VOL. 194, NO. 2 THE AMERICAN NATURALIST AUGUST 2019

Phenotypic Diversity Arises from Secondary Signal Loss in the Elaborate Visual Displays of Toucans and Barbets

Meredith C. Miles and Matthew J. Fuxjager*

Wake Forest University, Winston Salem, North Carolina 27101

Submitted October 26, 2018; Accepted February 13, 2019; Electronically published June 17, 2019

Online enhancements: appendix.

ABSTRACT: Complexity and diversity are fundamental characteristics of life, but the relationship between the two remains murky. For example, both gaining and losing complexity can support diversity—so how exactly does complexity influence the emergence of unique phenotypes? Here we address this question by examining how complexity underlies the diversity of elaborate visual displays in the avian clade Ramphastides (toucans and barbets). These species communicate in part by using body movement and colorful ornaments on the tail. We find that sexual size dimorphism predicts the evolution of one specific signal, the tail-cock gesture, implying that tail cocking is more likely to evolve under stronger sexual selection. We also discover process-level constraints on the evolution of complexity: signals are gained along a strict order of operations, where the tail-cock gesture arises before other colors and gestures. Yet virtually any signal can be lost at any time. As a result, many extant phenotypes were more likely to arise through loss of complexity, highlighting the importance of secondary signal loss to phenotypic diversity. Collectively, our results demonstrate how sexual selection catalyzes the evolution of complex phenotypes, which indirectly support diversity by allowing different traits to be modified or lost in the future.

Keywords: complexity, phenotypic diversity, animal communication, display, gesture, signal design.

Introduction

How does biodiversity arise? This question sits at the core of biology, as organic evolution is characterized by explosive growth in phenotypic diversity over surprisingly short time spans (Alfaro et al. 2009). One hypothesis is that diversification is facilitated by the emergence of complexity (Civelek and Lusis 2014), whereby traits interact across developmental, ecological, and evolutionary timescales to constitute an integrated phenotypic system (Hebets et al. 2016). This idea has its origins in the robust correlation between phenotypic diversity and complexity through time (Carroll 2001; Chambers 2014; Vaesen and Houkes 2017). However, the evolutionary mechanisms that explain this relationship remain unclear (Wiens 2017). Therefore, studying how phenotypic complexity itself evolved as traits were gained and lost provides a promising way to understand the origins of biodiversity.

Here we study the evolutionary pathways by which phenotypes gain and lose complexity, aiming to examine how these processes drive trait diversity. We specifically focus on animal displays, which are used for communication. Most displays are complex, which means they are the integrated product of many interacting component signals (Hebets and Papaj 2005). For example, male peacock spiders (Salticidae: *Maratus*) court females using dance displays, which often involve body movements that reveal and/or accentuate colorful ornaments (Girard et al. 2015). A receiver's evaluation of the display will therefore rely on the intersection of gesture and color, as many ornaments are invisible without being revealed by gesture. In various taxa, a similar display can be shared by dozens or even hundreds of species, but a display's complexity means that each signal can differentially evolve to generate a diverse array of unique phenotypes (Hebets and Papaj 2005; Soma and Garamszegi 2015; Miles and Fuxjager 2018*b*; Miles et al. 2018*a*). Moreover, because signaling traits directly mediate fitness-linked interactions like mate choice, they can play an outsized role to facilitate reproductive isolation and even speciation (Maia et al. 2013; Servedio 2016). Evolving increasingly complex displays may therefore support diversity, since complexity offers an expanded capacity for different signals to evolve in response to multiple selection pressures and/or constraints (Hebets et al. 2013; Miles et al. 2017). Indeed, theoretical models of sexual selection highlight the importance of gaining novel courtship traits to support incipient speciation (Yasukawa 1978; West-Eberhard 1983; Prum 2010, 2012; Kopp et al. 2017). But this raises the question: Do displays diverge and diversify only when new signals are gained? One non–mutually exclusive alternative is that some threshold of display com-

---

* Corresponding author; email: mfoxhunter@gmail.com.

ORCIDs: Miles, https://orcid.org/0000-0002-7307-0195; Fuxjager, https://orcid.org/0000-0003-0591-6854.

Am. Nat. 2019. Vol. 194, pp. 152–167. © 2019 by The University of Chicago. 0003-0147/2019/19402-58833$15.00. All rights reserved.
DOI: 10.1086/704088

plexity exists, which is sufficient to support phenotypic diversity. If this is the case, then signal loss may be an alternate trajectory to evolve unique phenotypes. Indeed, loss of traits and/or genes can open up new phenotypic possibilities by facilitating divergent evolution or reconfiguring evolutionary constraints (West-Eberhard 2003). This model may be particularly plausible for displays because signaling traits appear to be lost more often than they are gained (Maia et al. 2016).

Integrated visual displays are some of the most elaborate and diverse phenotypes in the animal world (to human observers, at least), encoding information through the combination of colorful ornaments and ritualized body movement or gesture (Miles et al. 2017; Miles and Fuxjager 2018*a*, 2018*b*). These displays provide an excellent platform to study diversity and complexity because closely related species have often gained/lost different ornaments and gestures. However, the interconnected nature of visual signal components does mean that coevolution between signals may also influence patterns of phenotypic diversity. For example, some authors point to the co-occurrence of color and gesture as an indication that gesture evolved to accentuate preexisting color ornaments (Hasson 1991; Galván 2008; Scholes et al. 2017; but see Prum 1990). This notion is certainly plausible, considering that gesture is often used to modify a signaler's appearance by revealing or concealing body ornaments (Girard et al. 2011; Hongjamrassilp et al. 2018). Still, behavioral traits like gesture appear to be more evolutionarily labile than static ornaments (Blomberg et al. 2003; Miles et al. 2018*b*) and are thus more likely to emerge first. Gestural signals can also function independently of ornamentation (Barske et al. 2011; Manica et al. 2016; Schuppe and Fuxjager 2018). Moreover, some ornaments even function to amplify an underlying signal produced with body movement (Hebets and Uetz 2000). Therefore, a compelling alternative to the "color first" model is one in which gesture evolves first and is later accentuated by color. Rarely (if ever) do studies test between these two ideas.

To address our study's main objective, we investigate a visual display system in the avian clade Ramphastides (fig. 1*a*), a tropical radiation that includes barbets (Megalaimidae, Lybiidae, Capitonidae), toucan-barbets (Semnornithidae),

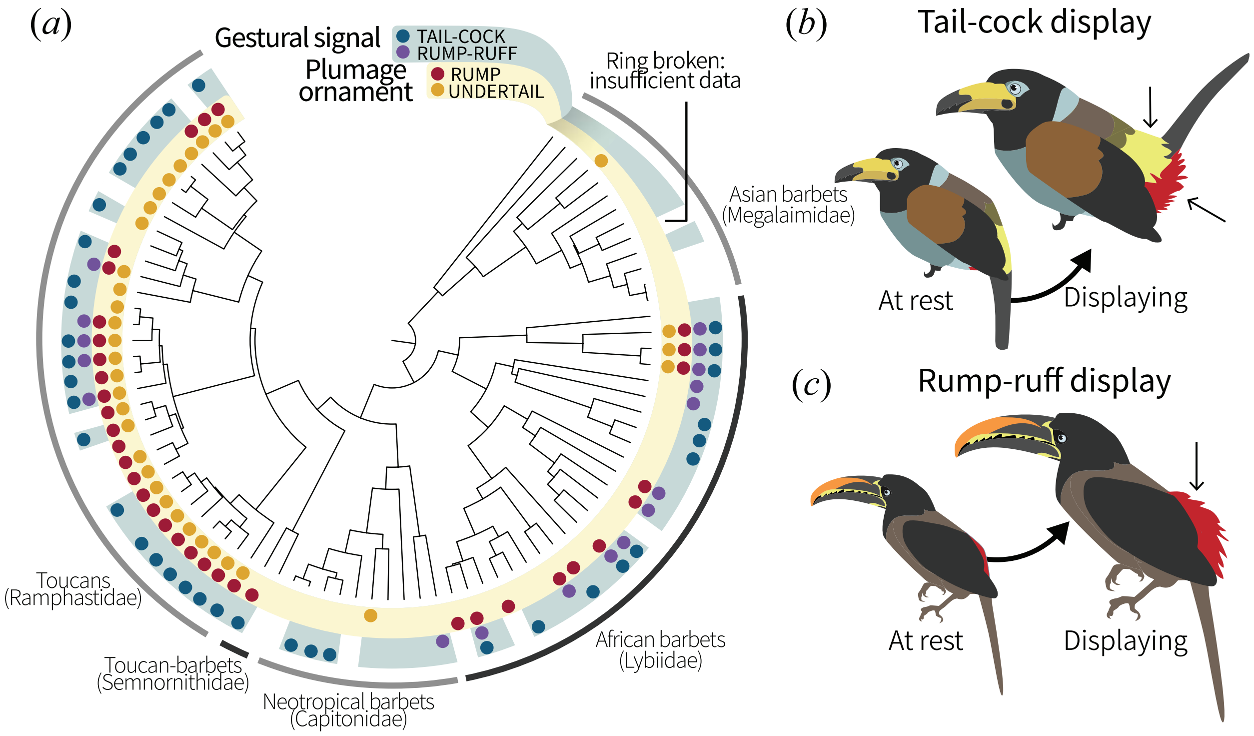


Figure 1: *a*, Phylogeny of the clade Ramphastides and the phenotypic patterning of tail displays shared by species in the group. Display phenotypes are categorized by the presence (circle) or absence (no circle) of each of four visual signals: color ornaments on the rump (red) and undertail (orange) plus the use of tail-cocking (blue) and rump-ruffing (purple) displays. Of 16 possible combinations of these four signals, 13 are found in species today, and each signal has evolved multiple times in different portions of the phylogeny. Each gestural signal is actuated by a distinct form of body movement and may coevolve with plumage ornaments that are revealed while displaying. The tail-cock gesture (*b*) is performed when an individual elevates the tail to be perpendicular to the body, which pushes the rump feathers above the wings and exposes the undertail (small arrows). By contrast, the rump-ruff gesture (*c*) is performed by raising feathers on the rump (ptiloerection), which changes only the appearance of rump plumage (small arrow).

and toucans (Ramphastidae). More than 50 species within this group perform elaborate tail and rump displays (Short et al. 2001). The most common is the tail-cock gesture, produced by raising the tail up to be perpendicular with the body (fig. 1*b*), which increases the exposure of brightly colored feathers on the rump and undertail. Another common gesture is the rump ruff, which is actuated by erecting the long rump feathers without moving the tail itself (fig. 1*c*). Whereas tail cocking will putatively influence how ornaments on both the rump and undertail are perceived, rump ruffing changes the appearance of only rump patches. Together, these signals appear to constitute a phenotypic system, in that they appear to have coevolved to some degree and are remarkably similar across species that are otherwise quite different in their appearance and ecology (Short et al. 2001). Indeed, examining the distribution of all four signals—tail-cocking and rump-ruffing gestures as well as ornaments on the rump and tail—across the clade reveals that nearly every combination of traits has evolved in the group (fig. 1*a*). This provides a compelling opportunity to explore how different pathways of gaining and losing signals explain such phenotypic diversity.

We first evaluated whether each visual signal's evolution is predicted by sexual size dimorphism (SSD), a common index of relative sexual selection pressure in comparative studies of birds and other vertebrates (e.g., Ord and Martins 2006; Klomp et al. 2016; Miles et al. 2018*a*). SSD tends to be highest in species for which competition among males for female mates is greater (Payne 1984; Székely et al. 2000; Dale et al. 2007) and/or when individual males can mate with many females (Webster 1992).

Next, we used discrete trait modeling to characterize the evolutionary process that best explains the phenotypic diversity observed in nature. We did this using a two-step approach. First, we accounted for coevolution between signals by comparing the rates of evolutionary gains and losses under all possible trait combinations, testing whether the presence/absence of one trait influences the probability that another will evolve. If colorful ornaments arose first and were later accentuated by body movement, then (1) a plain ancestor should be more likely to gain color; (2) gesture should be unlikely to evolve unless the color is already present; and (3) when color is lost, the loss of gesture should follow. Likewise, the inverse should be true if gesture evolved first and was later accented by colorful ornaments. Finally, we assemble an omnibus transition matrix that accounts for evolutionary dependencies among traits to explore how the dual processes of phenotypic gains (increasing complexity) and secondary loss (decreasing complexity) work together to explain the evolution of different signal combinations. From this approach, we assess the likelihood that phenotypic diversity is generated when different lineages gain distinct traits. Alternatively, secondary loss may drive diversity. In this case, the evolutionary pathway to high-complexity states should be highly canalized, and losses will dominate the transition network instead.

Methods

Data Collection

We gathered data from the published literature (Short et al. 2001) on gestural displays and plumage ornamentation of barbets and toucans (Aves: Ramphastides), a clade represented by ≈100 species distributed throughout the continental tropics. Of these species, 81 were included in the molecular phylogeny we used for this analysis (Jetz et al. 2012), and the species included represent an even sample across the group. Importantly, barbets evolved recently with a low extinction and high speciation rate (Den Tex and Leonard 2013; Lutz et al. 2013), so data collected from extant species alone are sufficient to support robust evolutionary inference.

Barbets and toucans of both sexes use visual displays for communication during courtship and communication, and for most species ($n = 56$) this includes some form of complex tail display encompassing both color and gesture (fig. 1*a*). Each signal in the tail display is distributed widely across the phylogeny, such that it was gained and lost multiple times in different lineages. Among species with a tail display, ornaments occur on the rump (i.e., the dorsal feather group posterior to the mantle) and/or undertail (i.e., the undertail coverts or vent). Typically, these ornaments are red or yellow, but some black species ($n = 7$) have a contrasting unpigmented (white) rump instead. However, white patches on a dark background are also important in visual signaling, so we still considered these to be cases of rump ornamentation (Pärt and Qvarnström 1997; Galván 2008). Although many bird species use visual signals in the form of ultraviolet-reflective plumage, those in the order Piciformes (which includes barbets and toucans) do not have ultraviolet-sensitive opsin (Ödeen and Håstad 2013). Therefore, any plumage patches that do reflect ultraviolet light could not function as a visual signal and need not be considered. Tail displays also typically involve two common gestures that change the appearance of ornaments on the rump and undertail: (1) tail cocking, in which the tail is raised to be perpendicular with the back (fig. 1*b*), and (2) rump ruffing, or ptiloerection (feather raising), of the rump plumage (fig. 1*c*). Both tail cocking and rump ruffing can accentuate ornamental plumage on the rump, if present, and the tail-cock gesture also uniquely reveals the undertail feathers. In this way, each gesture has the potential to interact with a different combination of ornaments.

To examine how these visual displays evolve, we first scored phenotypes using a simple binary system (table A1; tables A1–A5 are available online): for each trait, a species received either a 0 (trait absent) or a 1 (trait present). To score plumage ornaments, we referenced high-quality handbook plates (del Hoyo et al. 1999) and awarded species a 1 if they possessed a solid plumage patch that contrasted with the wings, back, and tail. We still considered these to be instances of ornamentation, as they occurred in species that were primarily black, and white ornaments on a black background are known to be functionally significant in visual signaling (Galván 2008). With the same system, we next scored presence or absence of gestural displays using reputable species accounts (Short et al. 2001), which we corroborated where possible by viewing videos of displaying birds archived at the Internet Bird Collection. We scored displays only for species with breeding information thoroughly described ($n = 86$), which helps avoid introducing potential false negatives by assuming a species does not have a given display when there simply is not sufficient information available to tell. All scores were conducted blind to phylogeny and other signaling traits. Of the 16 theoretically possible phenotypes, all but three are represented by at least one species (fig. A1; figs. A1, A2 are available online).

Finally, we computed SSD ($\mathrm{SSD} = (\mathrm{mass}_{\mathrm{male}} - \mathrm{mass}_{\mathrm{female}})/\mathrm{mass}_{\mathrm{species}}$; Dale et al. 2007) using sex-specific mass data gathered from published work (Short et al. 2001). In this formulation, higher values indicate species in which males are progressively larger than females, and $\mathrm{SSD} = 0$ corresponds to species in which males and females are identical in mass (and species with negative SSD have females larger than males). Previous work in birds finds that SSD is higher in species undergoing greater sperm competition (Dunn et al. 2001) or species in which males are more promiscuous (Payne 1984; Székely et al. 2000; Dale et al. 2007). This is consistent with the idea that males are larger in populations where competition for mates is more intense (Payne 1984; Price 1984). As a result, SSD is frequently used as a phenomenological index of relative strength of sexual selection pressure in comparative studies (e.g., Webster 1992; Ord and Martins 2006; Klomp et al. 2016; Miles et al. 2018*a*). This also likely applies to members of the clade Ramphastides, considering the high variation in SSD across the group (and particularly among highly aggressive and/or territorial species, such as the *Ramphastos* toucans; Short et al. 2001).

Phylogenetic and Statistical Approach

All analyses were performed in a comparative framework using a random sample of 1,000 trees constructed from sequenced species and based on the Hackett backbone for avian phylogenies (Hackett et al. 2008; Jetz et al. 2012), which represents our most up-to-date understanding of barbet evolution. From this sample, we also generated a maximum clade consensus (MCC) tree in the R package phangorn (Schliep 2011) for replicate use in Bayesian analyses and data visualization. We first used this tree to test whether SSD predicts display evolution, using phylogenetic logistic regression because of our binary response variables (Ives and Garland 2010). Each model was fit using the phyloglm function included in the R package phylolm (Ho and Ané 2014), with species body mass included as a covariate to account for any confounding effect of Rensch's rule (Dale et al. 2007). Because we tested the same hypothesis on four traits, we corrected all P values to control for the false discovery rate (Benjamini and Hochberg 1995). Other analyses were run in BayesTraits V3 (Pagel and Meade 2007) and replicated on both the 1,000-tree sample and MCC tree to ensure that our inference was robust to topological differences across phylogenetic hypotheses.

To model how the four-trait system of barbet tail displays evolved, we first needed to account for coevolution between signals. We triangulated our inference of coevolution with multiple statistical approaches, first conducting frequentist comparisons of model fit and then evaluating change in evolutionary transition rates in a Bayesian framework. We therefore first employed Pagel's test for coevolution, which is simply a fit comparison between different configurations of the general multistate model for inferring discrete character evolution. This approach has been previously described in detail elsewhere (Pagel et al. 2004; Pagel and Meade 2006). Briefly, multistate models are easily expanded to assess how two binary traits evolve with respect to one another by considering each feasible state combination—(0,0), (1,0), (0,1), or (1,1)—to be its own discrete state (although BayesTraits accepts a single data file with each binary trait in its own column as input). The discrete model first estimates a global rate of phenotypic change for the supplied trait and phylogenetic data, and it then estimates the transition rates (q_{ij}, where i is the initial state and j is the final state) between each supplied trait. Here we use the typical independent null model ($\mathrm{df} = 4$) for all fit comparisons, which estimates two rates for each trait (gain rate and loss rate). In its maximally complex form, however, the model estimates four rate parameters for each trait: gain ($0 \rightarrow 1$) when the other trait is absent (0), gain when the other trait is present (1), loss ($1 \rightarrow 0$) when the other trait is absent (0), and loss when the other trait is present (1). By allowing different transitions to vary according to another trait's ancestral state and then statistically comparing the fit of the expanded model to the independent null, we can thus infer how each signal influences the rate at which each other is lost or gained (Pagel et al. 2004; Pagel and Meade 2006). Importantly, the inference of coevolution between discrete traits is readily confounded by shared ancestral history among species, especially for smaller data sets and/or phylogenies (Maddison and FitzJohn 2015). To account for this fact, we also simulated data sets to estimate type I error rate under the null hypothesis of independent evolution for every trait (outlined in detail below). We then replicated our main analysis on the simulated data sets, an approach that allowed us to verify whether we needed to adopt a more conservative threshold for significance testing to account for phylogenetic pseudoreplication (Rabosky 2006; Cooney et al. 2017).

Coevolutionary Analysis

We tested for coevolution in a maximum likelihood framework, first fitting the null model as parameterized in the Discrete: Independent function in BayesTraitsV3. Next, for each pair of signals (e.g., rump ornament/rump-ruff gesture, undertail ornament/tail-cock gesture) we also fit a series of alternative models ($n = 4$ per signal pair) in the Discrete: Dependent module, each time allowing one transition rate (gain or loss in one trait) to adopt a different value when the other trait was present and absent. This allowed us to examine coevolution separately when traits are gained and lost and also independently assess how each trait influences transitions in another. We tested whether each alternative model better explained our data than the null using a likelihood ratio test and adjusted P values within each model family to control for the false discovery rate (Benjamini and Hochberg 1995).

Comparing the fit of maximum likelihood trait models only indicates which models better explain the phenotypic data, and this is not an informative approach to characterize the direction and magnitude of rate differences. For example, the best-fitting model for our data may be one in which rump ruffing evolves at a different rate when another signal is present or absent, but the actual magnitude of the rate change could be positive, negative, or near zero (and thus negligible). To determine the direction and magnitude of rate differences, we estimated transition rate posteriors from the full dependent model with Markov chain Monte Carlo (MCMC), run twice each on the 1,000-tree sample and the MCC tree. For each phylogenetic hypothesis we also sampled from chains conditioned on different hyperpriors (exponential and gamma distributions). Each chain ran for 1 million iterations after a 100,000-generation burn-in, and we concatenated the posterior sample (thinning interval = 10,000) for diagnostic analysis in the R package coda (Plummer et al. 1999). After verifying that chains were mixing (the BayesTraits V3 autotuner was sufficient to achieve a random walk through likelihood and parameter space and an acceptance rate ≈ 0.23), we log transformed each transition rate q and computed posterior rate difference samples ($\Delta q = q_{\text{trait present}} - q_{\text{trait absent}}$). In these instances, a $\Delta q > 0$ reflects a higher transition rate when a second trait is present, whereas $\Delta q < 0$ indicates a lower rate when present. We summarized the Δq sample by computing its mean and interquartile range (IQR) and finally calculated the posterior difference rate as the proportion of samples where $q_{\text{trait present}} > q_{\text{trait absent}}$.

In summary, we used a combination of frequentist fit comparison and Bayesian estimation of rate parameters to infer the degree to which signals evolved dependently on one another. From these two approaches, we adopted a conservative set of criteria to determine whether a given evolutionary transition was truly dependent. Thus, we considered an evolutionary transition (gain or loss in a signal) to depend on another signal's ancestral state (present or absent) only if each of the following was true: (1) the model where the transition rate was allowed to change better explained the data when compared with the null, (2) the posterior rate difference did not cross zero (mean and IQR are all positive or all negative), and (3) the same posterior difference (either positive or negative) was recovered in >85% of MCMC samples.

Accounting for Phylogenetic Pseudoreplication: Null Model Approach

On smaller data sets and/or phylogenies, testing for coevolution between discrete traits can incur inflated type I error if there is insufficient representation of independent evolutionary transitions. To ensure that our inference was robust to such phylogenetic pseudoreplication, we used a null model approach to estimate type I error and, if necessary, adjust thresholds for significance (Rabosky 2006; Cooney et al. 2017). We therefore first estimated evolutionary rates for each trait assuming no coevolution (i.e., each trait evolves independently). We then generated 100 simulated data sets on the basis of these independent rates using the R package phytools (Revell 2012) function sim.history. In all instances where we found statistical support for inferring coevolution in our original data set, we retested the same trait pair with the simulated data. This allowed us to compute type I error rate and construct a null distribution based on the χ^2 values from the 100 different likelihood ratio tests performed on the simulated data. Thus, the 95th quantile of this distribution represents the appropriate threshold for concluding the two traits coevolve (fig. A2). However, it appears that type I error was not appreciably inflated in our data set, as we recovered an average error rate of $3.93\% \pm 1.4\%$ (standard deviation; minimum = 2%, maximum = 6%). Nonetheless, we verified that χ^2 values obtained in our original analysis were greater than the adjusted threshold (table A2) before concluding that coevolution occurs between two traits.

Evolutionary Transition Network

Our central aim was to construct a complete transition matrix describing how complex display phenotypes evolved as different signals were gained or lost, which we could then use to characterize the trajectories by which phenotypic diversity in tail displays arose. However, simultaneously computing the transition rates required to encompass all 13 phenotypes is unlikely to be accurate for small data sets like this one. To address this, we instead ran series of two-trait discrete models and constructed a multimodel weighted-average matrix that took coevolution between traits into account.

In BayesTraits, we used reversible jump MCMCs to estimate posterior transition rates when modeling each pair of traits under the Dependent Discrete function. For each of these models, there are >21,000 possible parameter configurations, which makes finding the best-fit parameterization difficult (if not impossible) through iterative fit comparison alone. We therefore estimated rate posteriors using reversible jump MCMC, which integrates values drawn across all possible parameter configurations weighted by model likelihood (Pagel and Meade 2006). To ensure that our samples were robust to permutations in tree topology and prior selection, we ran three replicate models for each combination of two different hyperpriors (exponential and gamma distributions) and on the 1,000-tree sample as well as the maximum clade credibility tree (although model output was similar across priors and phylogenies). The posterior was drawn from 10 million samples drawn every 1,000 generations after a 1,000,000-iteration burn-in. Chain mixing was thorough when using the built-in BayesTraits autotuner, which we verified by confirming there was no autocorrelation and that acceptance rates were ≈23%.

To construct our final transition matrix (table A3), we integrated the parameters derived from our two-trait models into a framework that evaluates how the entire visual phenotype evolved. However, evaluating the simultaneous evolution of multiple dependent traits requires accounting for the possibility of multiple dependencies: if trait A is more likely to evolve when trait B $= 1$ and trait C $= 1$, which trait A parameter should we use in the situation when trait B $= 0$ and trait C $= 1$? We addressed this issue by computing weighted averages for each parameter based on the degree to which a dependent model explained the data better than an independent model (i.e., based on the level of support for coevolution of any two traits). Thus, if fitting a model when trait A is dependent on trait B (while ignoring trait C) provides a larger increase in fit compared with the independent model than one where trait A is dependent on trait C, the weighted average can take both dependencies into account while prioritizing the influence of trait B's state. The weights themselves were derived from log Bayes factors based on each model's log marginal likelihood ($\log \mathrm{BF} = 2 \times (\log \mathrm{likelihood}_1 - \log \mathrm{likelihood}_2)$), which provides a continuous measure of how much information was gained when switching from an independent to a dependent model. We estimated the log marginal likelihood for models using the stone sampling method (Xie et al. 2011; 100 stones sampled for 10,000 iterations, $\alpha = 0.4$, $\beta = 1.0$) and then averaged this across the 12 replicates of each dependent and independent model. Altogether, the transition matrix was constructed using transition rates drawn from all possible models, and transition rates for traits that are highly dependent on one another were given more weight in the final output.

Complexity and Diversity

Complex displays are made up of more component signals compared with simple displays (Hebets et al. 2016; Miles et al. 2017, 2018*a*); thus, displays gaining signals over evolutionary time are becoming more complex, whereas displays transitioning into lower-complexity states would do so via secondary signal loss. Because the processes by which phenotypes evolve to be more or less complex are partitioned into trait gains and losses, we compared the distribution of transitions between the two. Transition rates were heterogeneous across the matrix (table A3), so we used the standardized residuals from a χ^2 test (table A4) to eliminate transitions that were uncommon enough to be considered nonsignificant (Egge et al. 2011; described in Green and Patek 2018). Briefly, only cells with standardized residuals <1.96 are considered significant at $\alpha = 0.05$, and others are removed from the transition network. The remaining values are therefore the dominant pathways by which the integrated phenotype evolves.

After removing nonsignificant values ($n = 17$ gains and $n = 5$ losses), we used the reduced matrix to ask whether more phenotypic combinations were accessed by gaining or losing signals (table A5). To do this, we compared the observed count distribution of nonsignificant and remaining transitions among gains and losses with a χ^2 test (computed with Yates correction) to assess whether the distribution of transitions differed among the two processes. If phenotypic complexity underlies diversity, then there should be a difference in the number of transitions remaining when signals are gained compared to when they are lost. Finally, we ran post hoc χ^2 tests on gains and losses individually to test whether an underlying difference between gains and losses was driven by an increase or decrease in nonsignificant transition count.

Results

Sexual Size Dimorphism

We first aimed to uncover whether sexual selection might be a driver that explains why different visual signals evolve. To do this, we used SSD ($\mathrm{SSD} = (\mathrm{mass}_{\mathrm{male}} - \mathrm{mass}_{\mathrm{female}})/\mathrm{mass}_{\mathrm{species}}$) as a phenomenological indicator of relative sexual selection pressure. When SSD is zero, males and females are exactly the same size. Large positive values reflect proportionally larger males, while greater negative values reflect larger females instead. SSD does not provide an absolute estimate of sexual selection pressure on phenotypic evolution (various other drivers can also underlie sex-specific size variation). However, it is effective as a relative index when applied across many species that exhibit variation in SSD and life-history traits relevant to sexual selection. This is certainly the case for clade Ramphastides, where both SSD and social behavior varies from species to species (Short et al. 2001).

With this in mind, we calculated SSD for each species on the basis of published body mass values (Short et al. 2001). SSD positively predicted whether species would evolve a tail-cock gesture (fig. 2*a*; phylogenetic logistic regression: $z = 2.67$, $P = .012$), where species with higher SSD were more likely to have the signal than those that did not. Because SSD can be confounded by overall body size, we also verified that this effect was independent of species body mass ($z = 0.51$, $P = .607$). By contrast, we found no evidence that differences in SSD predict whether a species would evolve the rump-ruff gesture (fig. 2*b*; $z = 1.72$, $P = .115$) or plumage ornamentation on the rump (fig. 2*c*; $z = 1.97$, $P = .193$) or undertail (fig. 2*d*; $z < 0.00$, $P = .999$). These data therefore suggest that tail-cocking gestures are more likely to be gained by species where males are proportionally larger than females (rather than being similar or smaller in size).

Coevolution

Because each component in a visual display can influence how the others function, we tested for evolutionary dependency between each pair of signal components. We did this using a series of discrete models, each of which reconstructs how a pair of binary traits evolved along the phylogeny. For each pair of traits (e.g., tail-cock gesture and rump ornament), we first fit a null model with four parameters that described the rate of gain and loss in each trait and then com-

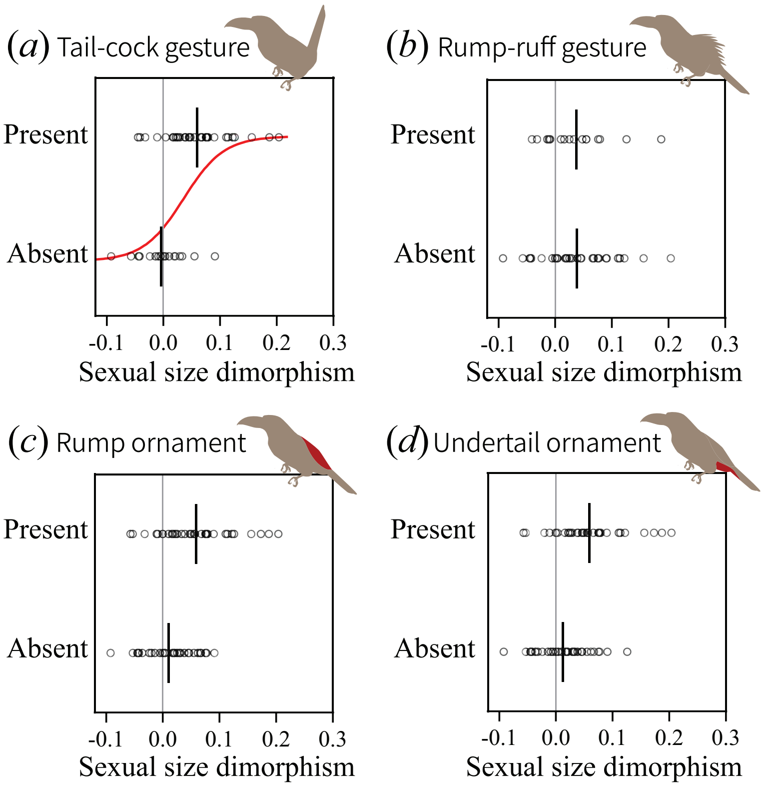


Figure 2: Variation in sexual size dimorphism (SSD) among species, partitioned by the presence or absence of signals in the tail display system: tail cocking (*a*), rump ruffing (*b*), rump ornament (*c*), and undertail ornament (*d*). Each circle represents one species, with the vertical line reflecting the multispecies mean. SSD characterizes size differences between the sexes: at 0, males and females are identical in mass, while greater positive values reflect larger males and greater negative values reflect larger females. In a phylogenetic logistic regression that accounts for species body mass, SSD significantly predicts the probability that a species has the tail-cock gesture (red line; $z = 2.67$, $P = .012$; *a*).

pared this with an alternative model where we allowed a focal transition (e.g., rate of gaining rump ornament or rate of losing tail cocking) to take a different rate value when the second trait was present and absent.

We initially examined whether coevolution influences how traits are gained (table 1; fig. 3*a*), the type of transition marked by an increase in visual display complexity. First, our data were best explained by models where rump orna-

Table 1: Statistical summary of models evaluating whether a given trait's presence (predictor) influences the rate of gaining (+) or losing (−) another trait, either a gesture (G) or an ornament (O)

Ancestral trait and focal transition	Log likelihood$_0$	Log likelihood$_A$	χ^2	P	Δq (IQR)
Gesture:					
Rump ruff:					
+G: tail cock	−70.89	−70.52	.74	.390	2.99 (1.5–4.5)
+O: rump	**−72.27**	**−68.94**	**6.7**[a]	**.020***	**3.22 (1.9–4.5)**
+O: undertail	−56.15	−56.02	.26	.742	.61 (−.67 to 1.9)
Tail cock:					
+G: rump ruff	−70.89	−70.95	.12	.836	.78 (−.77 to 2.2)
+O: rump	**−74.80**	**−70.45**	**8.7**[a]	**.006****	**2.67 (1.4–3.8)**
+O: undertail	−64.04	−60.33	7.4	.013*	−1.01 (−2.2 to .26)
Ornament:					
Rump:					
+G: rump ruff	**−72.27**	**−68.69**	**7.2**[a]	**.015***	**1.86 (.77–2.8)**
+G: tail cock	**−74.80**	**−67.28**	**15.0**[a]	**.0002*****	**3.83 (2.5–5.1)**
+O: undertail	−65.42	−62.97	4.9	.027*	−.21 (−1.2 to .76)
Undertail:					
+G: rump ruff	−56.15	−56.15	.01	.923	−.68 (−1.9 to .49)
+G: tail cock	−64.04	−63.40	1.3	.261	.45 (−.65 to 1.7)
+O: rump	−65.42	−65.30	.25	.618	−.60 (−1.6 to .58)
Gesture:					
Rump ruff:					
−G: tail cock	−70.89	−70.33	1.1	.292	2.2 (1.2–3.3)
−O: rump	−72.27	−70.19	4.2	.041*	.96 (−.39 to 2.2)
−O: undertail	−56.15	−56.14	.03	.908	.68 (−.51 to 1.9)
Tail cock:					
−G: rump ruff	−70.89	−70.68	.41	.696	.69 (−.64 to 1.9)
−O: rump	−74.80	−70.44	8.7	.003**	−.65 (−1.9 to .31)
−O: undertail	**−64.04**	**−61.92**	**4.2**[a]	**.040***	**−2.5 (−3.9 to −1.2)**
Ornament:					
Rump:					
−G: rump ruff	−72.27	−68.95	6.7	.015*	−.29 (−1.4 to .58)
−G: tail cock	−74.80	−73.25	3.1	.078	1.6 (.41–2)
−O: undertail	**−65.42**	**−60.66**	**9.5**[a]	**.004****	**−3.01 (−3.96 to −1.8)**
Undertail:					
−G: rump ruff	−56.15	−56.12	.06	.883	−.63 (−1.8 to .42)
−G: tail cock	−64.04	−61.68	4.7	.060	−1.4 (−2.4 to −.57)
−O: rump	−65.42	−63.69	3.5	.125	−.86 (−3.9 to −.02)

Note: In each case, we compared the fit of a null model to an alternative where the focal transition rate was estimated separately when the predictor rate was present and when it was absent. Log transition rate difference (Δq) was estimated from the alternative model via Markov chain Monte Carlo. Positive Δq values thus indicate that a focal transition was more frequent when the predictor trait was present, whereas negative values indicate that the transition was less likely, and values near 0 suggest there is no difference. Finally, for each case supporting coevolution, we also simulated data sets under the null hypothesis (traits evolve independently) to verify that type I error was <5% or alternatively compute an adjusted threshold for significance. Thus, we infer that two traits coevolve only if (1) the alternative model better explains the data ($P < .05$ after correction for multiple testing), (2) Δq is nonzero in that model, and (3) the alternative model is still supported when compared with a simulated null distribution. Boldface reflects models supporting an inference of coevolution according to our criteria. IQR = interquartile range.

[a] Test statistic also exceeds critical value computed from null distribution for a given ancestral trait/focal transition pair (table A2 [available online]; for more information, see "Methods").

* $P < .05$.

** $P < .01$.

*** $P < .001$.

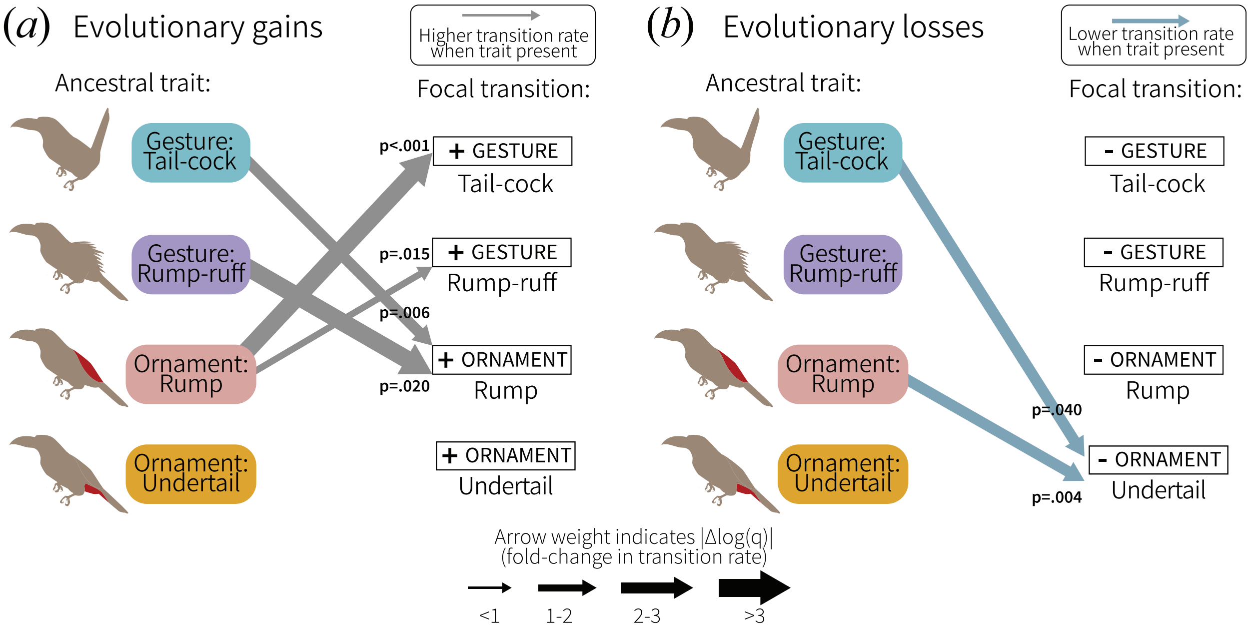


Figure 3: Summary of coevolution between signals in the tail display system. We separately assessed whether the presence of one signal influences the probability that another will be gained (*a*) or lost (*b*). In each panel, ancestral signals are aligned on the left (shaded boxes), and evolutionary transitions are on the right (unshaded boxes). Arrows pointing from an ancestral trait to a given transition reflect statistical support for coevolution, whereby the ancestral trait's presence predicts an increased rate of gaining another signal (*a*) or a decreased rate of losing one (i.e., increased rate of retention; *b*). Arrow thickness corresponds to the magnitude of the difference between the log-transformed transition rate (i.e., fold change) when a given trait is present and when it is absent. For each significant result depicted, we verified that our conclusions were robust to phylogenetic pseudoreplication with a null model simulation (fig. A2 [available online]).

ments were more likely to evolve in lineages that already had a tail-cock gesture ($\chi^2 = 8.71$, $P = .006$) or rump-ruff gesture ($\chi^2 = 6.67$, $P = .020$). In other words, rump ornaments evolved approximately 2.7 times (IQR, 1.4–3.8) more often when a tail-cock gesture evolved first (posterior difference = 94.3%) and 3.2 times (IQR, 1.9–4.5) more frequently when the rump-ruff gesture was already present (posterior difference = 89.0%). At the same time, species with an ancestral rump ornament were also more likely to gain a gesture: they were 3.8 times (IQR, 2.5–5.1) more likely to begin tail cocking ($\chi^2 = 15.0$, $P = .0002$; posterior difference = 96.5%) and 1.9 times (IQR, 0.8–2.8) more likely to gain rump ruffing ($\chi^2 = 7.17$, $P = .015$; posterior difference = 94.0%). Color ornaments on the rump thus tended to be gained alongside both gestures, but it is impossible to distinguish whether one precedes the other from coevolution alone. By contrast, undertail ornaments appear to be gained independently of the other signals; lineages were also no more likely to gain another visual signal when undertail ornaments were present (table 1).

For the most part, both gestural signals and color ornaments were lost independently of one another (fig. 3*b*; table 1), although we did uncover evidence that secondary loss of undertail ornamentation is influenced by the presence of a tail-cocking gesture ($\chi^2 = 4.2$, $P = .04$) and/or rump ornament ($\chi^2 = 9.5$, $P = .004$). More specifically, lineages that had both an undertail ornament and the tail-cock gesture were 2.5 times (IQR, 3.9–1.2) less likely to lose the signal (posterior difference = 87.5%). In other words, coloration on the undertail was retained more often in species that also used a gesture that can accentuate the ornament. Similarly, our data are best explained by a model in which undertail ornaments were retained more frequently when a species also had a rump ornament ($\Delta q = -3.01$ [IQR, -4.0 to -1.8]; posterior difference = 98.9%).

For each model above where we uncovered statistical support for coevolution, we also ensured our results were robust to phylogenetic pseudoreplication (Maddison and FitzJohn 2015) using a null modeling approach (modeled after Rabosky 2006). After simulating data sets where each trait evolved independently on the basis of realistic rates, we reanalyzed the null data and computed type I error rate (mean: 3.93% ± 1.4%; maximum: 6%). With this output, we also computed adjusted χ^2 thresholds for significance testing and verified that our original test statistics exceeded this new critical value (table 1). Each original effect suggesting coevolution takes place was preserved.

In summary, each signal in the tail display system appears to undergo some form of correlated evolution with another. When gaining novel traits, it was more common for lineages that use either tail cocking or rump ruffing to gain a rump ornament. However, they were also more likely to gain one of the two gestures when rump ornaments were already present. Meanwhile, color ornaments appear to be lost more often in lineages that lack a gesture.

Underpinnings of Phenotypic Patterning

The evolutionary process driving visual signal complexity in these birds appears to be one marked by sexual selection for a gestural display (tail cocking) and apparent coevolution among different signaling traits—but how does this process generate the diversity of phenotypes observed in nature today (fig. 1*a*)? Of the extant barbets and toucans, all theoretical combinations of our four focal signals (n = 16) have been documented in the wild except for three (table A1), which highlights the variety of feasible phenotypic outcomes. At the same time, some phenotypic transition rates were influenced by the state of another trait (fig. 3)—in other words, a species' chance of gaining a given signal change is conditional on the presence/absence of another—which means that reconstructing how an entire visual display evolves must synthesize information from multiple two-trait models. To do this we used reversible jump MCMC, which samples all possible parameterizations of the models describing how each trait pair evolved. We then constructed a multimodel transition matrix (table A3) that synthesized the results from each model, weighting parameters by the log Bayes factor (computed using log marginal likelihood estimates from the full dependent model and a constrained independent null). This allowed us to ensure that parameters from models in which one trait was highly dependent on another contributed more to the final average.

From the final transition matrix, we eliminated rare (statistically nonsignificant) transitions from the network using χ^2 standardized residuals (Egge et al. 2011; Green and Patek 2018), which left 42 of 64 transitions remaining (66%; table A4). The majority of rare transitions were phenotypic gains (n = 17), which resulted in a network with few available routes to gain new signals. Instead, a plain ancestor was likely to gain only the tail-cock gesture but not any of the other three signals. Color ornaments were gained only in lineages that already used tail cocking. In a similar vein, the rump-ruff signal emerged only after color ornaments were present, suggesting that color (and thus also the tail-cock gesture) were a prerequisite for this gesture to arise. Importantly, the small number of common gains means that several phenotypes were unlikely to evolve through a process of signal gains alone (n = 6; fig. 4*a*). Because these orphaned phenotypes were inaccessible by gaining traits, this

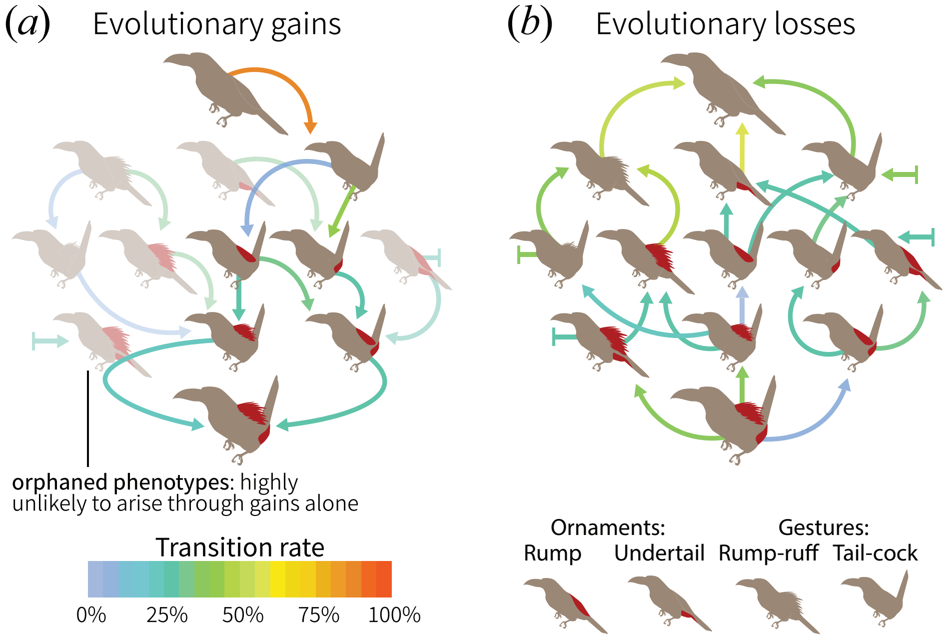


Figure 4: After accounting for evolutionary dependence between signals, we assembled an omnibus transition network describing how tail displays evolved in toucans and barbets. Each silhouette represents one of 16 unique phenotypes, and arrows indicate evolutionary transitions that commonly occurred. When signals are gained (*a*), a relatively large number of transitions (n = 17) are statistically nonsignificant (χ^2 standardized residuals <1.96), which restricts the trajectory by which displays evolve to become more complex. As a result, many phenotypes are orphaned (faded silhouettes; *a*) and unlikely to evolve through gaining signals alone. By contrast, nearly every trait loss (*b*) is common, and orphaned phenotypes can therefore arise via secondary loss.

suggests that secondary signal loss instead may have been an important process underlying their evolution. Altogether, this restricted route by which species gained complex displays suggests that the evolutionary trajectory toward complexity was highly canalized.

By contrast, most trait losses were statistically significant, meaning that few were eliminated in our analysis. This generated a highly diverse transition network characterizing how visual displays evolved from states of high to low phenotypic complexity (fig. 4*b*; table A3). Interestingly, the five transitions eliminated from this network all involved losing undertail ornamentation in species that had a tail-cock gesture and/or rump ornament. Considering that we found undertail ornaments to be retained more frequently in these cases (fig. 3*b*), the rare transitions are likely a result of evolutionary dependency.

Finally, we determined whether the processes of gaining and losing signals differed in the number of available (statistically significant) transitions. As such, we tested whether eliminated transitions were equally distributed between the two processes, and we found that the share of nonsignificant transitions was significantly different ($\chi^2 = 8.38$, $\mathrm{df} = 1$, $P = .004$ with Yates correction; table A5). In other words, there were more gains/fewer losses eliminated than predicted by chance alone ($\chi^2 = 4.98$, $\mathrm{df} = 1$, $P = .023$ for each), which supports the idea that gaining signals—but not losing them—is a canalized process. Therefore, despite the limited number of ways for complex display phenotypes to arise by gaining signals, phenotypic diversity can still arise when ancestrally complex displays subsequently lost different signals.

Discussion

Here we show how gaining and losing traits each contributes to phenotypic diversity in the elaborate tail displays of toucans and barbets. First, SSD predicts whether species use one of four signals (tail-cock gesture) in the display system, suggesting that sexual selection directly favors this gesture's emergence. At the same time, our coevolution analysis suggests that gaining new signals—the process by which displays become more complex—is highly ordered; many signals will not be gained unless others are already present in the ancestral population. By contrast, the same is not true for signal loss, as most signals are lost independently of each other. Each of these findings is robust to subsequent null model analyses, which we conducted to account for potential phylogenetic pseudoreplication due to low sample size. Next, we employed a phenomic approach to reveal that trait gains are limited to a few transitions, whereas nearly every opportunity for signal loss is common. Altogether, these findings corroborate the notion that sexual selection underlies the evolution of complex displays. However, complexity arises following a restricted order of trait gains that is unlikely to generate diverse phenotypes. Instead, novel phenotypes emerge when complex displays lose different signals along an unconstrained sequence of transitions.

Procedural Constraint Guides the Evolution of Complexity

One key insight from our data is that an evolutionary order of operations dictates how signals in the tail display are gained. The first line of evidence is that specific pairs of signals coevolve, as they tend to be gained together. For example, colorful ornaments on the rump are more likely to evolve when tail cocking is present—but at the same time, tail cocking is also gained more frequently in species that have a rump ornament. The same is true for rump ruffing and rump ornaments. By interpreting these relationships in the context of the trait transition network based on the deep evolutionary history of the tail display (fig. 4), we can distill how coevolution shapes the order of evolutionary events. First, only tail cocking is likely to evolve from a plain ancestor, resolving the chicken and egg conundrum of whether gesture or color emerged first. After a lineage gains tail cocking, then plumage ornament on the rump or undertail can evolve. However, simply evolving a tail-cock gesture is not sufficient to unlock all possible signal gains. This is because tail-cocking species are next likely to gain only a color ornament on the rump or undertail but not a rump-ruffing gesture. Coevolution again appears to limit which transitions will take place, as rump ruffing evolves only in lineages that have an ancestral rump ornament. Importantly, this is the only common path by which a species can evolve the most complex display, at least through gains alone. These analyses collectively uncover a canalized route to phenotypic complexity.

Compared with gaining signals, the evolutionary path for signal loss is unconstrained. We found only two cases of potential evolutionary dependence for signal loss, where undertail ornaments are retained more often in species with a tail-cock gesture and/or rump ornament. This means that undertail ornaments are gained independently but evolutionarily stabilized by the other signals. Again, the impact of this coevolution is apparent in the transition matrix, where the only nonsignificant transitions were those in which a species loses its undertail ornament when tail cocking and/or rump ornamentation are present. Beyond this, the rest of the network is marked by near-endless possibilities for signal loss to occur. Numerous mechanisms could act as drivers for this process, including drift, shifts in the context of sexual selection, signal costliness, and functional similarities between signals at the systems level. In the context of this study, we cannot tell which of these factors (if any) plays a dominant role in triggering signal loss.

The frequency of signal loss stands in stark contrast to the limitations on displays gaining complexity, where coevolution between signals restricts the range of likely phenotypic outcomes. In this respect, the process by which visual displays evolve greater complexity appears to be procedurally constrained. Just as developmental and functional constraints limit the range of evolutionary outcomes for a trait (Alexander 1985; Briffa and Sneddon 2007), such procedural constraint restricts a visual display from evolving specific signal combinations if they are out of order.

Sexual Selection and the Order of Signal Evolution

Importantly, our data point to the mechanism by which sexual selection can indirectly act as a catalyst for the evolution of complex displays. SSD predicts which species have gained tail cocking but not the other three signals. Tail cocking is also the first gesture to evolve from a plain ancestor and facilitates future evolution and retention of other visual signals; thus, sexual selection for this gesture may start the process by which elaborate tail displays emerge and diversify. In this case, we suspect that the tail-cock gesture was historically the functional core of barbet tail displays. Subsequent gains of colorful ornaments could augment the original gesture's function in various ways, such as (1) allowing one display to accomplish multiple functions (multiple messages; Johnstone 1996); (2) matching the original signal's function (degeneracy), which can enhance evolvability (Hebets et al. 2016); (3) amplifying a receiver's response to the original signal (Hebets and Uetz 2000); and/or (4) increasing the chance that a receiver can perceive the signal (Ord et al. 2007). Each proposed role for color evolving after gesture is non–mutually exclusive and highlights why complexity is a common feature of displays used in multiple contexts (Langmore 1997; Delaney et al. 2007) or evolving under the influence of competing selection pressures (Endler 1992; Halfwerk et al. 2014).

The idea that tail cocking is a precursor to subsequent ornaments (each of which can accomplish any of the aforementioned functions) is supported by the fact that barbets and toucans use tail cocking specifically in both courtship and competition. The dominant male in a group of *Lybius* barbets, for example, will cock its tail during social displays (Short and Horne 1982; Short et al. 2001), and multiple taxa use tail cocking when confronting rivals or greeting a mate (Short and Horne 1980; Short et al. 2001). Of course, this is not to suggest that color ornaments and additional gestures are not also functional on their own. Some species do not tail cock (e.g., *Pogoniulus* tinkerbirds; *Capito* and *Megalaima* barbets), and the widespread occurrence of secondary signal loss suggests that color ornaments and rump ruffing may take on a signaling function of their own even after tail cocking is lost. If true, a subsequent ornament that overlaps functionally with gesture has the capacity to be a degenerate component in the system or one that overlaps in function with another in such a way that either could be lost without compromising function (Hebets et al. 2016). Because both color and gesture can each be effective signals for mediating courtship and competition, a tail display with both could plausibly lose one or the other and still be functional. For this reason, we suspect that degeneracy could underlie instances of secondary signal loss in the tail display system. Functional origins of additional signals aside, our data certainly undermine a pervading (and typically untested) assumption that behavioral displays evolve as mere accessories to ornamentation (e.g., Hasson 1991; Galván 2008; Scholes et al. 2017) rather than functioning as independent signals.

Why would gesture evolve before ornamentation in a visual display? We hypothesize that gestural evolution preceded color for several (non–mutually exclusive) reasons. First, gesture is an important signaling mechanism in its own right, used widely by species in every major animal group (Girard et al. 2011; Mangiamele et al. 2016; Miles et al. 2017; Hongjamrassilp et al. 2018). In contrast to better-studied signaling mechanisms like coloration and vocalization, which mostly encode information on a single modality (visual and acoustic, respectively), gesture can facilitate visual, acoustic, and/or mechanical (tactile) communication (Elias et al. 2012; Garcia et al. 2012; Miles et al. 2018*a*). The functional flexibility of gesture may underlie its early emergence in a signaling system. Second, gesture's prominence may lie in its capacity to act as an effective indicator of an individual's capacity to perform challenging motor tasks, which can convey valuable information about a signaler's quality (Byers et al. 2010; Barske et al. 2011; Manica et al. 2016).

Alternatively, gesture's early emergence may stem from larger patterns in the pace of behavioral evolution. Considering that behavioral traits are comparatively labile (Blomberg et al. 2003; Miles et al. 2018*b*), tail cocking may have evolved first simply as a matter of probability. If this is the case, gesture may evolve more frequently because the body movements used for signaling can be co-opted from preexisting motor patterns that already functioned in other contexts. For example, several groups of frogs have independently evolved a foot-flagging display, which appears to be a ritualized form of the kicking movements used by fighting males (Preininger et al. 2013; Mangiamele et al. 2016). Interestingly, some foot-flagging species have also evolved bright toe webbing that is displayed while waving, whereas younger lineages have plain webs instead (Preininger et al. 2013). To this end, we also know that color ornaments evolve relatively slowly, likely because gaining new ornaments often necessitates changes in an organism's capacity to metabolize pigments into usable forms and/or deposit those pigments in a new area as a colorant (Badyaev et al. 2015).

A Phenomics Perspective on Complexity and Diversity

Our findings provide new insight into the evolutionary mechanisms that underlie the omnipresent (yet murky) relationship between phenotypic complexity and diversity (Carroll 2001). Because complexity and diversity both exhibit explosive parallel growth throughout evolutionary history, researchers have speculated for decades that complexity is a direct driver of diversity (Carroll 2001; Chambers 2014; Vaesen and Houkes 2017). Tests of this idea have provided mixed results. Most phenotypic studies focus on morphological complexity, which appears to support diversity in some organisms and traits but not others (Adamowicz et al. 2008; Azevedo et al. 2018). Cases corroborating the complexity-diversity link specifically focus on functional traits with an outsized potential to evolve in response to selection (Lenski et al. 2003; Adamowicz et al. 2008). As a result, one unifying feature of this work is an emphasis on versatility and adaptability, because complex traits offer a wider range of opportunities for modification in response to various selection pressures. However, here we leverage a historically oriented phenomic approach to reveal that procedural constraints make it unlikely for diversity to arise as displays become more complex alone. Instead, the procedural asymmetry between gaining and losing traits suggests that evolving complexity plays a facultative role in supporting phenotypic diversity. Whereas the typical model of diversity from complexity emphasizes iterative and often irreversible changes, our data demonstrate how reversing the evolution of complexity can play its own potent role in bringing about new phenotypes.

Finally, the facultative manner in which phenotypic complexity supports diversity also sheds light on an emerging link between complexity, sexual selection, and speciation. Studies suggest that complexity itself is a target of sexual selection (Hebets et al. 2013; Miles and Fuxjager 2018*b*). Moreover, species evolving in a context characterized by high rates of secondary contact tend to have visual displays that are more complex or diverse (Martin et al. 2010; Miles et al. 2017). If complexity indeed amplifies a phenotype's potential to diverge among populations, then highly complex traits should be more likely to support the emergence of critical reproductive barriers (Grant and Grant 2010; Gilman et al. 2018). Typically, the process by which mate choice for courtship traits facilitates diversification is described through the lens of phenotypic novelty (i.e., when a trait is gained; West-Eberhard 1983; Servedio and Bürger 2014). When the novel trait is associated with new patterns of mate choice, it can then drive assortative mating (or trait preference matching) and later speciation, regardless of whether the shift in mate choice is adaptive or arbitrary (Prum 2012; Kopp et al. 2017). If this model applies to barbet visual displays, however, then courtship novelty may functionally arise when a population loses an ancestral signal instead. In turn, phenotypic diversification through secondary loss may also provide the impetus for future speciation if assortative mating is strong, even without the emergence of novel traits altogether.

Conclusions

In summary, we used a phenomic approach to disentangle the evolutionary processes that generate critical elements of phenotypic pattern—complexity and diversity—in the elaborate visual displays of toucans and barbets. Procedural constraint on the evolution of complexity restricts the potential for many display phenotypes to arise. Instead, phenotypic diversity is supported by secondary signal loss from lineages with complex displays. The entire process appears to be catalyzed by sexual selection for communication with body movement (gesture), highlighting the potential for behavior to influence how other traits and even species evolve.

Acknowledgments

This work was supported by a National Science Foundation grant (IOS-1655730) to M.J.F. We thank two anonymous peer reviewers and associate editor Rebecca J. Safran for providing valuable feedback on the manuscript.

Statement of authorship: M.C.M. and M.J.F. conceived of the study, analyzed and interpreted the data, and wrote the manuscript; and M.C.M. collected the data.

Literature Cited

Adamowicz, S. J., A. Purvis, and M. A. Wills. 2008. Increasing morphological complexity in multiple parallel lineages of the Crustacea. Proceedings of the National Academy of Sciences of the USA 105:4786–4791.

Alexander, R. M. N. 1985. The ideal and the feasible: physical constraints on evolution. Biological Journal of the Linnean Society 26:345–358.

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, et al. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences of the USA 106:13410–13414.

Azevedo, G. H. F., C. E. Griswold, and A. J. Santos. 2018. To complicate or to simplify? phylogenetic tests of complexity trends and genital evolution in ground spiders (Araneae: Dionycha: Gnaphosidae). Zoological Journal of the Linnean Society 84:673–694.

Badyaev, A. V., E. S. Morrison, V. Belloni, and M. J. Sanderson. 2015. Tradeoff between robustness and elaboration in carotenoid networks produces cycles of avian color diversification. Biology Direct 10:45.

Barske, J., B. A. Schlinger, M. Wikelski, and L. Fusani. 2011. Female choice for male motor skills. Proceedings of the Royal Society B 278:3523–3528.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple comparisons. Journal of the Royal Statistical Society 57:289–300.

Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745.

Briffa, M., and L. U. Sneddon. 2007. Physiological constraints on contest behaviour. Functional Ecology 21:627–637.

Byers, J., E. Hebets, and J. Podos. 2010. Female mate choice based upon male motor performance. Animal Behaviour 79:771–778.

Carroll, S. B. 2001. Chance and necessity: the evolution of morphological complexity and diversity. Nature 409:1102–1109.

Chambers, G. K. 2014. Understanding complexity: are we making progress? Biology and Philosophy 30:747–756.

Civelek, M., and A. J. Lusis. 2014. Systems genetics approaches to understand complex traits. Nature Reviews Genetics 15:34–48.

Cooney, C. R., J. A. Tobias, J. T. Weir, C. A. Botero, and N. Seddon. 2017. Sexual selection, speciation and constraints on geographical range overlap in birds. Ecology Letters 20:863–871.

Dale, J., P. O. Dunn, J. Figuerola, T. Lislevand, T. Székely, and L. A. Whittingham. 2007. Sexual selection explains Rensch's rule of allometry for sexual size dimorphism. Proceedings of the Royal Society B 274:2971–2979.

Delaney, K. J., J. A. Roberts, and G. W. Uetz. 2007. Male signaling behavior and sexual selection in a wolf spider (Araneae: Lycosidae): a test for dual functions. Behavioral Ecology and Sociobiology 62:67–75.

del Hoyo, J., A. Elliott, and J. Sargatal. 1999. Handbook of the birds of the world. Vol. 5. Lynx Edicions, Barcelona.

Den Tex, R. J., and J. A. Leonard. 2013. A molecular phylogeny of Asian barbets: speciation and extinction in the tropics. Molecular Phylogenetics and Evolution 68:1–13.

Dunn, P. O., L. A. Whittingham, and T. E. Pitcher. 2001. Mating systems, sperm competition, and the evolution of sexual dimorphism in birds. Evolution 55:161–175.

Egge, A. R., Y. Brandt, and J. G. Swallow. 2011. Sequential analysis of aggressive interactions in the stalk-eyed fly *Teleopsis dalmanni*. Behavioral Ecology and Sociobiology 65:369–379.

Elias, D. O., W. P. Maddison, C. Peckmezian, M. B. Girard, and A. C. Mason. 2012. Orchestrating the score: complex multimodal courtship in the *Habronattus coecatus* group of *Habronattus* jumping spiders (Araneae: Salticidae). Biological Journal of the Linnean Society 105:522–547.

Endler, J. A. 1992. Signals, signal conditions, and the direction of evolution. American Naturalist 139(suppl.):S125–S153.

Galván, I. 2008. The importance of white on black: unmelanized plumage proportion predicts display complexity in birds. Behavioral Ecology and Sociobiology 63:303–311.

Garcia, M., I. Charrier, D. Rendall, and A. N. Iwaniuk. 2012. Temporal and spectral analyses reveal individual variation in a nonvocal acoustic display: the drumming display of the ruffed grouse (*Bonasa umbellus*, L.). Ethology 118:292–301.

Gilman, R. T., K. Fowler-Finn, and E. A. Hebets. 2018. A probable case of incipient speciation in *Schizocosa* wolf spiders driven by allochrony, habitat use, and female mate choice. American Naturalist 192:332–346.

Girard, M. B., D. O. Elias, and M. M. Kasumovic. 2015. Female preference for multi-modal courtship: multiple signals are important for male mating success in peacock spiders. Proceedings of the Royal Society B 282:20152222.

Girard, M. B., M. M. Kasumovic, D. O. Elias, R. Hoy, and A. Mason. 2011. Multi-modal courtship in the peacock spider, *Maratus volans* (O.P.-Cambridge, 1874). PLoS ONE 6:e25390.

Grant, B. R., and P. R. Grant. 2010. Songs of Darwin's finches diverge when a new species enters the community. Proceedings of the National Academy of Sciences of the USA 107:20156–20163.

Green, P. A., and S. N. Patek. 2018. Mutual assessment during ritualized fighting in mantis shrimp (Stomatopoda). Proceedings of the Royal Society B 285:20172542.

Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski, et al. 2008. A phylogenomic study of birds reveals their evolutionary history. Science 320:1763–1768.

Halfwerk, W., P. L. Jones, R. C. Taylor, M. J. Ryan, and R. A. Page. 2014. Risky ripples allow bats and frogs to eavesdrop on a multisensory sexual display. Science 343:413–416.

Hasson, O. 1991. Sexual displays as amplifiers: practical examples with an emphasis on feather decorations. Behavioral Ecology 2:189–197.

Hebets, E. A., A. B. Barron, C. N. Balakrishnan, M. E. Hauber, P. H. Mason, and K. L. Hoke. 2016. A systems approach to animal communication. Proceedings of the Royal Society B 283:20152889.

Hebets, E. A., and D. R. Papaj. 2005. Complex signal function: developing a framework of testable hypotheses. Behavioral Ecology and Sociobiology 57:197–214.

Hebets, E. A., and G. W. Uetz. 2000. Leg ornamentation and the efficacy of courtship display in four species of wolf spider (Araneae: Lycosidae). Behavioral Ecology and Sociobiology 47:280–286.

Hebets, E. A., C. J. Vink, L. Sullivan-Beckers, and M. F. Rosenthal. 2013. The dominance of seismic signaling and selection for signal complexity in *Schizocosa* multimodal courtship displays. Behavioral Ecology and Sociobiology 67:1483–1498.

Ho, L. S. T., and C. Ané. 2014. Phylolm: phylogenetic linear regression. Version 2.1. https://cran.r-project.org/web/packages/phylolm/index.html.

Hongjamrassilp, W., A. P. Summers, and P. A. Hastings. 2018. Heterochrony in fringeheads (*Neoclinus*) and amplification of an extraordinary aggressive display in the sarcastic fringehead (Teleostei: Blenniiformes). Journal of Morphology 279:626–635.

Ives, A. R., and T. Garland. 2010. Phylogenetic logistic regression for binary dependent variables. Systematic Biology 59:9–26.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. Nature 491:444–448.

Johnstone, R. A. 1996. Multiple displays in animal communication: 'backup signals' and 'multiple messages.' Philosophical Transactions of the Royal Society B 351:329–338.

Klomp, D. A., T. J. Ord, I. Das, A. Diesmos, N. Ahmad, and D. Stuart-Fox. 2016. Ornament size and colour as alternative strategies for effective communication in gliding lizards. Journal of Evolutionary Biology 29:1689–1700.

Kopp, M., M. R. Servedio, T. C. Mendelson, R. J. Safran, R. L. Rodríguez, M. E. Hauber, E. C. Scordato, et al. 2017. Mechanisms of assortative mating in speciation with gene flow: connecting theory and empirical research. American Naturalist 191:1–20.

Langmore, N. E. 1997. Song switching in monandrous and polyandrous dunnocks, *Prunella modularis*. Animal Behaviour 53:757–766.

Lenski, R. E., C. Ofria, R. T. Pennock, and C. Adami. 2003. The evolutionary origin of complex features. Nature 423:139–144.

Lutz, H. L., J. D. Weckstein, J. S. L. Patané, J. M. Bates, and A. Aleixo. 2013. Biogeography and spatio-temporal diversification of Selenidera and Andigena toucans (Aves: Ramphastidae). Molecular Phylogenetics and Evolution 69:873–883.

Maddison, W. P., and R. G. FitzJohn. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Systematic Biology 64:127–136.

Maia, R., D. R. Rubenstein, and M. D. Shawkey. 2013. Key ornamental innovations facilitate diversification in an avian radiation. Proceedings of the National Academy of Sciences of the USA 110:10687–10692.

———. 2016. Selection, constraint, and the evolution of coloration in African starlings. Evolution 70:1064–1079.

Mangiamele, L. A., M. J. Fuxjager, E. R. Schuppe, R. S. Taylor, W. Hödl, and D. Preininger. 2016. Increased androgenic sensitivity in the hind limb muscular system marks the evolution of a derived gestural display. Proceedings of the National Academy of Sciences of the USA 113:5664–5669.

Manica, L. T., R. H. Macedo, J. A. Graves, and J. Podos. 2016. Vigor and skill in the acrobatic mating displays of a Neotropical songbird. Behavioral Ecology 28:164–173.

Martin, P. R., R. Montgomerie, and S. C. Lougheed. 2010. Rapid sympatry explains greater color pattern divergence in high latitude birds. Evolution 64:336–347.

Miles, M. C., S. Cheng, and M. J. Fuxjager. 2017. Biogeography predicts macro-evolutionary patterning of gestural display complexity in a passerine family. Evolution 71:1406–1416.

Miles, M. C., and M. J. Fuxjager. 2018*a*. Animal choreography of song and dance: a case study in the Montezuma oropendola, *Psarocolius montezuma*. Animal Behaviour 140:99–107.

———. 2018*b*. Synergistic selection regimens drive the evolution of display complexity in birds of paradise. Journal of Animal Ecology 87:1149–1159.

Miles, M. C., E. R. Schuppe, R. M. Ligon, and M. J. Fuxjager. 2018*a*. Macroevolutionary patterning of woodpecker drums reveals how sexual selection elaborates signals under constraint. Proceedings of the Royal Society B 285:20172628.

Miles, M. C., M. N. Vitousek, J. F. Husak, M. A. Johnson, L. B. Martin, C. C. Taff, C. Zimmer, et al. 2018*b*. Standing variation and the capacity for change: are endocrine phenotypes more variable than other traits? Integrative and Comparative Biology 58:751–762.

Ödeen, A., and O. Håstad. 2013. The phylogenetic distribution of ultraviolet sensitivity in birds. BMC Evolutionary Biology 13:36.

Ord, T. J., and E. P. Martins. 2006. Tracing the origins of signal diversity in anole lizards: phylogenetic approaches to inferring the evolution of complex behaviour. Animal Behaviour 71:1411–1429.

Ord, T. J., R. A. Peters, B. Clucas, and J. A. Stamps. 2007. Lizards speed up visual displays in noisy motion habitats. Proceedings of the Royal Society B 274:1057–1062.

Pagel, M., and A. Meade. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. American Naturalist 167:808–825.

———. 2007. BayesTraits. Version 3. http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.1/BayesTraitsV3.0.1.html.

Pagel, M., A. Meade, and D. Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. Systematic Biology 53:673–684.

Pärt, T., and A. Qvarnström. 1997. Badge size in collared flycatchers predicts outcome of male competition over territories. Animal Behaviour 54:893–899.

Payne, R. B. 1984. Sexual selection, lek and arena behavior, and sexual size dimorphism in birds. Ornithological Monographs 33. American Ornithologists' Union, Washington, DC.

Plummer, M., N. Best, K. Cowles, K. Vines, D. Sarkar, D. Bates, R. Almond, and A. Magnusson. 1999. coda: output analysis and diagnostics for MCMC. R package.

Preininger, D., M. J. Stiegler, K. V. Gururaja, S. P. Vijayakumar, V. R. Torsekar, M. Sztatecsny, and W. Hödl. 2013. Getting a kick out of it: multimodal signalling during male-male encounters in the foot-flagging frog *Micrixalus* aff. *saxicola* from the Western Ghats of India. Current Science 105:1735–1739.

Price, T. D. 1984. The evolution of sexual size dimorphism in Darwin's finches. American Naturalist 123:500–518.

Prum, R. O. 1990. Phylogenetic analysis of the evolution of display behavior in the Neotropical manakins (Aves: Pipridae). Ethology 84:202–231.

———. 2010. The Lande-Kirkpatrick mechanism is the null model of evolution by intersexual selection: implications for meaning, honesty, and design in intersexual signals. Evolution 64:3085–3100.

———. 2012. Aesthetic evolution by mate choice: Darwin's really dangerous idea. Philosophical Transactions of the Royal Society B 367:2253–2265.

Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152–1164.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Schliep, K. P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593.

Scholes, E., J. M. Gillis, and T. G. Laman. 2017. Visual and acoustic components of courtship in the bird-of-paradise genus Astrapia (Aves: Paradisaeidae). PeerJ 5:e3987.

Schuppe, E. R., and M. J. Fuxjager. 2018. High-speed displays encoding motor skill trigger elevated territorial aggression in downy woodpeckers. Functional Ecology 32:450–460.

Servedio, M. R. 2016. Geography, assortative mating, and the effects of sexual selection on speciation with gene flow. Evolutionary Applications 9:91–102.

Servedio, M. R., and R. Bürger. 2014. The counterintuitive role of sexual selection in species maintenance and speciation. Proceedings of the National Academy of Sciences of the USA 111:8113–8118.

Short, L., J. Horne, and J. F. M. Horne. 2001. Toucans, barbets, and honeyguides: Ramphastidae, Capitonidae and Indicatoridae. Oxford University Press, Oxford.

Short, L. L., and J. F. M. Horne. 1980. Vocal and other behaviour of the green barbet in Kenya. Ostrich 51:219–229.

———. 1982. Vocal and other behaviour of Kenyan black-collared barbets *Lybius torquatus*. Ibis 124:27–43.

Soma, M., and L. Z. Garamszegi. 2015. Evolution of courtship display in Estrildid finches: dance in relation to female song and plumage ornamentation. Frontiers in Ecology and Evolution 3:4.

Székely, T., J. D. Reynolds, and J. Figuerola. 2000. Sexual size dimorphism in shorebirds, gulls, and alcids: the influence of sexual and natural selection. Evolution 54:1404–1413.

Vaesen, K., and W. Houkes. 2017. Complexity and technological evolution: What everybody knows? Biology and Philosophy 32:1245–1268.

Webster, M. S. 1992. Sexual dimorphism, mating system and body size in New World blackbirds (Icterinae). Evolution 46:1621–1641.

West-Eberhard, M. J. 1983. Sexual selection, social competition, and speciation. Quarterly Review of Biology 58:155–183.

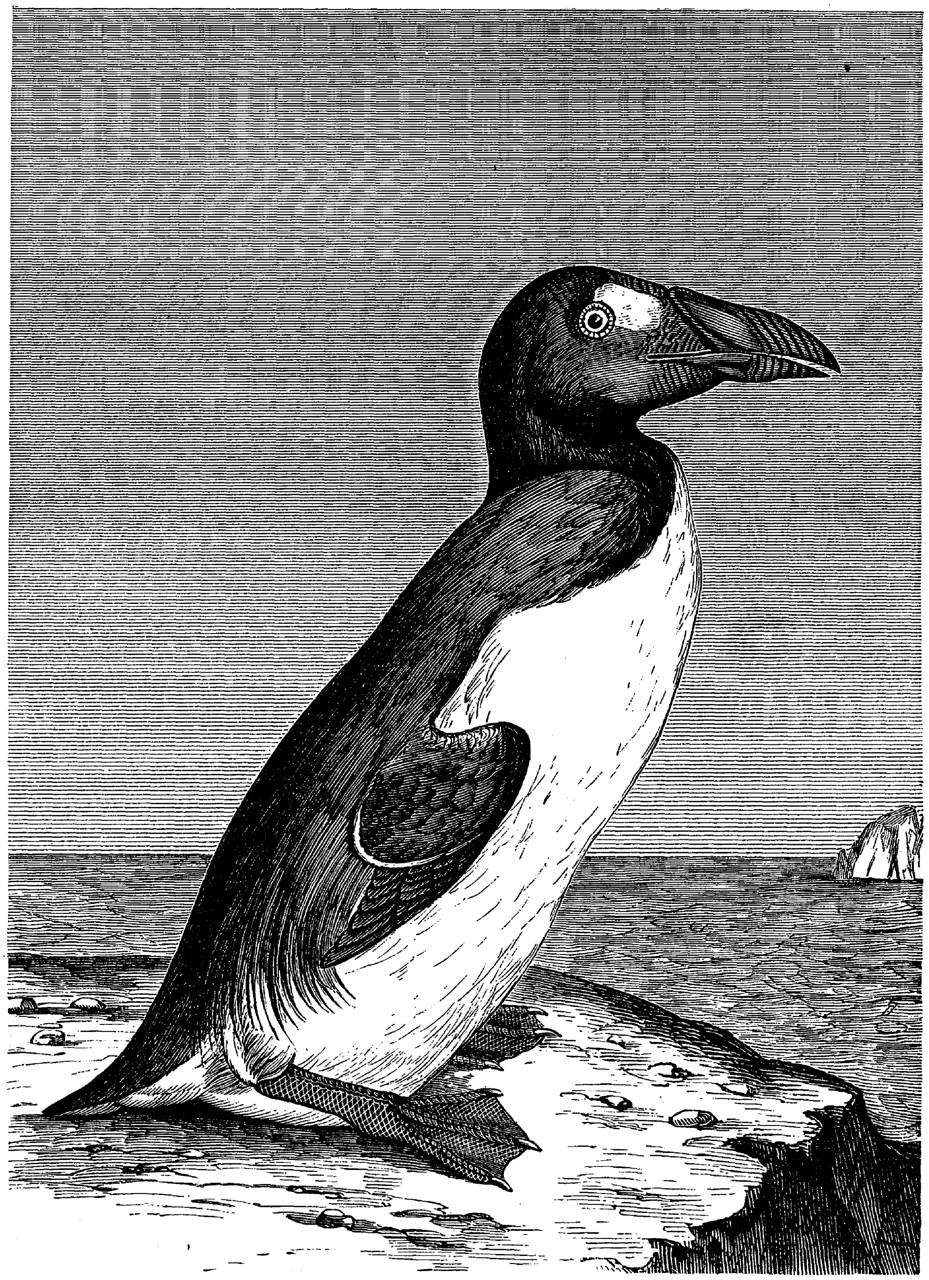
———. 2003. Developmental plasticity and evolution. Oxford University Press, Oxford.

Wiens, J. J. 2017. What explains patterns of biodiversity across the Tree of Life? new research is revealing the causes of the dramatic variation in species numbers across branches of the Tree of Life. BioEssays 39:1600128.

Xie, W., P. O. Lewis, Y. Fan, L. Kuo, and M. H. Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Systematic Biology 60:150–160.

Yasukawa, K. 1978. Aggressive tendencies and levels of a graded display: factor analysis of response to song playback in the red-winged blackbird (*Agelaius phoeniceus*). Behavioral Biology 23:446–459.

Associate Editor: Rebecca J. Safran
Editor: Daniel I. Bolnick



“The Great Auk or Gare-fowl . . . was about the size of a goose, with a large head, a curved, grooved and laterally flattened bill; wings rudimental, adapted to swimming only, approaching in this respect the penguins of the southern hemisphere.” From “The Great Auk” by James Orton (*The American Naturalist*, 1869, 3:539–542).