamically change within a year, as in the optimal plasticity model, due to differences in density of lizards among years and local openings of territories due to mortality of the territory holder. Although the optimal plasticity model allows for the dynamic transformation from yellow to blue-morphs that has been observed in the field (14, 21), the model is based upon a global optimum of fitness based on the population-level frequency of the morphs, whereas in the field there is likely only local optimization of morph frequencies. Given these considerations, we explored a set of semi-optimal models for the frequencies of the blue and yellow-morphs in which the optimal proportion of the two morphs was adjusted by a constant value (i.e. $p_{\text{optimal}} + c_{\text{morph}}$) to maintain a certain proportion of one morph. We explored different values of c_{morph} in the models. Positive values of c_{morph} led to an increased fixed proportion of the yellow-morph whereas negative values led to an increased fixed proportion of the blue-morph.

We assessed how well the two-locus models with plasticity and the one-locus three-allele model maintained the polymorphism when there was genetic drift. The three-allele model had the same structure as the two-allele model except that the frequencies of the three morph phenotypes were determined based on the prior hypothesis that the yellow allele was dominant to the blue allele and the orange allele was dominant to the other two alleles in determining the mating strategies of the morphs (37, 95):

$$F_O = f_{oo} + f_{ob} + f_{oy}$$

$$F_B = f_{bb}$$

$$F_Y = f_{by} + f_{yy}$$

Genetic drift was simulated using multinomial sampling of offspring genotypes using the expected offspring genotype frequencies ($f_{oo.\text{new}}$, $f_{ob.\text{new}}$, $f_{bb.\text{new}}$) and a fixed population size. The simulations were started with equal allele frequencies. We varied the population size from 50-250 individuals with a step size of 10. Stability was assessed by determining the proportion of 1000 trials in which all alleles were maintained after 1000 generations for simulations with the paternity data fitness matrix and after 100 generations for simulations with the territory overlap fitness matrix. For simulations with the paternity data fitness matrix, we used yellow-morph

proportions from 0.2-0.9, which bracketed the range of situations in which the two-allele model was more stable than the three-allele model. For simulations with the territory overlap fitness matrix, we used yellow-morph proportions from 0.5-0.95, given that the polymorphism is maintained under a more restrictive set of proportions with this fitness matrix.

We used the two and three-allele models to generate predictions about the expected relationships between the frequencies of the morphs. We simulated data with the three-allele model for 10,000 generations with a population size of 250 individuals, an equal starting frequency of all three alleles, with genetic drift, and with fitness values derived from genetic paternity estimates of siring success (15, 19). We then only used data from every 100 generations to avoid potential autocorrelation in the data. We used linear models to determine the slope and intercept of the correlation of the frequencies of the different morphs, which formed the basis of the predictions for the three-allele model. The relationships among the frequencies of the morphs could be obtained analytically for the two-allele model with fixed plasticity. The frequencies of all the morphs could all be related to each other based on the equations for the phenotypes of the morphs given above (*i.e.* F_O, F_B, F_Y) and that $F_O = f_{oo} = 1 - (f_{bb} + f_{ob})$, which leads to the following relationships:

$$F_Y = \frac{p}{1-p} F_B$$

$$F_O = 1 - \frac{1}{1-p} F_B$$

$$F_O = 1 - \frac{1}{p} F_Y$$

Data analysis and visualization

The data were analyzed in R (80) and visualized using the packages dplyr (97), scales (98), ggplot2 (99), ggthemes (100), and gridExtra (101).

Supplementary Text

Morph phenotypes

We measured the coloration of 12 orange, 12 blue, and 9 yellow-morphs with a spectrometer to provide detailed data on their phenotypes. The three morphs exhibited distinct throat colors, with the orange and blue-morphs being the furthest apart in color space (Fig. S1 E, F).

Prior genetic model and prior genetic evidence

The mating strategies and throat color polymorphism of *Uta stansburiana* were hypothesized to be governed by three alleles at a single locus (30, 31, 37, 95, 102). Under this three-allele model, putative alleles were co-dominant on throat color so a lizard phenotype was scored as *ob* if it had an orange and blue throat, *by* if it had blue and yellow, and *oy* if it had orange and yellow. The *o* and *y* alleles were hypothesized to be dominant to *b* in their effects on male behavioral strategy: Orange strategy = *oo, ob, oy* phenotypes; Blue strategy = *bb*, Yellow strategy = *yy, by*.

Prior field studies suggested that there were genetic differences that gave rise to the morphs. Here we summarize the two studies that provided the best estimates of the heritability of the color polymorphism (30, 31). The first study reported high heritability of orange ($h^2 = 1.06$) and blue ($h^2 = 1.40$) coloration from sire-progeny regressions (31). Heritability estimates for yellow-morph individuals were not reported. However, yellow-morph individuals were included in the regressions for heritability of orange or blue, with the pure-yellow individuals scored as having zero blue or orange "alleles" and the blue-yellow individuals scored as having 0.5 blue "alleles." Thus, these analyses were conducted under the hypothesized one-locus three-allele model and did not explicitly test for the heritability of the yellow-morph. The second study applied an "animal model" approach to estimate heritability from pedigree data collected for lizards in the field (30). This study reported that $h^2 = 0.46$ for orange coloration for data from both males and females. No estimates of heritability were reported for yellow or blue.

Previous studies of the polymorphism were more explicit about the plasticity of the yellow-morph than its heritability. Sinervo et al (2000) reported that: 1) some individuals began the breeding season with a yellow throat early in the breeding season but then acquired a small blue patch of color on their throats later in the breeding season, 2) males that developed blue coloration had higher testosterone levels than males that kept all-yellow throats, and 3) transformation from yellow to blue was more likely for males that had an opportunity to acquire a territory (14). Mills et al. 2008 studied the effects of exogenous gonadotropins (follicle-stimulating hormone and luteinizing hormone) on the expression of morph-related traits and found that treatments with gonadotropins: 1) increased blue throat color in both all-yellow and blue-yellow males, 2) increased endurance and sprint speed in all-yellow males, which could enhance their ability to defend territories, and 3) increased courtship behaviors directed at all-yellow and blue-yellow males (21). Mills et al. 2008 also found that yellow-morphs had considerable flexibility in their physiology and postulated that yellow-morphs "may adopt a season-dependent strategy with plasticity from nonterritorial to territorial strategies."

We see four reasons why the one-locus three-allele model was adopted for the throat color polymorphism in *U. stansburiana* despite a growing body of evidence supporting a role for plasticity. First, it was assumed that the rock-paper-scissors game required three genetically determined strategies that can replicate themselves across generations (6, 17, 19). The first model of the rock-paper-scissors game for *U. stansburiana* was based on game theory in which strategies perfectly replicate themselves, which could occur for asexual species or for alleles on a Y-chromosome (6, 19). A natural way to translate this simplified model to a diploid sexual system is to model the polymorphism with one locus that has three alleles, with some alleles that are dominant to the others, which was done by Sinervo (2001) (19). The assumption of morph strategies being replicators and breeding true has continued to be used in more recent models of the polymorphism in *U. stansburiana* (17), as well as general models of the stability of alternative reproductive strategies (51) and general models of the rock-paper-scissors game (103). Second, multiple different genetic models were explored for the polymorphism in *U. stansburiana* (19). For example, a one locus, two allele model was considered in which two morphs were generated from the homozygotes and a third morph was generated from the heterozygotes. This model was rejected because fixation of the alleles was common and frequency cycles of the alleles were not observed. The one-locus three-allele model was favored over other models because the polymorphism could be maintained and there were cycles in allele frequency that were similar to changes in morph frequency that were observed in the field. Plasticity of the yellow-morph was considered by Sinervo (2001) (19), but only in the context of the one-locus three-allele model. Third, *U. stansburiana* sometimes exhibit throats with multiple colors (e.g. blue and yellow). The individuals with multiple colors were hypothesized to be

heterozygotes for codominant color alleles (31). Fourth, determining the genetics of the throat color polymorphism was quite difficult with the methods that were available prior to our ability to collect genomic data. Controlled crosses amongst morphs were performed in the lab, but the progeny failed to express throat colors in the lab setting. Therefore, the progeny of controlled crosses had to be released back into the field to score coloration, but mortality in the field drastically reduced sample sizes. For example, of 410 progeny released in 1999, only 34 could later be scored for throat coloration (31). Although few offspring survived, this experiment was able to show that orange and blue coloration was heritable. However, the inheritance of yellow was likely difficult to study because of: 1) the potential plasticity of this morph and 2) that sneaker morphs do not defend a territory and are therefore more difficult to capture, which may have exacerbated low sample size issues.

In summary, multiple factors likely led to the adoption of the one-locus three-allele model as an explanation for the throat color polymorphism in *U. stansburiana*. Although prior work did support genetic differences between orange and blue-morphs, there was little direct evidence for the genetic distinctiveness of yellow-morphs from the other two morphs.

Orange and blue genomes

The genome assembled from the blue-morph was 2.116 Gb across 482,585 scaffolds. The genome assembled from the orange-morph was 2.096 Gb across 176 contigs, with 9 contigs ≥ 100 Mb, 35 ≥ 10 Mb, and 71 ≥ 1 Mb.

Orange versus blue-morph GWAS

We made a fine-scale examination of the local region associated with the blue and orange-morph phenotypes (Fig. S4). In the mapping to the orange-morph genome, there was a second peak of association $\sim 100,000$ bp away from the main peak of association near the *SPR* gene (Fig. S4A). This second peak was not observed in the mapping to the blue-morph genome (Fig. S4B). The region associated with the second peak was quite narrow and was only 464 bp. We used RepeatMasker (65) to search for vertebrate repetitive elements within this second peak region and found that it was made up of an L3 element and a BovB-1_DR element. We hypothesized that the second association peak could be due to mismapping of repetitive element sequences. We had recalled that there was a region of sequence found on B-haplotypes, but not O-haplotypes, that had both an L3 element and a BovB-1_DR element (Fig. 1C; the L3 element is part of the CR1 family and the BovB-1_DR element is part of the RTE-BovB family). The orange-morph reference genome would not have these B-haplotype only repetitive elements sequences near *SPR*, so we hypothesized that they would erroneously map to repetitive sequences elsewhere in the genome.

We tested our hypothesis in two ways. First, we extracted the second peak region DNA sequence from the orange-morph genome and tried aligning it to the *SPR* B-haplotype from the blue-morph genome. The DNA sequence of the second peak region was found to align to a region of the B-haplotype that is not found in the O-haplotype, which was where there was the sequence of the L3 element and the BovB-1_DR element as we had hypothesized. Second, we extracted the B-haplotype DNA sequence from the area that aligned with the orange-morph genome's second peak region. We used this B-haplotype sequence as a query in a blastn (63) search of the entire orange-morph genome. The best blast match was the second peak region in the orange-morph genome, which had an evalue of 3.13×10^{-169} (the next best matching region had an evalue 1.03×10^{-49}). Therefore, the second peak region in orange-morph genome was clearly the best place for mapping the repetitive elements restricted to B-haplotype. Therefore, we

conclude that the second-peak is an erroneous association due to mismapping of repetitive element sequences restricted to the B-haplotype.

We observed three different structural genetic variants in *SPR* exon1 in the *SPR*-region haplotypes obtained from PacBio sequencing. These structural variants (i.e. indels) could potentially affect the phenotype, but would not have been included in our initial genetic scan on polymorphic nucleotides. We used individuals with diploid PacBio haplotype data to test if the structural variants were associated with orange or blue phenotypes. With this PacBio dataset, the upstream region to *SPR* had a stronger association ($P = 9 \times 10^{-14}$) than any of the exon1 structural variants (top association $P = 1 \times 10^{-5}$). This observation supports the hypothesis that regulatory changes in *SPR* underlie the phenotypes, because the strongest genetic associations are upstream of *SPR*.

Yellow-morph GWAS

We further explored the genetic basis of the yellow-morphs by conducting a GWAS with 10 individuals with all-yellow throats, which were separately compared to the 44 orange and 45 blue-morph samples. Neither GWAS had any polymorphic nucleotide positions with an association sufficiently strong that it passed the threshold of a Dunn-Šidák correction for multiple tests. However, we examined the top associated sites to see if they corresponded to any genes that might be relevant to the phenotypes. The two sites with the strongest association in the comparison of yellow versus orange-morphs were in the *SPR* region on contig 9 (Fig. S3C) and were the same two sites that showed the strongest association in the comparison of orange versus blue-morphs. This association is consistent with 6/10 of these yellow-morphs having at least one B-haplotype and all the orange-morphs having only O-haplotypes. The comparison of yellow versus blue-morphs had association peaks on contigs 5 and 7 (Fig. S3D). The top blast hit for the region on contig 5 was the Wnt ligand secretion mediator (*WLS*) gene, which is a conserved gene whose protein is responsible for transporting Wnt proteins (*104*). The top blast hit for the region on contig 7 was a repetitive element, specifically a long interspersed nuclear element (LINE) that has been found in birds (i.e. an avian retrotransposon-like element (*105*)). We did a RepeatMasker search of the associated region on contig 7 and found that 71% of it was repetitive elements. Overall, neither the contig 7 nor the contig 9 region contained candidate genes that seemed likely to influence coloration or behavior.

Yellow-morphs in the European common wall lizard (*Podarcis muralis*) were associated with genetic differences in the beta-carotene oxygenase 2 (*BCO2*) gene (*8*). We therefore examined the genetic associations around this candidate gene for yellow coloration in *U. stansburiana*. We did not find any suggestive associations with *BCO2*, because the associations in this gene were far below the multiple testing threshold of significance and were not outliers when compared to other positions on the same contig (Fig. S4C, D). Thus, *U. stansburiana* and *P. muralis* do not share a similar genetic basis for their yellow throat color phenotypes, despite them convergently utilizing the *SPR* gene for generating orange coloration.

Morph genotypes and phenotypes

We observed an almost perfect association between orange and blue-morph phenotypes and their genotypes at the *SPR* locus. We observed that all orange-morphs (44/44) had two O-haplotypes and that almost all blue-morphs had at least one B-haplotype (42/45). However, our sample contained three individuals classified as blue-morphs that were homozygous for the O-haplotype. We tried to vet these samples to determine if there was an error associated with the field data or genetic sampling of these individuals that could explain the discordance between the genotype and phenotype of these individuals. All three of these lizards were caught during years

when Barry Sinervo (the principal investigator of our focal study population) was unable to work at the field site. There are no photos associated with the discordant individuals that can be used to verify their phenotype (note that there is a perfect association between genotype and phenotype for all the lizards for which we do have photos). We are thus unable to reconcile the discordance between genotype and phenotype for these three individuals except to say that there are factors associated with these samples that may have made errors more likely.

We tested whether there were size differences among the morphs and *SPR* genotypes to evaluate whether there were phenotypic differences that distinguished the three anomalous OO blue-morphs and the four OO yellow-morphs, since these seven individuals were the only exceptions to our two-allele model. The analysis of body size ($\log_{10}$ snout-vent length (SVL), Fig. S7A) found significant differences among morphs ($F_{3,101} = 3.62$, $P = 0.016$) and genotypes ($F_{2,101} = 4.72$, $P = 0.011$), but no significant morph $\times$ genotype interaction ($F_{3,101} = 0.42$, $P = 0.741$). The analysis of standardized weight ($\log_{10}$ weight divided by $\log_{10}$ SVL, Fig. S7B) found significant differences among morphs ($F_{3,101} = 4.73$, $P = 0.004$), but no significant effect of the genotypes ($F_{2,101} = 2.34$, $P = 0.101$) or the morph $\times$ genotype interaction ($F_{3,101} = 0.60$, $P = 0.618$). The category of all-yellow phenotype yellow-morphs with OO genotypes stood out as being different in both body size and standardized weight (Fig. S7). If these four OO yellow-morphs were excluded from the analysis, there were no longer any significant differences in the analyses of body size or standardized weight (all $P > 0.194$). Thus, aside from these four yellow-morphs, there were generally no differences in body size or standardized weight among any of the morphs or genotypes. Thus, the anomalous OO blue-morphs were comparable in size to both blue and orange-morphs. However, when only the individuals with OO genotypes were analyzed together, the OO yellow-morphs were significantly smaller than orange-morphs in body size (Tukey posthoc test $t = -2.77$, $P = 0.020$) and standardized weight (Tukey posthoc test $t = -3.48$, $P = 0.003$). The small body size and standardized weight of the OO yellow-morphs suggests that they may be younger than the other males and still developing their color. Thus, we hypothesize that the yellow color of OO yellow-morphs may be a transient phenotype and they may express orange coloration later in life. The hypothesis that these individuals have a transient yellow coloration, perhaps because they are just beginning to express orange pigments, is consistent with our analyses of their throat color phenotypes that showed that they were more similar to orange-morphs in their coloration than other yellow-morphs (see below the section: “Evidence for the one-locus, two-allele model with plasticity”).

We tested whether BB and OB *SPR* genotypes differed in producing blue-morphs versus yellow-morphs. We observed 9 BB (21%) and 33 OB (79%) blue-morphs and 6 BB (37.5%) and 10 OB (62.5%) yellow-morphs. The frequencies of the two genotypes did not differ between blue and yellow-morphs (Fisher’s exact test $P = 0.31$). Thus, blue-morphs and yellow-morphs do not predominantly arise from particular genotypes, which could have occurred if there were differences in plasticity between BB and OB genotypes.

SPR pathways

The *SPR* gene is involved in two metabolic pathways relevant to morph phenotypes in terms of coloration and behavior (Fig. S5). First, *SPR* affects the formation and conversion of sepiapterin (Fig. S5A), which is a yellow-colored pteridine pigment (106, 107). *SPR* also affects the levels of 6-pyruvoyl-H₄pterin, which can be converted into drosopterin, which is an orange or red pteridine pigment (36, 107). Second, *SPR* helps synthesize tetrahydrobiopterin (BH₄), which is an essential cofactor for the enzymes tryptophan hydroxylase (TPH), phenylalanine hydroxylase (PAH), tyrosine hydroxylase (TH), and nitric oxide synthase (NOS) (Fig. S5A, B) (34, 35). These enzymes play key roles in the synthesis of neurotransmitters (Fig. S5B), because

TPH begins the modification of tryptophan into serotonin or melatonin, NOS produces nitric oxide, PAH produces tyrosine, and tyrosine is modified by TH into dopa, which is the precursor of dopamine, noradrenaline and adrenaline (34, 35). These neurotransmitters and hormones can have profound effects on behavior including aggression, sexual behavior, and social status (108-111). Thus, pleiotropic effects of *SPR* in both pigmentation and neurotransmission/hormonal pathways may allow behavioral differences to be coupled to a color signal to generate distinct morph phenotypes.

Both the GWAS results and the RNAseq data show that the *SPR* gene is likely to affect the development of orange versus blue throat coloration. The orange throat coloration is likely to be derived from an orange drosopterin pteridine (8, 20), but the blue throat coloration is likely structural coloration in which light is reflected and scattered from guanine crystals (32, 33). The mechanisms that allow *SPR* to mediate a transition between a pigment versus a structural color are currently unknown. However, we hypothesize that the answer may lie in a link between the guanine salvage pathway (Fig. S5C) and the BH4 synthesis pathway (Fig. S5A). The guanine salvage pathway shows that guanosine-5'-triphosphate (GTP) can be synthesized from a guanine precursor and GTP can be turned back into guanine. Orange and yellow pteridines are synthesized from GTP, with the rate limiting step of the pathway being the action of GTP cyclohydrolase I (GTP-CH), which turns GTP into H₂neopterin triphosphate (36). We posit that changes in expression of *SPR* could mediate a switch between the production of orange pigments and blue structural color as follows. Increased *SPR* expression could mean increased conversion of 6-pyruvoyl-H₄opterin into BH4, which would thereby decrease the synthesis of orange drosopterins, which are formed from 6-pyruvoyl-H₄opterin (36, 112). In addition, the increased formation of BH4 would in turn repress GTP cyclohydrolase I feedback regulatory protein (GFRP), which would lead to the reduced conversion of GTP by GTP-CH (36, 113). This would increase levels of GTP, which could be converted to guanine via the salvage pathway. This guanine could be used for guanine crystals that form the basis of structural blue coloration in lizards.

Our hypothesis about how changes in expression of *SPR* could mediate a switch between the BH4 and the guanine salvage pathways is speculative at this point and detailed studies of the molecular mechanisms are needed to resolve how orange and blue colors are generated. However, our hypothesis is meant to illustrate that a switch between an orange pteridine pigment and a blue structural color is possible through established biochemical pathways and changes in the expression of the *SPR* gene.

Expression of the *SPR* gene

We quantified the expression of the *SPR* gene in the main color patch of throat skin (Fig. S1), but also in white or very lightly colored skin outside the main color patch, which we refer to as "noncolored skin". There were no significant differences among morphs in the expression of the *SPR* gene in the noncolored skin (Fig. S9B, $F_{2,21} = 2.2$, $P = 0.132$) unless a single outlying blue-morph datapoint was removed, in which case blue-morphs had significantly higher expression than orange-morphs ($P = 0.0128$). We next tested if there were differences among the morphs and genotypes in the levels of expression of the colored skin relative to the non-colored skin. Significant differences were observed among morphs in the expression of the colored skin relative to the noncolored skin (Fig. S9C, $R^2 = 0.23$, $F_{2,21} = 4.5$, $P = 0.0234$). Blue-morphs had a significantly higher ratio of expression than orange-morphs (Fig. S9C, $P = 0.0202$), but no significant differences were observed in the other two morph comparisons (yellow vs. blue: $P = 0.144$; yellow vs. orange $P = 0.601$). The levels of expression of the colored skin relative to the noncolored skin were even better explained by a model that included

SPR region genotypes (Fig. S9D, $R^2 = 0.32$, $F_{2,21} = 6.5$, $P = 0.0065$). OB genotypes had a significantly higher ratio of expression than OO genotypes (Fig. S9D, $P = 0.0081$), no significant differences were observed between OB and BB genotypes ($P = 0.996$), and the comparison of OO to BB genotypes had a P -value of 0.0714. Overall, our analysis of the expression of *SPR* in the skin suggests that blue-morphs and individuals with a blue haplotype have higher expression of *SPR* than orange-morphs and that this difference in expression was primarily within the colored region of the skin.

We next tested whether there was *SPR* haplotype-specific expression within the skin in individuals that were heterozygous for the two haplotypes (i.e. 8 blue-morphs and 2 yellow-morphs). We observed that B-haplotypes were expressed at significantly higher levels than O-haplotypes in both colored (Fig. S9E, $P = 0.00007$) and noncolored (Fig. S9F, $P = 0.0003$) skin. In both types of skin and all ten individuals, the B-haplotype had higher expression than the O-haplotype within an individual lizard. The B-haplotype was expressed on average 7.7 times higher (range = 1.9 – 20.9) than the O-haplotype in colored skin and 5.9 times higher in noncolored skin (range = 1.5 – 18.4). Thus, there was a clear pattern of haplotype-specific expression in the skin. We next tested if the expression of each haplotype increased in colored skin compared to noncolored skin. We observed significantly higher expression in colored skin for both the B-haplotype (Fig. S9G, $P = 0.0008$) and the O-haplotype (Fig. S9H, $P = 0.0375$). The B-haplotype was expressed on average 1.6 times higher (range 0.9 – 2.25) higher in colored skin and the O-haplotype was expressed on average 1.2 times higher (range 0.8 – 1.5). Thus, both haplotypes showed increased expression in colored skin compared to noncolored skin. We plotted the haplotype-specific expression data for colored skin on to a phylogeny of *SPR*-region haplotypes to evaluate if there were any systematic differences in expression among clades within the phylogeny (Fig. S9I). There were clear differences in expression between the O and B-haplotypes. However, there were no systematic differences in expression among subclades within the O and B-haplotypes that would suggest that particular subclades have differences in expression, with the caveat that we had low sample sizes for evaluating such expression differences.

We also quantified the expression of the *SPR* gene in the brain. There were no significant differences among morphs in the expression of the *SPR* gene in the brain (Fig. S9J, $F_{2,21} = 0.538$, $P = 0.592$). However, these results reflect expression across the whole brain, so we cannot rule out localized expression differences among morphs. We next tested whether there was *SPR* haplotype-specific expression within the brain in individuals that were heterozygous for the two haplotypes. We observed that B-haplotypes were expressed at significantly higher levels than O-haplotypes in the brain (Fig. S9K, $P = 0.0334$). In nine of the ten individuals, the B-haplotype had higher expression than the O-haplotype within an individual lizard. The B-haplotype was expressed on average 1.9 times higher (range = 0.5 – 4.1) than the O-haplotype in the brain. Thus, we observed haplotype specific differences in expression in both the brain and the skin.

Evidence for the one-locus, two-allele model with plasticity

Several observations in addition to the ones described in the main text suggest that a one-locus, two-allele model with plasticity is a sufficient and parsimonious explanation for the three morph phenotypes. First, genetic variation at the *SPR* region is important for explaining the yellow-morph phenotype, because although yellow-morphs do not have a unique third haplotype, most had a B-haplotype, which explains their ability to transform to a blue-morph phenotype under certain conditions. Second, *SPR* genotypes were correlated with subtle coloration differences among yellow-morphs. Yellow-morphs with one or two B-haplotypes were

significantly closer to blue-morphs in color space than yellow-morphs with two O-haplotypes (Fig. S1F, $F_{1,7} = 6.83$, $P = 0.035$), which tended to be closer to orange-morphs in color space ($F_{1,7} = 4.51$, $P = 0.071$). This shows that individuals with two O-haplotypes classified as yellow-morphs have phenotypes with subtle similarities in coloration to orange-morphs. We hypothesize that as pigments build up on the throats of orange-morphs, they go through a transient phase where they resemble yellow-morphs, which leads to some errors of classification for individuals with the pure-yellow throat color phenotype. We showed above (in the section: Morph genotypes and phenotypes) that the OO yellow-morphs had relatively small body sizes, so they may have still been developing their throat color phenotypes. Regardless of the mechanism, *SPR* genotypes are predictive of phenotypic variation in the coloration of yellow-morphs. Third, *SPR* genotypes helped explain gene expression differences among yellow-morphs (Fig. S9), which provides further evidence that *SPR* haplotypes functionally affect the phenotype of yellow-morphs and that individuals with two O-haplotypes classified as yellow-morphs have phenotypes similar to orange-morphs. Fourth, the hypothesis that yellow-morphs suppress the expression of blue throat coloration is consistent with them also lacking male-specific coloration on their dorsum, which has led to them being described as female mimics. Overall, these data are consistent with the hypothesis that the yellow-morphs predominantly result from plasticity in individuals with a B-haplotype, which can either express yellow coloration and adopt a sneaker strategy or can express blue coloration and defend a territory.

Population genetic modeling of the maintenance and invasion of the alleles

Our deterministic two-allele models with a fixed proportion of blue and yellow-morphs arising from individuals with a B-haplotype showed that a rare blue allele (Fig. S12 A, D, G, J) and a rare orange allele (Fig. S12 C, F, I, L) could spread in a population dominated by the other allele (i.e. invade) and attain the same equilibrium frequencies. However, there was a transient effect of the starting allele frequency. Models that started with an orange allele frequency of 50% or 99% converged on the same equilibrium allele frequencies by generation 100 (Fig. S12 A, B, G, H), but models started with 1% orange alleles had very different allele frequencies at generation 100, with high levels of the blue allele (Fig. S12 C, I). However, all starting conditions converged to the same equilibrium allele frequencies by generation 10,000 (Fig. S12 D-F, J-L). The allele frequencies converged more slowly to an equilibrium for the 1% orange allele starting condition, because the formation of individuals homozygous for the O-haplotype that have the orange-morph phenotype will be quite rare when the orange allele frequency is low. These models also showed that the polymorphism is stably maintained across a much broader range of yellow-morph proportions with the paternity data fitness payoffs (Fig. S12 A-F) than with the territory data fitness payoffs (Fig. S12 G-L). One difference between the two matrices is that with the paternity data fitness matrix, a rare orange-morph in an environment where yellow-morphs are common has slightly higher fitness (payoff = 1.2) than a yellow-morph (payoff = 1) in that environment, whereas with the territory overlap fitness matrix, the rare orange-morph has much lower fitness (payoff = 0.4). Thus, with the paternity data fitness matrix, orange-morphs have an advantage when rare against yellow-morphs and a disadvantage when common, which favors the persistence of these strategies. For example, orange-morphs can persist at a yellow-morph proportion of 1 (i.e. all individuals with a blue haplotype stay as yellow-morphs) with the paternity data fitness matrix (Fig. S12E).

We examined a model in which the proportion of blue and yellow-morphs in each generation was optimized based on the frequency of the orange-morphs. This optimal-proportion model maintained the polymorphism with both payoff matrices (Fig. S13). Simulations started with an orange allele frequency of 99%, 50%, or 1% all converged to the

same equilibrium frequencies, although this took longer when the orange allele started at 1% frequency (Fig. S13 A-F). Thus, both the blue and orange alleles can increase in frequency when rare and the proportion of blue and yellow-morphs converges to a fixed proportion in these deterministic models due to stabilization of the allele frequencies. When genetic drift is added to the model, the polymorphism can also be maintained with the optimal-proportion model (Fig. S13 G, H). Notably, the version of this model that used territory data fitness payoffs exhibited cycles in the frequencies of the morphs when there was genetic drift (Fig. S13 H). This model exhibited the frequency changes expected with rock-paper-scissors dynamics among the color-morphs, which were a high frequency of orange, followed by a high frequency of yellow, followed by a high frequency of blue, and followed by repeated changes in frequencies following this pattern. When the paternity data fitness payoffs were used (Fig. S13 G), there could be periods with ordered changes in frequency from high orange to high yellow to high blue (e.g. generations 1-30), but there were also periods with unordered changes in the morph frequencies in which cycling was not observed (e.g. generations 30-50).

We examined a model in which the proportion of blue and yellow-morphs in each generation was semi-optimized (i.e. the optimal proportion was calculated and then offset by a constant value (c_{morph})). This model provided further insight into the frequency cycles of the morphs (Fig. S14). The version of the semi-optimized model that used territory data fitness payoffs exhibited cycles in the frequencies of the morphs when there was an increased proportion of blue-morphs (i.e. negative values of c_{morph} , Fig. S14B, D), but no cycles were observed when there was an increased proportion of yellow-morphs (i.e. positive values of c_{morph} , Fig. S14F, H). No frequency cycles were observed when the paternity data fitness payoff matrix was used in the model (Fig. S14A, C, E, G). This pattern is similar to what was observed with the original three-allele model for the morph dynamics, where frequency cycles were observed with the territory data fitness matrix (Fig. S14J), but not the paternity data fitness matrix (Fig. S14I).

Game theoretical modeling of the maintenance and invasion of the alleles

The rock-paper-scissors game is a classic model in game theory, so we also conducted a game theoretical treatment for the side-blotched lizards with our new model with one type with a fixed strategy and one plastic type that can play two strategies. We consider a population of individuals with three possible strategies: Orange (O), Blue (B), and Yellow (Y). There are two types of individuals in the population: (1) individuals that follow strategy O (O individuals) and (2) individuals that follow strategy B with probability $1 - p$ and strategy Y with probability p (Y/B individuals). The pay-off for an individual playing strategy i against a player with strategy j is a_{ij} . We will assume that $a_{OO} = a_{BB} = a_{YY} = 1$, but the results can easily be generalized to an arbitrary pay-off matrix. In a population where the current frequency of O individuals is f , the frequency (q) of individuals with the three strategies are $q_O = f$, $q_B = (1 - f)(1 - p)$, and $q_Y = (1 - f)p$. The expected pay-off for individuals of type O is

$$w_O = q_O a_{OO} + q_B a_{OB} + q_Y a_{OY}$$

and the expected pay-off in the population averaged over all individuals is

$$\bar{w} = \sum_{i \in \{O, B, Y\}} q_i (q_O a_{iO} + q_B a_{iB} + q_Y a_{iY})$$

We set the replicator equation equal to zero and solve for f to find possible equilibrium states:

$$\dot{f} = f(w_O - \bar{w}) = 0$$

which after a bit of algebra reveals find three solutions $\hat{f} = 0$, $\hat{f} = 1$, and

$$\hat{f} = \frac{a_{OB}p - a_{OB} - a_{OY}p - a_{BY}p^2 + a_{BY}p - a_{YB}p^2 + a_{YB}p + 2p^2 - 2p + 1}{a_{OB}p - a_{OB} - a_{OY}p + a_{BO}p - a_{BO} - a_{BY}p^2 + a_{BY}p - a_{YO}p - a_{YB}p^2 + a_{YB}p + 2p^2 - 2p + 2} \quad (\text{Equation 1}).$$

As there is at most one solution for $f \in (0,1)$, we can fully describe the behavior of the system using this solution and an analysis of the behavior at the boundaries. Let the expected payoff of Y/B individual be $w_{Y/B}$, then a population of only O individuals can be invaded by Y/B individuals if (CONDITION 1):

$$\lim_{f \uparrow 1} (w_{Y/B} - w_O) = (1 - p)a_{BO} + pa_{YO} - 1 > 0$$

Similarly, a population of only Y/B individuals can be invaded by O individuals if (CONDITION 2):

$$\lim_{f \downarrow 0} (w_O - w_{Y/B}) = (1 - p)a_{OB} + pa_{OY} - (1 - p)(1 - p + pa_{BY}) - p((1 - p)a_{YB} + p) > 0$$

These results imply the following behavior of the system:

- (1) If CONDITION 1 and CONDITION 2 are both true, then there is a single Evolutionary Stable State given by $\hat{f}$ in Equation 1.
- (2) If CONDITION 1 is true and CONDITION 2 is false, then Y/B is an Evolutionary Stable Strategy (ESS) and there is a single Evolutionary Stable State given by $f = 0$.
- (3) If CONDITION 1 is false and CONDITION 2 is true, then O is an ESS and there is a single Evolutionary Stable State given by $f = 1$.
- (4) If CONDITION 1 is false and CONDITION 2 is false then both O and Y/B are ESSs, there are two Evolutionary Stable States at $f = 0$ and $f = 1$, and the system will end in $f = 0$ or $f = 1$ if the initial value of f is larger or smaller, respectively, than $\hat{f}$ in Equation 1. An exception is the trivial case where the two conditions hold if the inequalities are replaced with equalities, in which case the system does not evolve.

We applied these four results to the pay-off matrices (see Methods section) of male fitness estimated from either overlap with female territories (6, 19) or from genetic paternity estimates of siring success (15, 19). The Evolutionary Stable State as a function of the proportion of sneaker morphs (p) is shown in Figure S15. The results of the ESS analysis are very similar to the equilibrium states obtained from our population genetic simulations (e.g. Fig. S12E vs. Fig. S15A; Fig S12K vs. S15B). In both sets of analyses, the polymorphism is stably maintained across a much broader range of sneaker morph proportions with the paternity data fitness payoffs than with the territory data fitness payoffs.

Note that the ESS analysis does not take matings between males and females and segregation of alleles in diploid organisms into account. Also note that by relaxing the assumption of $a_{OO} = a_{BB} = a_{YY} = 1$ it would similarly be possible to model an intrinsic fitness difference between morphs/strategies, for example due to the energetic cost or increased exposure to predation of the O strategy.

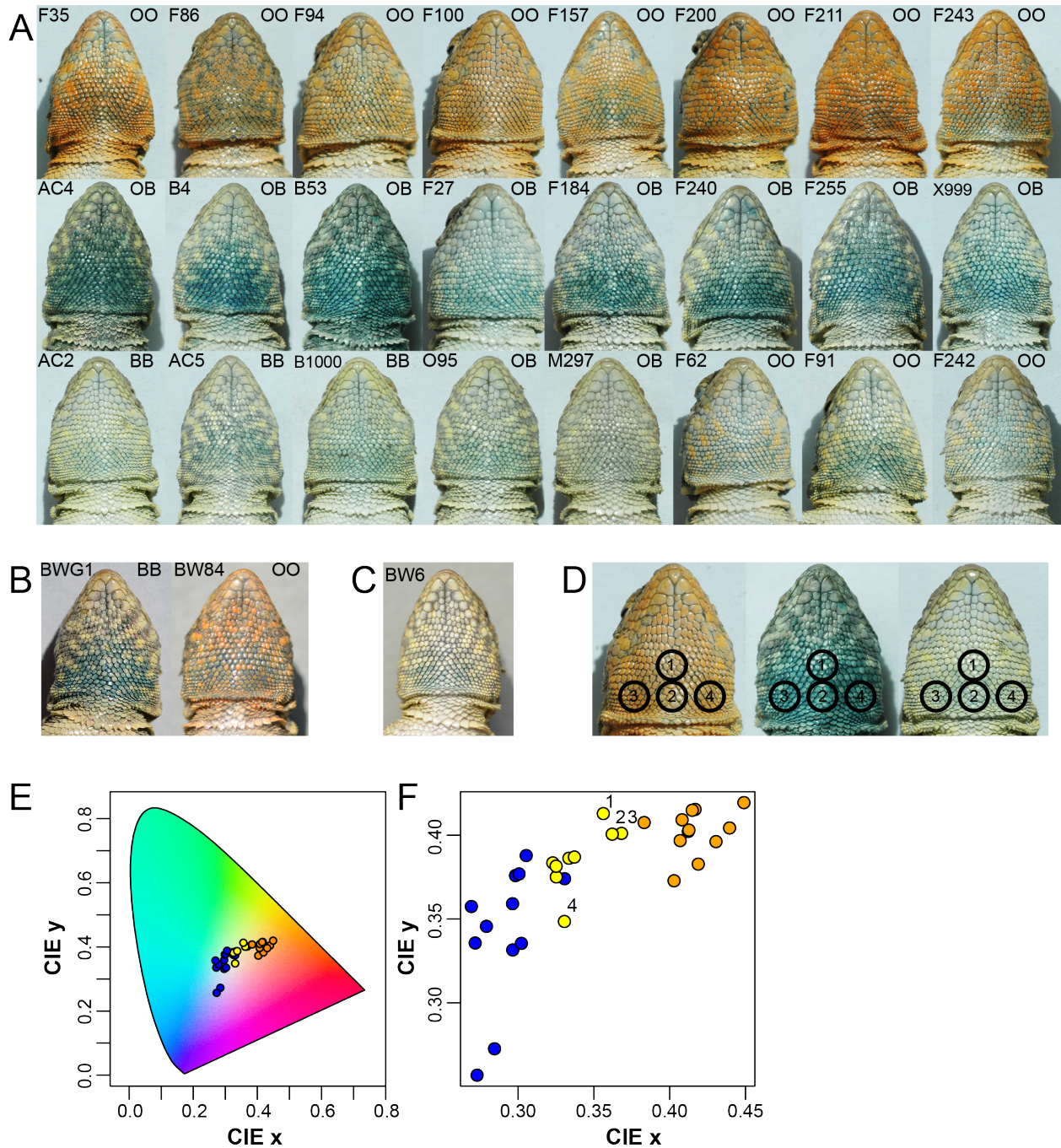
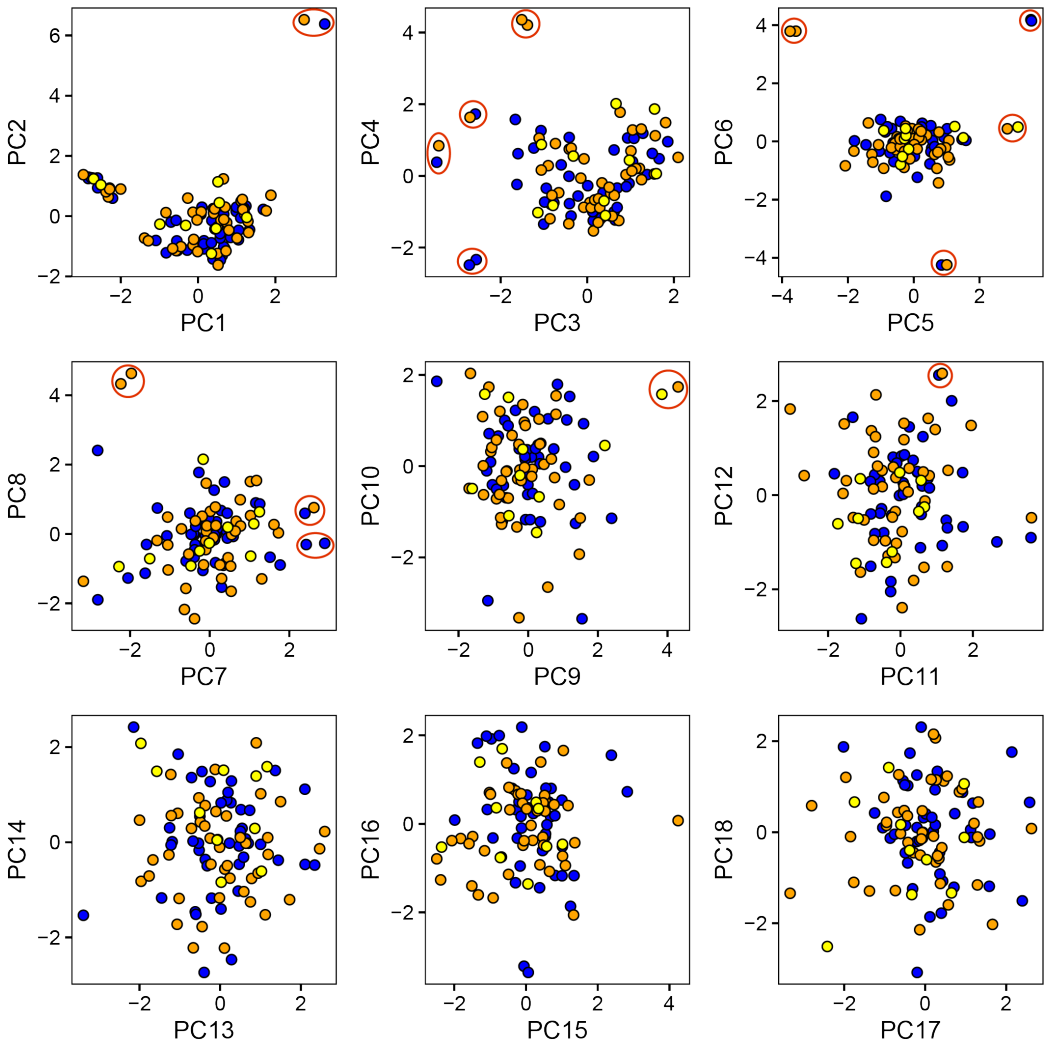


Fig. S1. Throat color phenotypes. **A.** Photos of the 24 lizards used in the study of gene expression. Orange-morphs are in the top row, blue-morphs in the middle row, and yellow-morphs in the bottom row. **B.** Photos of the blue-morph (left) and orange-morph (right) that were used for constructing reference genomes. In panels A-B, the field identification code of the lizard is in the top left corner and the genotype of the lizard at the *SPR* region is in the top right corner, with OO = homozygous for the orange haplotype, BB = homozygous for the blue haplotype, OB = heterozygous. **C.** Photo of a yellow-morph male with the blue-yellow phenotype. Note the light blue patch at the base of the throat and the yellow color at the sides and upper part of the throat. **D.** The four regions of the throat where color measurements were taken with a spectrometer. **E.** Measurements of the throat colors of 12 blue, 9 yellow, and 12 orange-morphs plotted in CIE xyz color space. Color measurements from the lizards in panel A

1777 are included in this figure. **F.** Zoomed in view of the throat color data. The four yellow-morphs
1778 with two O-haplotypes at the *SPR* locus are shown with numbers. Points in E and F are colored
1779 by their morph classification.
1780

1781



1782

1783

1784

1785

1786

1787

1788

1789

Fig. S2. Genetic similarity of 44 orange, 45 blue, and 10 yellow-morph lizards used in the genome-wide association study. Individual lizards are colored by their morph phenotype and are plotted on principal component (PC) axes of the genetic variation observed amongst the lizards. The outlying pairs of points that are circled in red are six pairs of individuals that have pairwise relatedness values above 0.1, many of which are outliers in multiple plots.

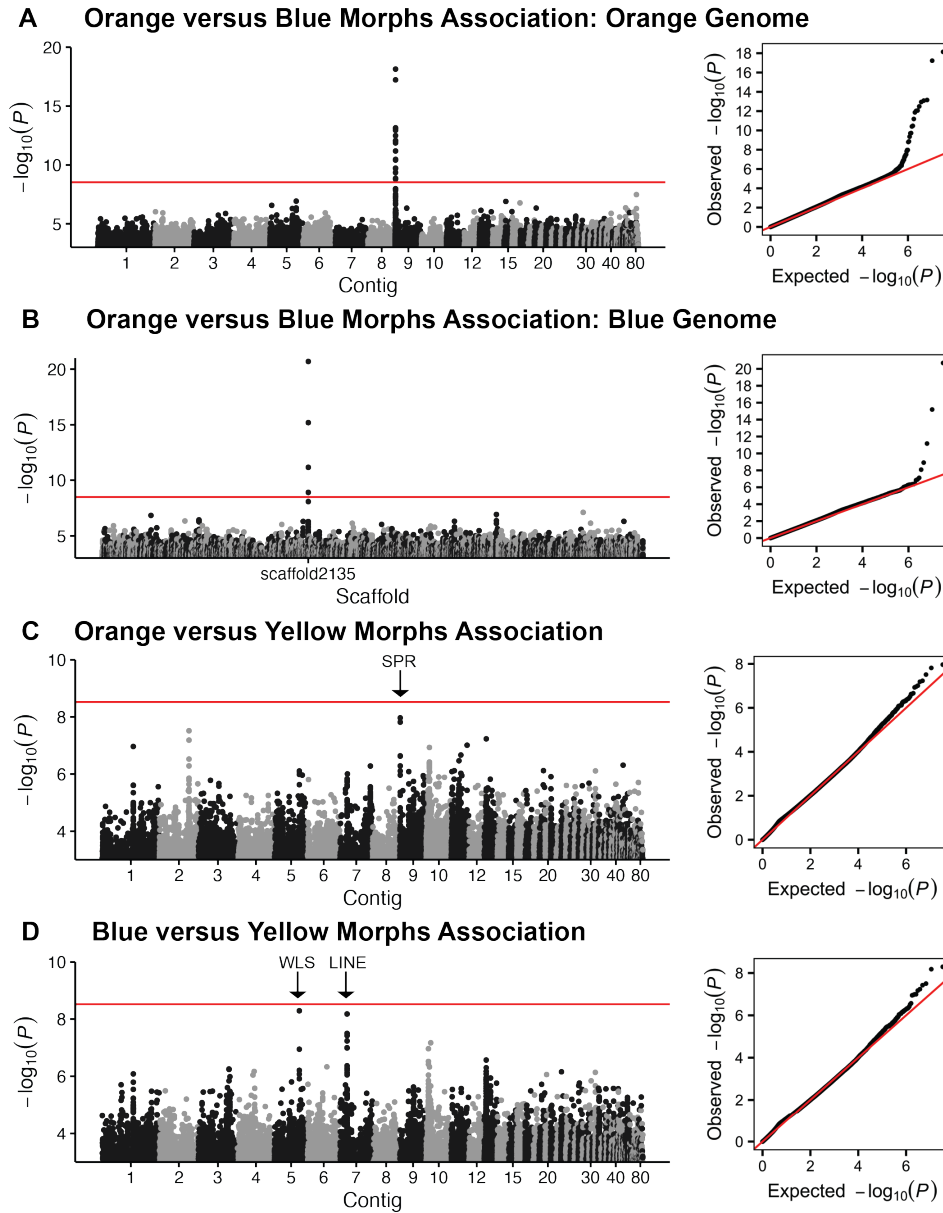


Fig. S3. Genetic associations with the morph phenotypes. **A.** Genetic associations for the orange (N= 44) versus blue (N= 45) morph phenotypes with the data mapped to the orange-morph genome. **B.** Genetic associations for the orange (N= 44) versus blue (N= 45) morph phenotypes with the data mapped to the blue-morph genome. The scaffold with the *SPR* gene is labeled. **C.** Genetic associations for the orange (N= 44) versus yellow (N = 10) morph phenotypes with the data mapped to the orange-morph genome. The location of the *SPR* gene is shown with an arrow. **D.** Genetic associations for the blue (N= 45) versus yellow (N = 10) morph phenotypes with the data mapped to the orange-morph genome. The location of the Wnt ligand secretion mediator (*WLS*) gene and the long interspersed nuclear element (LINE) are shown with arrows. In the Manhattan plots on the left, points alternate between black and gray to highlight different scaffolds/contigs when plotting the associations of the polymorphic sites. The red line in the Manhattan plots shows the threshold of significance based on a Dunn-Šidák correction for multiple tests. The qqplots on the right show the distribution of observed *P*-values and whether they deviate from the expected distribution based on random chance. The red line in the qqplots indicates the position of a one-to-one match between the distributions.

1806

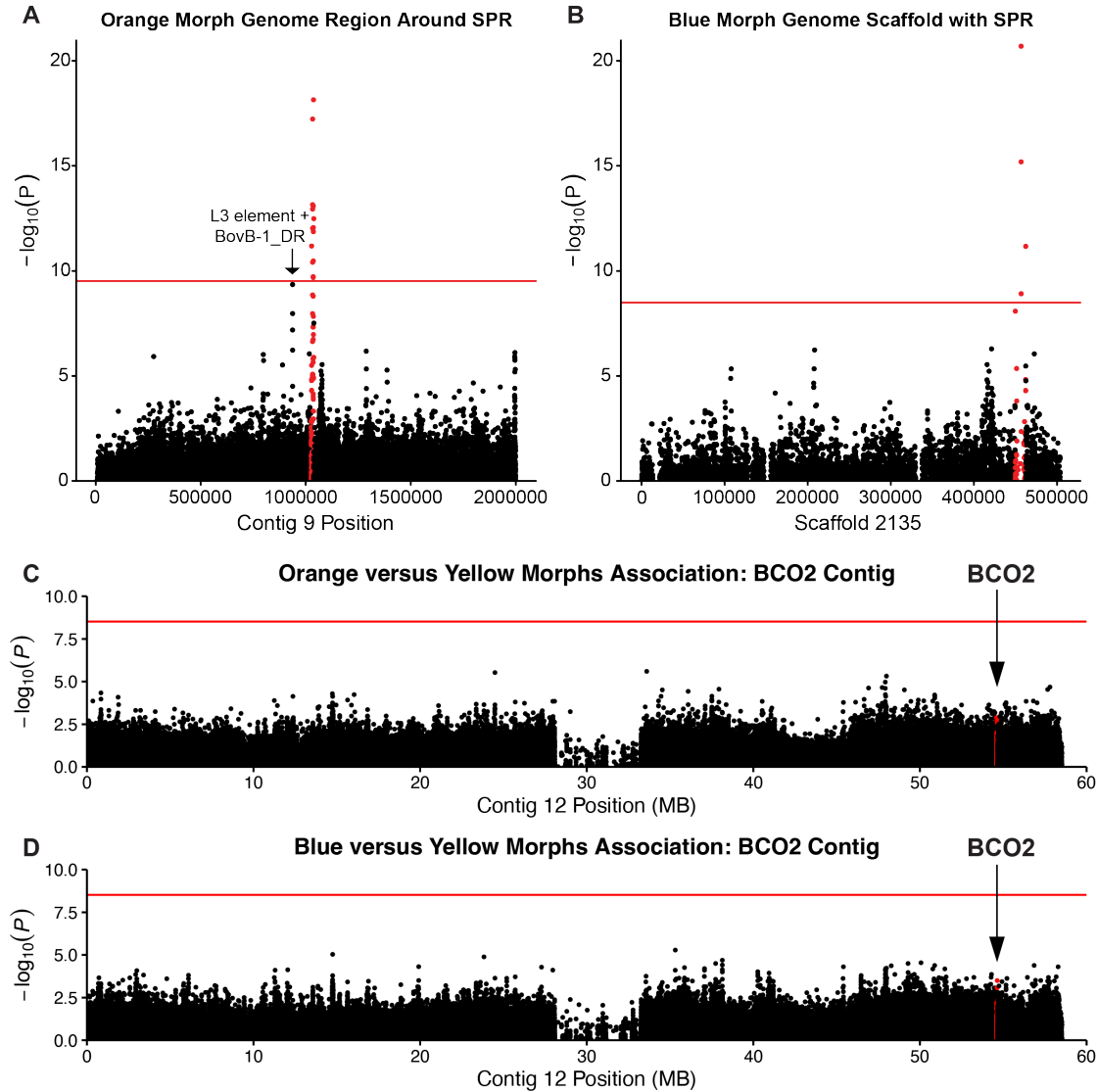


Fig. S4. Local views of the genetic associations with morph phenotypes. Panels A and B show genetic associations with the orange versus blue-morph phenotypes around the *SPR* region. Panels C and D show associations with the yellow-morph phenotype on Contig 12 where the beta-carotene oxygenase 2 (*BCO2*) gene is located, which is associated with yellow-morphs in (*Podarcis muralis*) (8). The red line in all the Manhattan plots shows the threshold of significance based on a Dunn-Šidák multiple-testing correction for all the sites that were tested in the genome. **A.** Association plot of the data mapped to the orange-morph genome, with the plot restricted to the first 2 million base pairs of contig 9, which brackets the position of the *SPR* gene. The location of the L3 element and BovB-1_DR element discussed in the Supplementary Text is shown with an arrow. **B.** Association plot of the data mapped to the blue-morph genome, with a view of all the associations on scaffold 2135, which contains the *SPR* gene. In panels A and B, red points highlight the *SPR* coding sequence and the upstream divergent region amplified by long-range PCR. **C.** Genetic associations with the orange versus yellow-morph phenotypes along Contig 12, with the position of the *BCO2* gene highlighted in red. **D.** Genetic associations with the blue versus yellow-morph phenotypes along Contig 12, with the position of the *BCO2* gene highlighted in red.

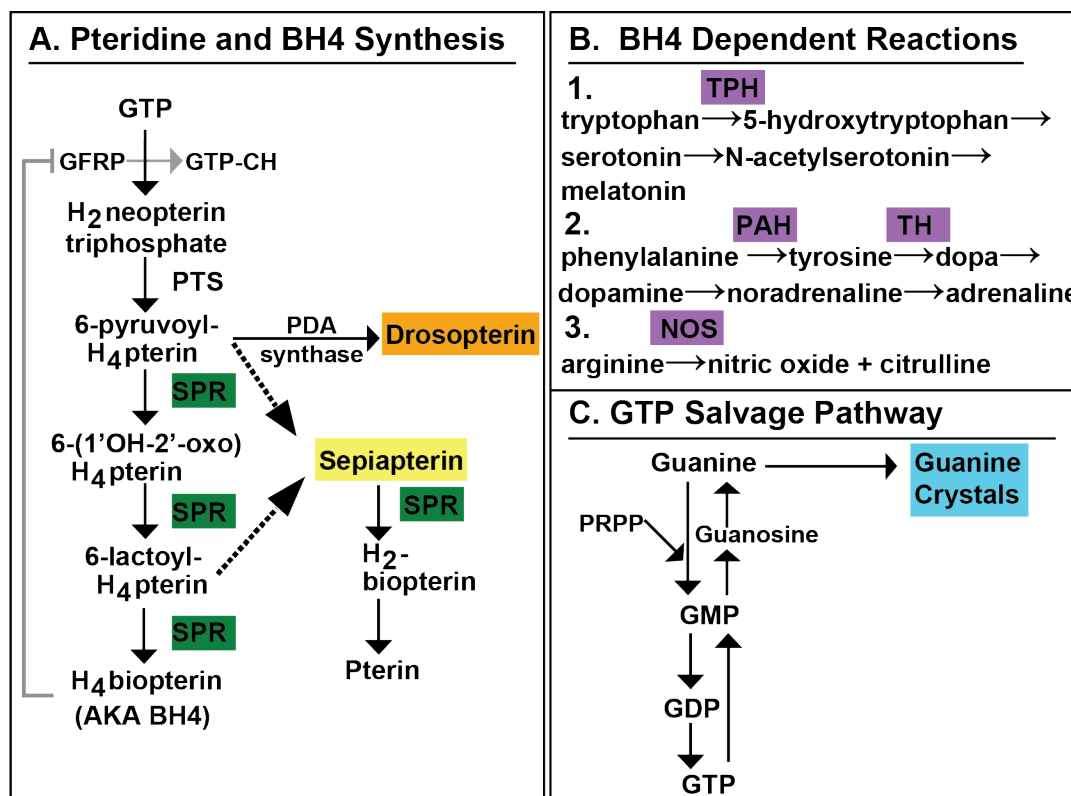


Fig. S5. Metabolic pathways related to the *SPR* gene and morph phenotypes. **A.** The role of sepiapterin reductase (*SPR*) in the synthesis of pteridines and tetrahydrobiopterin (BH4). The pathway begins with the conversion of guanosine-5'-triphosphate (GTP) into H₂neopterin triphosphate by GTP cyclohydrolase I (GTP-CH), which is the rate limiting enzyme for the pathway. H₂neopterin triphosphate is converted to 6-pyruvoyl-H₄pterin by 6-pyruvoyltetrahydropterin synthase (PTS). *SPR* (highlighted in green boxes) acts upon 6-pyruvoyl-H₄pterin and downstream products to catalyze the final stages of BH4 synthesis. *SPR* also acts upon sepiapterin, which is a yellow pigment. The orange pigment drosopterin is known to form from 6-pyruvoyl-H₄pterin in *Drosophila* by the action of pyrimidodiazepine (PDA) synthase, but the synthesis pathway for this pigment has yet to be characterized vertebrates (36, 112). End-product feedback regulation of the BH4 pathway is shown with gray lines. BH4 inhibits GTP cyclohydrolase I feedback regulatory protein (GFRP), which regulates the activity of GTP-CH. Hypothetical enzymatic reactions are shown with dashed arrows. Pathway diagrams follow from (36, 106, 112, 113). **B.** Three pathways for the synthesis of neurotransmitters and hormones that depend upon BH4. BH4 as an essential cofactor for the enzymes (highlighted in purple) tryptophan hydroxylase (TPH), phenylalanine hydroxylase (PAH), tyrosine hydroxylase (TH), and nitric oxide synthase (NOS) (34, 35). **C.** The GTP salvage pathway shows that guanine and GTP can be converted into each other (114-116). Guanine crystals are the basis of blue coloration in many vertebrates (32, 33). Abbreviations: phosphoribosyl diphosphate (PRPP), guanosine monophosphate (GMP), and guanosine diphosphate (GDP).

1849

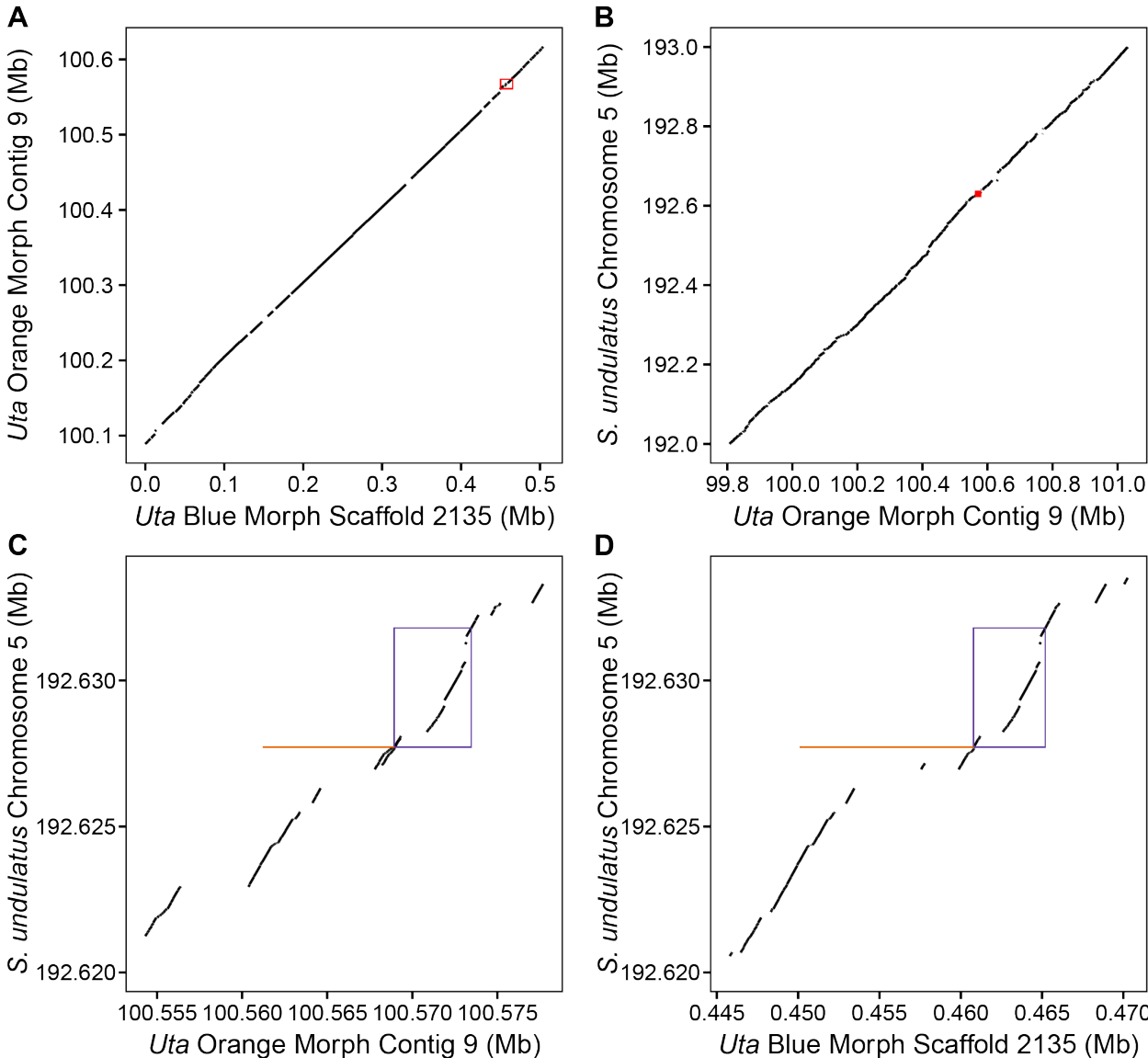


Fig. S6. Alignments between regions with the *SPR* gene in different genomes. A. Alignment between the *Uta stansburiana* blue-morph genome scaffold with *SPR* and the orange-morph genome contig with *SPR*. The red box shows the *SPR* region, which includes both the *SPR* coding region and the divergent region upstream of the gene. **B.** Alignment between a 1 megabase (Mb) region in the *Sceloporus undulatus* chromosome 5 with *SPR* with the orange-morph genome contig with *SPR*. The red box shows the *SPR* coding region. **C and D.** Alignment between the *Sceloporus undulatus* chromosome 5 and the *U. stansburiana* orange-morph contig 9 (**C**) and the blue-morph scaffold 2135 (**D**) in the local region encompassing *SPR*. The purple rectangle shows the *SPR* coding region and the orange line shows the extent of the upstream region to the *SPR* gene studied in *U. stansburiana*. Black lines in all of the plots indicate regions where the sequences between the two genomes could be aligned and whether those regions were in the same orientation relative to one another.

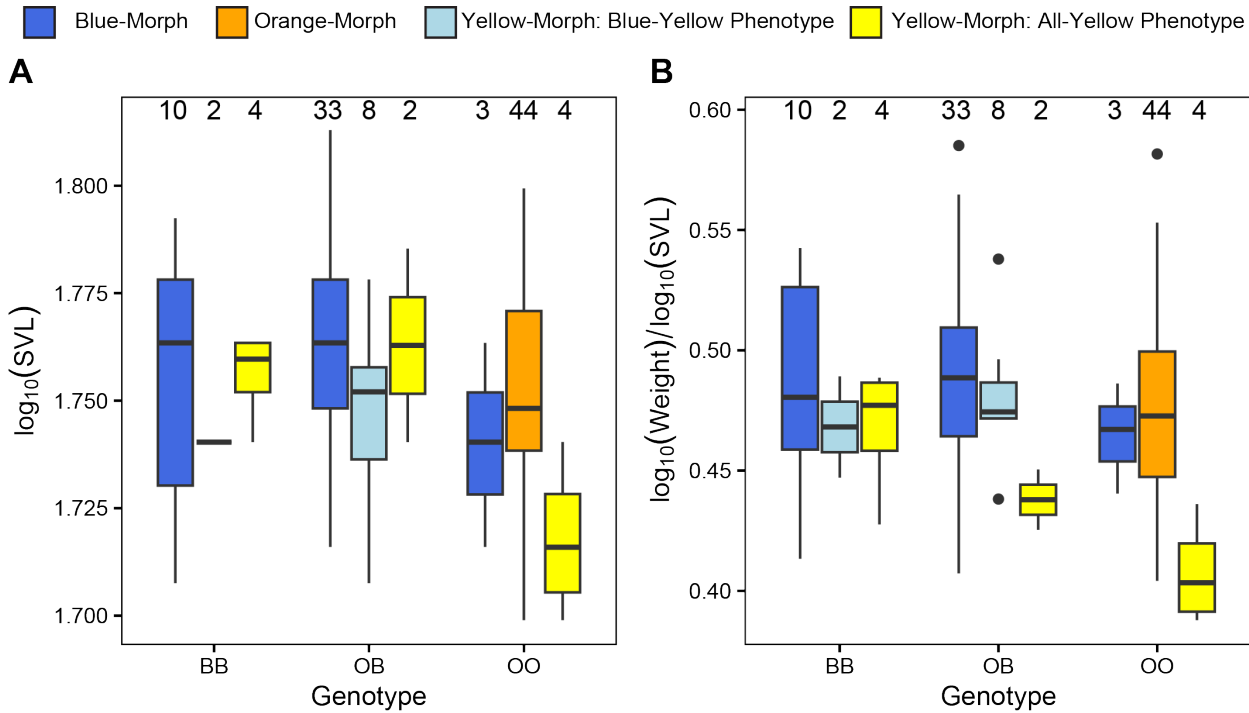


Fig. S7. Variation in body size among the morphs and *SPR* genotypes. **A.** Body size as measured by $\log_{10}$ snout-vent length (SVL) in millimeters. **B.** Standardized weight as measured by $\log_{10}$ weight in grams divided by $\log_{10}$ SVL. Boxplots are colored by the throat color morph, with the boxes showing the 25th and 75th percentiles, a thick line for the median, whiskers for the minimum and maximum, and outliers as separate points. The two different phenotypes of yellow-morphs are shown in separate boxplots. The sample sizes of each category are above each boxplot.

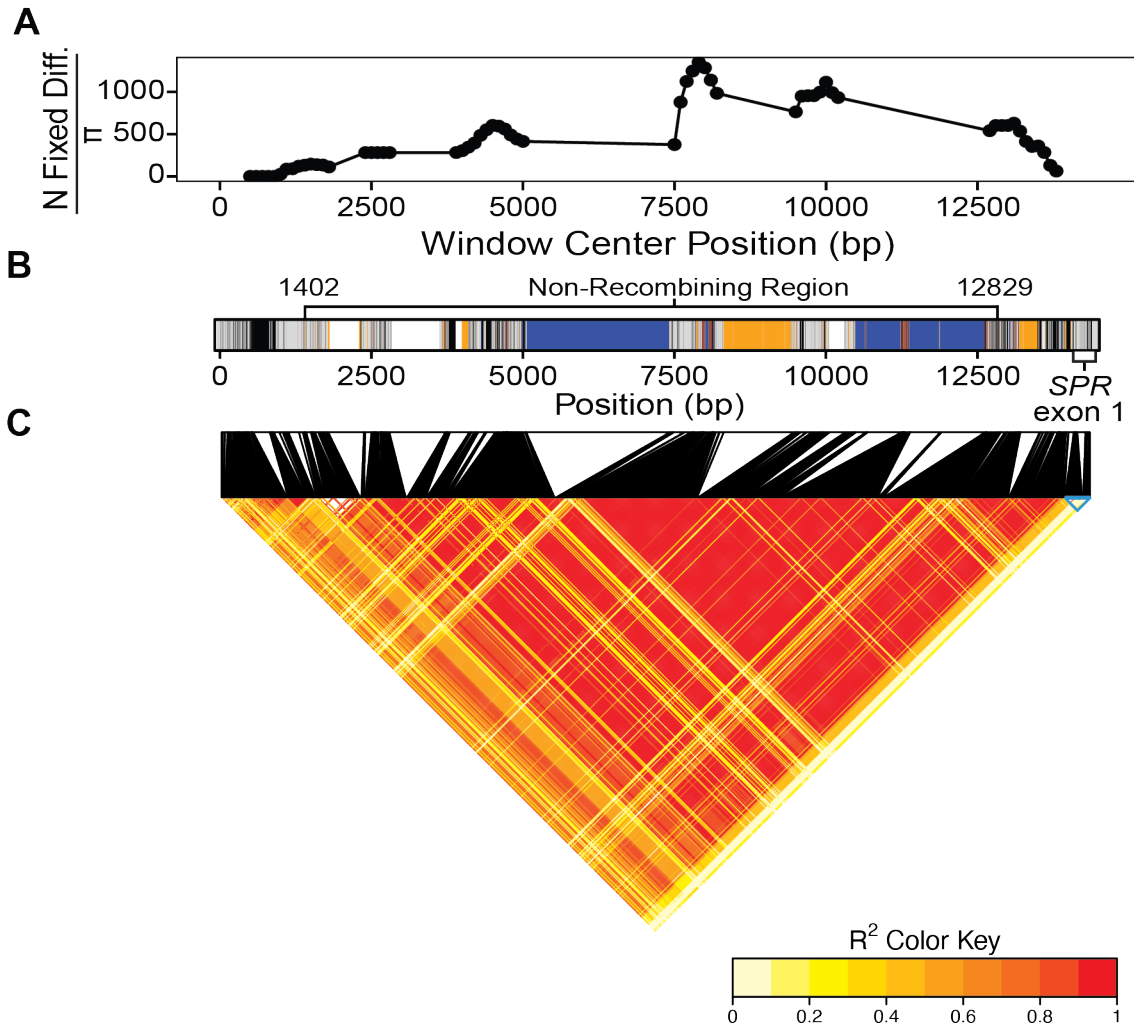


Fig. S8. Sequence characteristics of the alignment of 48 B-haplotypes and 108 O-haplotypes. **A.** The number of fixed genetic differences between the haplotypes standardized by nucleotide diversity (π). **B.** Overview of the haplotype alignment showing positions that were: invariant (gray), variable (black), primarily gaps (white), unique to blue/orange haplotypes (blue/orange), and fixed differences (red). The line above the plot shows the extent of the non-recombining region that lacked shared polymorphism between the O and B-haplotypes. **C.** Linkage disequilibrium amongst single nucleotide polymorphisms (SNPs) in the alignment. Colors in the heatmap from yellow to red indicate low to high levels of linkage disequilibrium. Black lines above the heatmap indicate the physical position of each of the SNPs and the light blue triangle shows the position of exon 1 of the *SPR* gene.

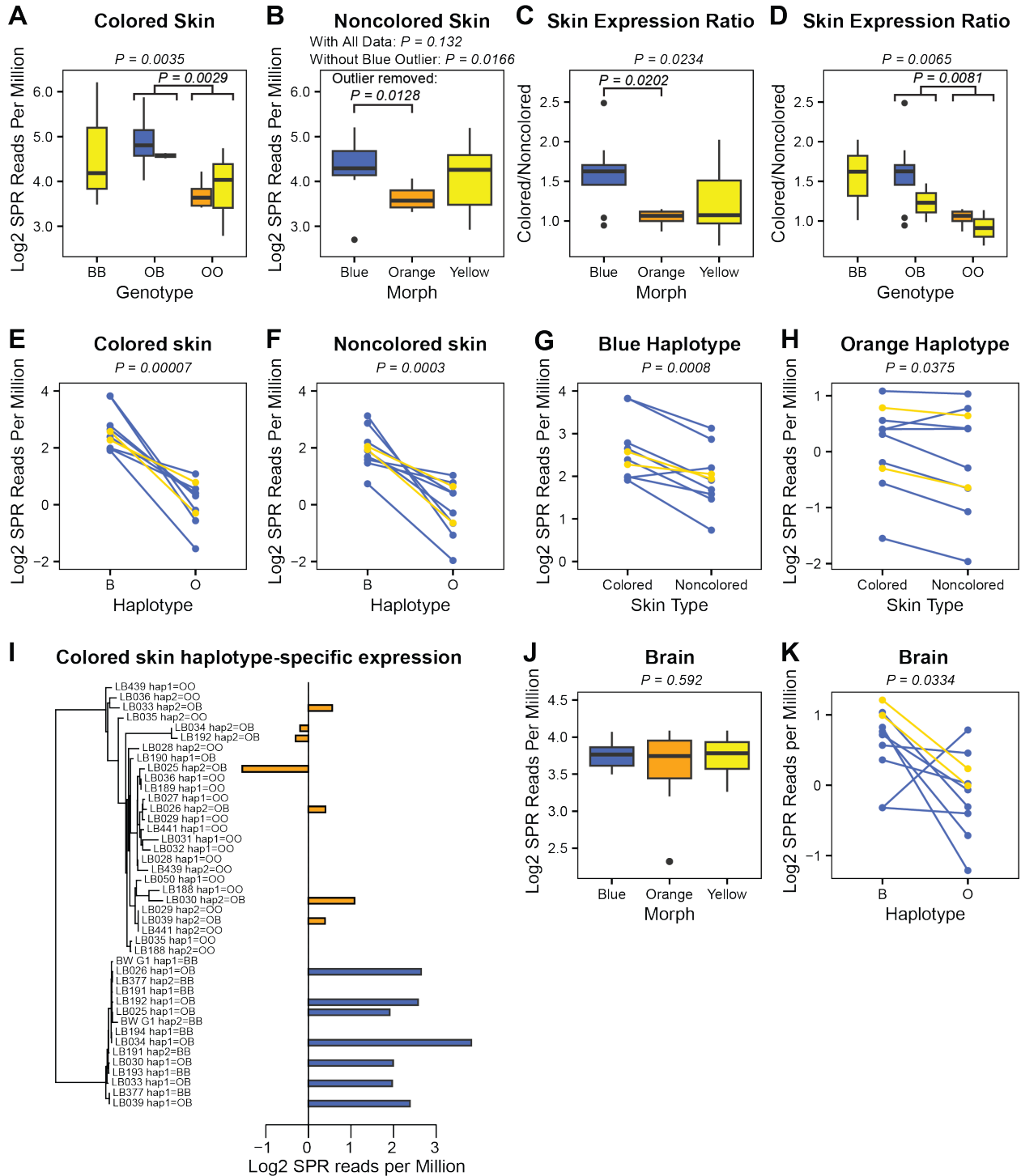


Fig. S9: Expression of the *SPR* gene in different tissues. In panels A-D and panel J, boxplots are colored by the lizard throat color morph, with the boxes showing the 25th and 75th percentiles, a thick line for the median, whiskers for the minimum and maximum, and outliers as separate points. In panels E-H and panel K, points depict levels of haplotype-specific expression within individuals heterozygous for the O and B-haplotypes, with lines connecting the data points for the same individual, and the points/lines are colored by the morph phenotype of the individual (i.e. a yellow or blue-morph). The P -value above a particular panel is the results of a test for differences among genotypes (A, D) or morphs (B, C, J) or haplotypes (E, F, K) or skin

types (G, H). *P*-values within panels reflect significant differences in post-hoc tests among the groups indicated by the lines. **A.** Comparisons of the expression of *SPR* in colored throat skin among lizards with different *SPR*-region genotypes. All 8 orange-morphs had OO genotypes, all 8 blue-morphs had OB genotypes (coincidentally none had BB genotypes), and yellow-morphs had 3 BB individuals, 2 OB individuals (boxplot to the right of that for blue-morphs), and 3 OO individuals. The data is broken down by morph phenotypes in this figure to highlight the importance of the variation in genotypes in yellow-morphs, but the test for expression differences was only based upon the genotypes of the lizards. **B.** Comparisons among morphs in the expression of *SPR* in noncolored throat skin adjacent to the main color patch. Significant differences were only observed if the outlying point for blue-morphs was excluded. **C.** Comparisons among morphs in the expression of *SPR* in the colored throat skin relative to the noncolored throat skin. **D.** Comparisons among genotypes (as in panel A) in the expression of *SPR* in the colored throat skin relative to the noncolored throat skin. **E and F.** Haplotype-specific expression of B versus O-haplotypes in colored (panel **E**) and noncolored skin (panel **F**). **G and H.** Comparisons of the haplotype-specific expression of the blue (panel **G**) and orange haplotype (panel **H**) in colored versus noncolored skin. **I.** Haplotype-specific expression of *SPR* in colored skin plotted on to a phylogeny of the haplotypes. The expression data is colored by whether it came from an O or B-haplotype. Expression data was only plotted for the seven OB individuals that had both of their haplotypes sequenced. **J.** Comparisons among morphs in the expression of *SPR* in brain. **K.** Haplotype-specific expression in the brain.

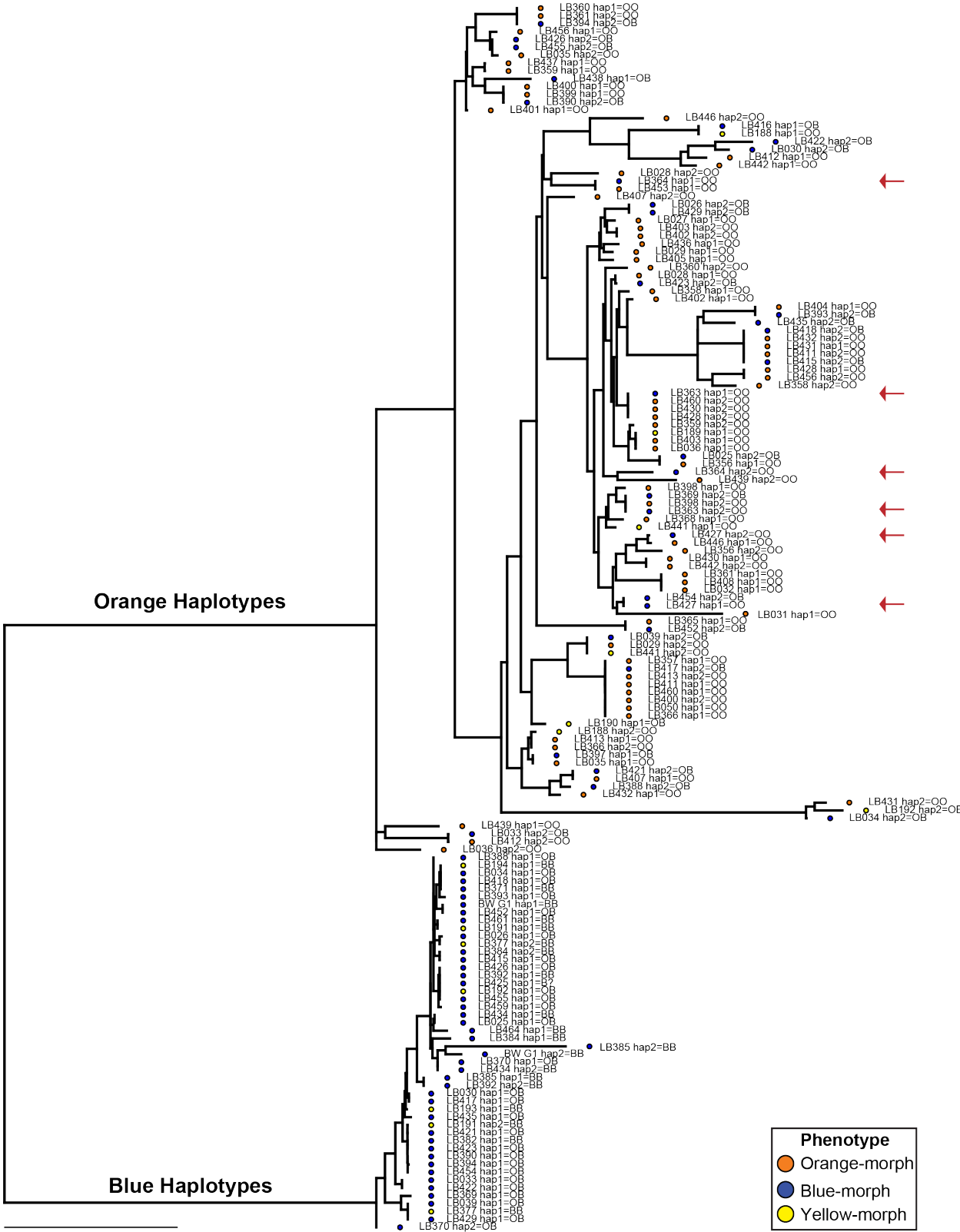


Fig. S10: Maximum likelihood phylogeny of the *SPR* region associated with morph phenotypes. Colored circles at the tips of the branches show the morph phenotype of the lizard associated with each haplotype. The individual identity code of the sample is provided next to

each circle followed by whether it was the first or second haplotype of that individual and then the *SPR* region genotype is given after the equals sign. This phylogeny includes DNA sequences derived from individuals in the genome-wide association study and from the blue-morph used for making a reference genome. Red arrows highlight the position of the haplotypes from the three anomalous blue-morphs with OO genotypes (*i.e.* LB363, LB364, and LB427).

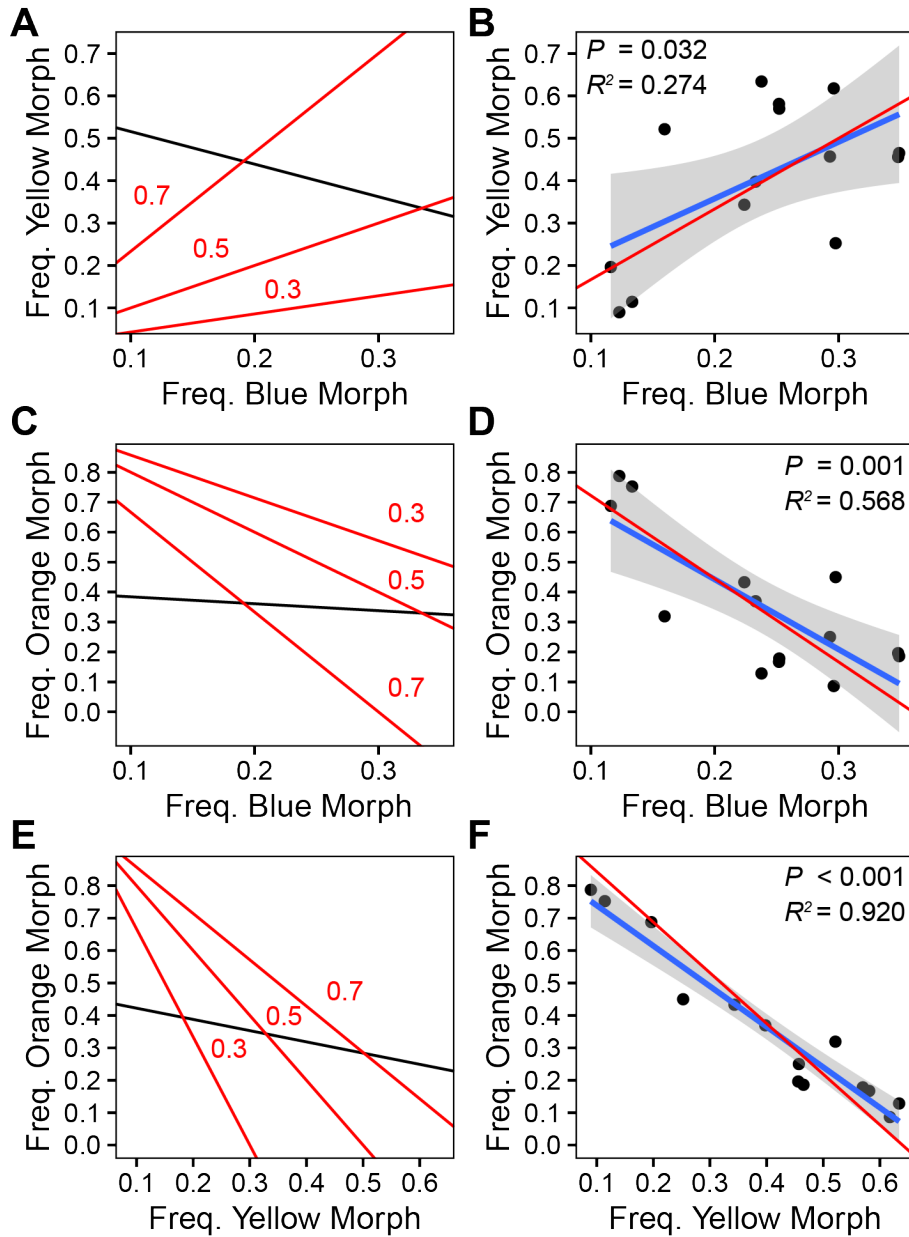


Fig. S11: The predicted (left) and observed (right) correlations between the frequencies of the morphs. A, C, E show predictions from the 3-allele model (black) and the 2-allele model with different proportions of yellow-morphs arising from individuals with the B-haplotype (red lines with the yellow-morph proportion given above the line). B, D, F show the frequencies of the morphs for 14 years of field data (black points) and the correlations observed amongst the frequencies (blue line with 95% confidence intervals in gray). The red line is the two-allele model prediction using an estimate of the proportion of yellow-morphs generated from B-haplotypes from those years of the field data (i.e. 0.64 yellow-morphs). The P -values in the panels are for the correlation observed in the field data.

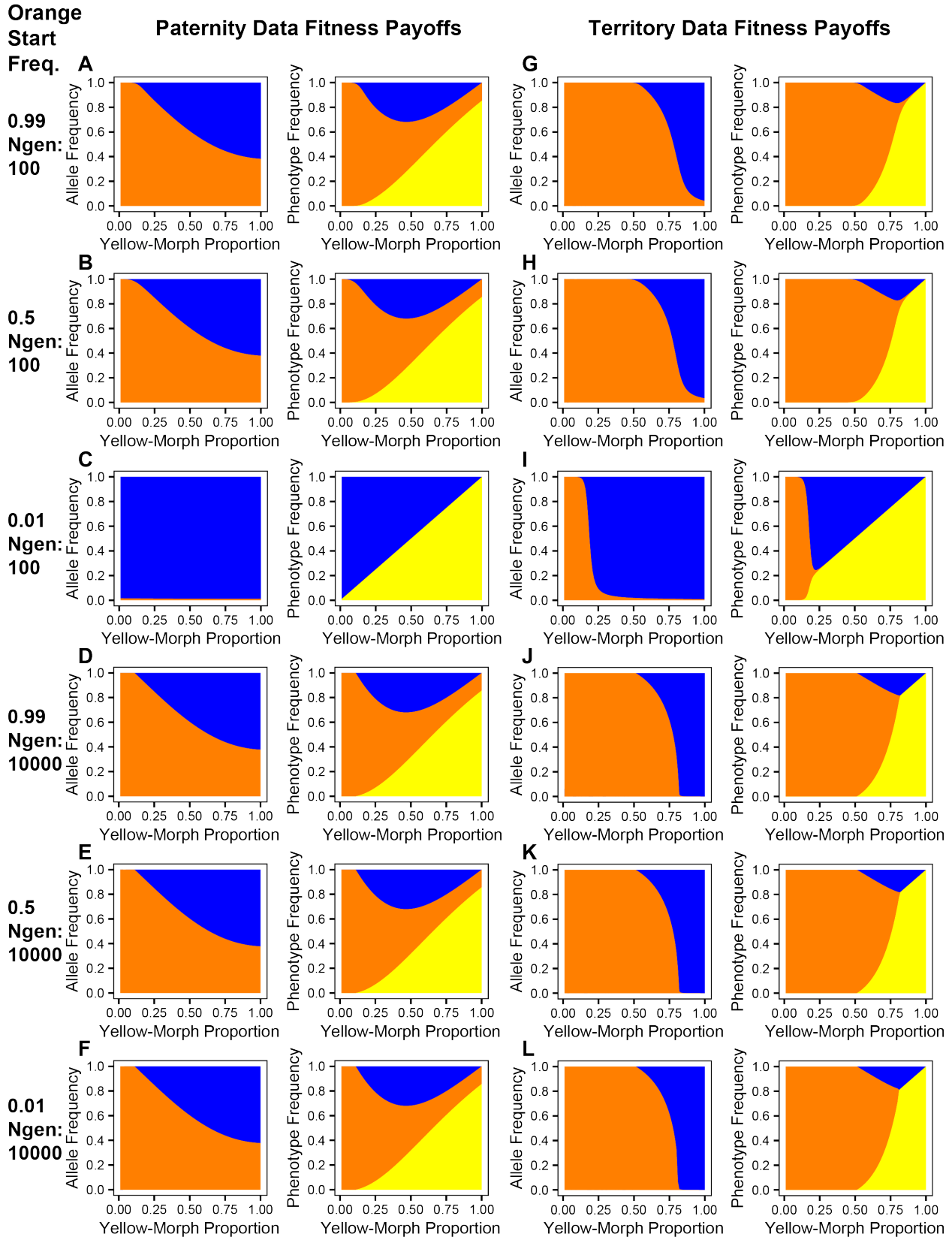


Fig. S12: The allele and phenotype frequencies resulting from simulations varying the proportion of yellow-morphs in the two-allele model and beginning with different allele starting frequencies. For each value of the yellow-morph proportion, the proportion was fixed

across generations. Two different payoff matrices for the morphs were used, with panels **A-F** based on male fitness estimates from paternity data (15, 19) and panels **G-L** based on male fitness estimates from overlap with female territories (6, 19). Panels **A-C** and **G-I** show the allele frequencies observed after 100 generations. Panels **D-F** and **J-L** show the allele frequencies observed after 10,000 generations. The starting allele frequency of the orange allele is shown in bold along the left side of the graphs.

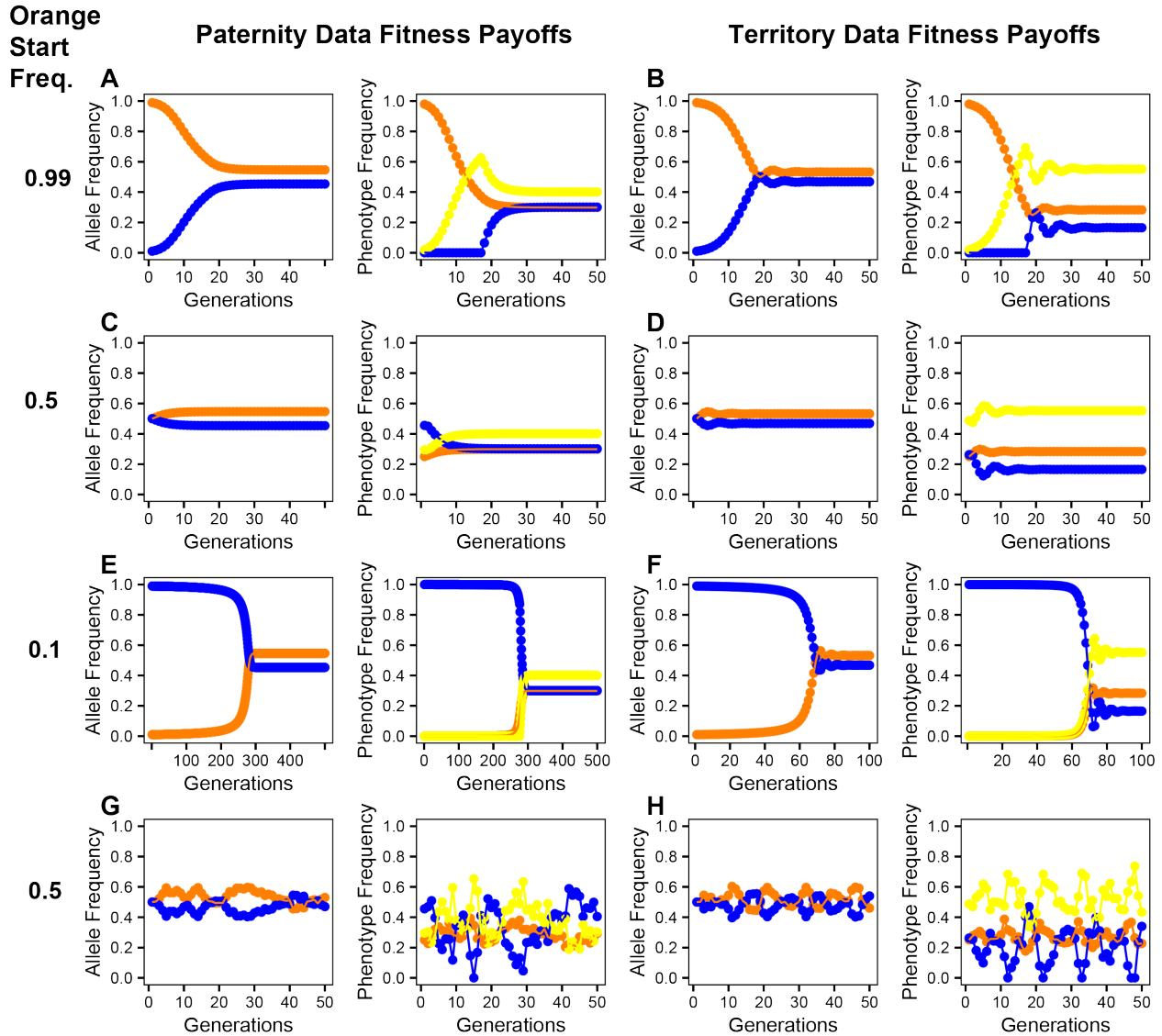


Fig S13: Two-allele models in which the proportion of yellow and blue-morphs within a generation was set to an optimal level given the frequency of the orange-morphs. The frequencies of the two alleles (O vs. B-haplotypes) and the three phenotypes (orange, blue, and yellow-morphs) are shown in their respective colors. Two different payoff matrices for the morphs were used, with the pair of panels on the left based on male fitness estimates from paternity data (15, 19) and the pair of panels on the right based on male fitness estimates from overlap with female territories (6, 19). The starting allele frequency of the orange allele is shown in bold along the left side of the graphs. Panels A-F are deterministic models whereas panels G-H are stochastic because they incorporate genetic drift with a population size of 300 individuals.

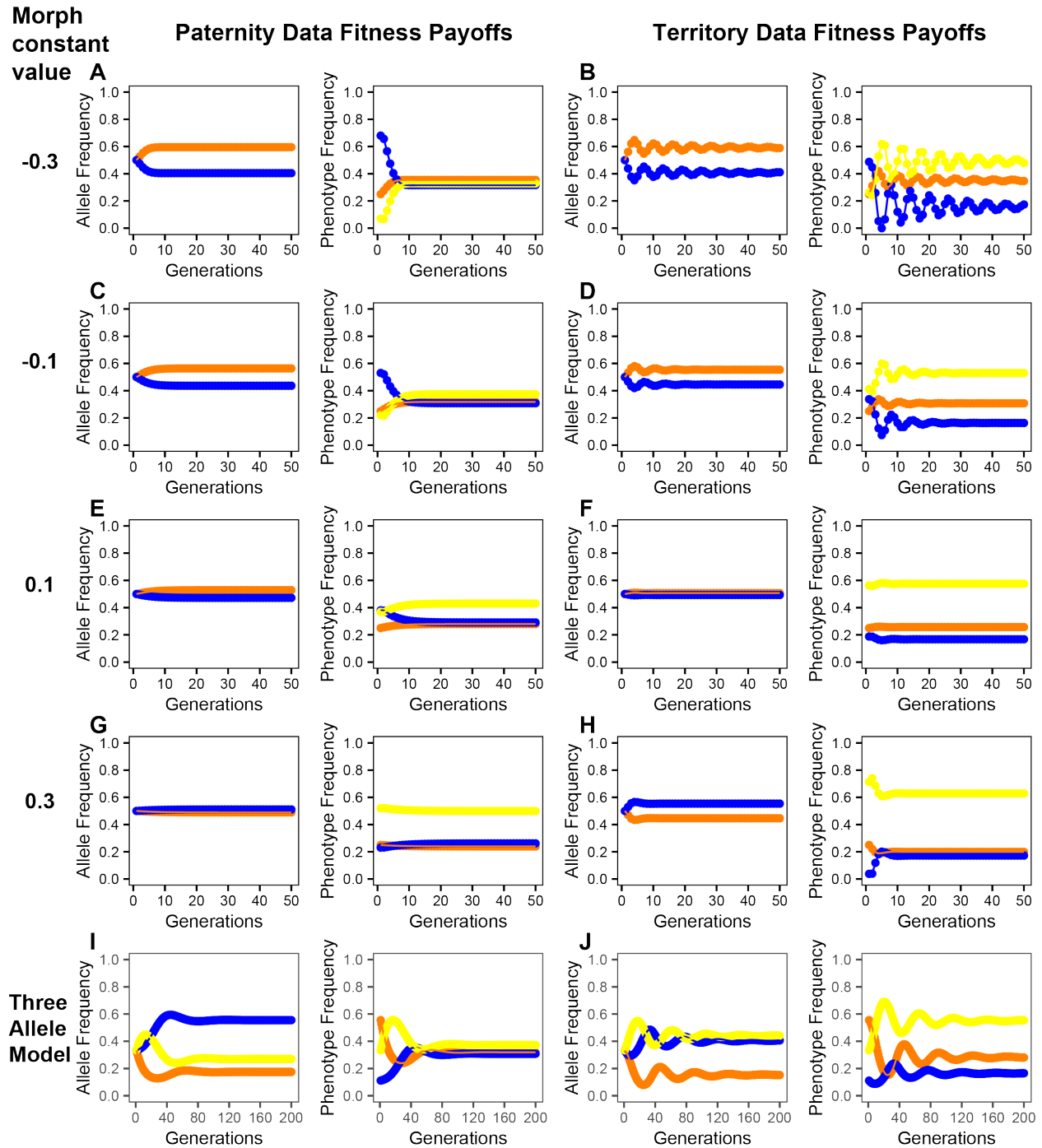


Fig. S14: Two-allele models in which the proportion of yellow and blue-morphs within a generation was set to a semi-optimal level given the frequency of the orange-morphs (A-H) and, for comparison, the three-allele model (I-J). In the two-allele models, the optimal proportion of blue and yellow-morphs was calculated, but was then adjusted by a constant value (the “morph constant value” shown in bold along the left side of the graphs) to maintain a certain proportion of one morph. Positive values of the constant led to an increased fixed proportion of the yellow-morph whereas negative values led to an increased fixed proportion of the blue-morph. The frequencies of the alleles (A-H: O vs. B-haplotypes, I-J: *o*, *b*, *y* alleles) and the three phenotypes (orange, blue, and yellow-morphs) are shown in their respective colors. Two different payoff matrices for the morphs were used, with the pair of panels on the left based on

1971 male fitness estimates from paternity data (*15, 19*) and the pair of panels on the right based on
1972 male fitness estimates from overlap with female territories (*6, 19*). Note the number of
1973 generations differs for the two-allele and three-allele models.
1974

1975

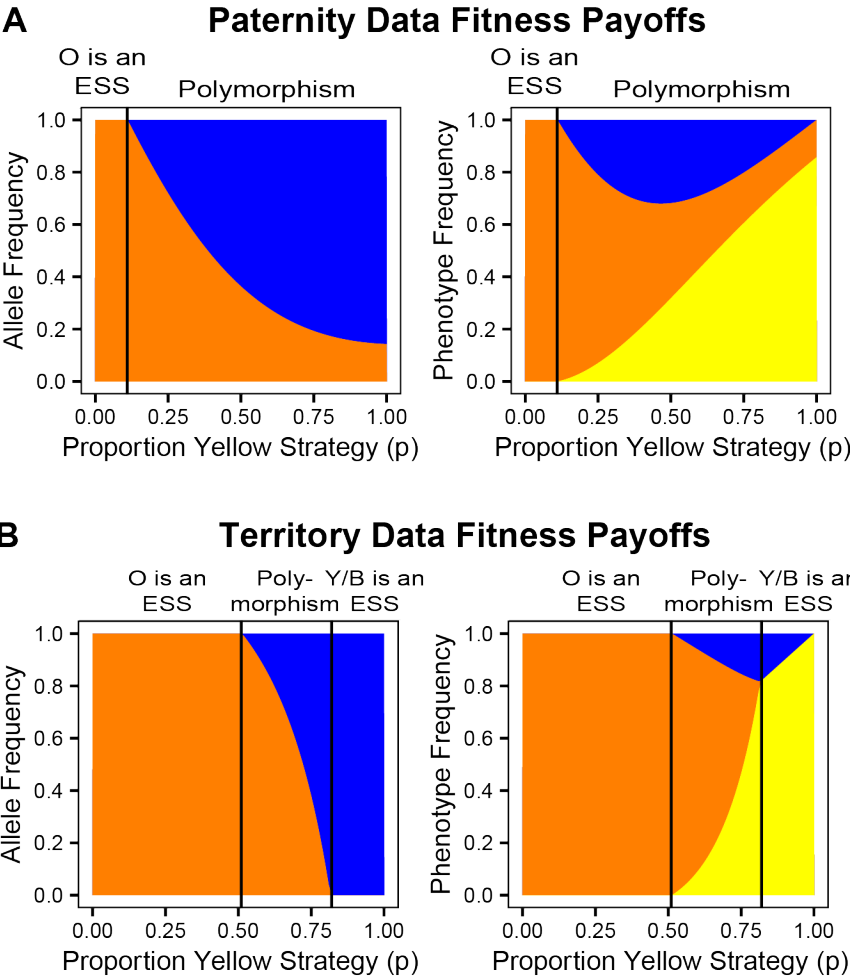


Fig. S15: Game theoretical analysis of the Evolutionarily Stable Strategies (ESS) and the evolutionarily stable states that maintain polymorphism. Vertical black lines indicate transition points between areas where one strategy is an ESS and where there is a stable polymorphism. O individuals play the orange strategy and Y/B individuals play the yellow strategy with probability p and the blue strategy with probability $1 - p$. **A.** Models parameterized with male fitness estimates from paternity data (15, 19). **B.** Models parameterized with male fitness estimates from overlap with female territories (6, 19).

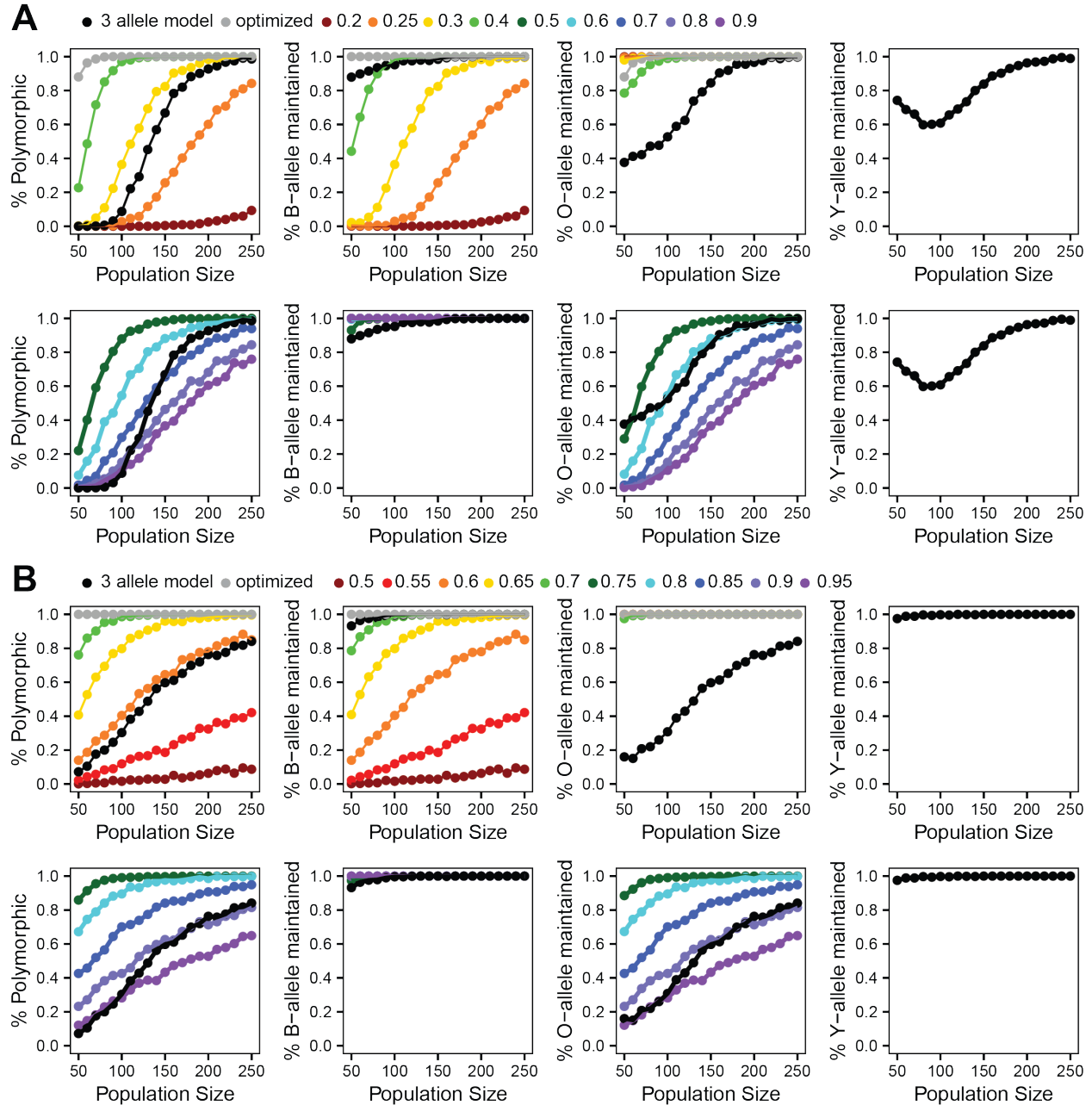


Fig. S16: Maintenance of the polymorphism under different models that incorporated genetic drift. **A.** Models parameterized with male fitness estimates from paternity data (15, 19). **B.** Models parameterized with male fitness estimates from overlap with female territories (6, 19). The sets of graphs in **A** and **B** both show the stability at different population sizes of the two-allele model with fixed proportions of yellow-morphs across generations (colored lines) or optimized within a generation (gray) compared to the three-allele model (black). Note that the fixed proportions of yellow-morphs associated with a particular color differ between panels **A** and **B**. The Y-axis is the proportion of 1000 trials in which the polymorphism was maintained for 1000 generations (panel **A**) and for 100 generations (panel **B**). The panels on the far left show the percentage of trials in which all alleles are maintained whereas the panels going to the right show the percentage of trials in which the B-allele, O-allele, and Y-allele are maintained. Note that the loss of the Y-allele is only relevant for three-allele model.