

Research



Cite this article: Araya-Salas M, Smith-Vidaurre G, Mennill DJ, González-Gómez PL, Cahill J, Wright TF. 2019 Social group signatures in hummingbird displays provide evidence of co-occurrence of vocal and visual learning. *Proc. R. Soc. B* **286**: 20190666. <http://dx.doi.org/10.1098/rspb.2019.0666>

Received: 21 March 2019

Accepted: 2 May 2019

Subject Category:

Behaviour

Subject Areas:

behaviour, evolution, neuroscience

Keywords:

visual displays, geographical variation, vocal learning, hummingbirds, lek mating system, motor theory for vocal learning

Author for correspondence:

Marcelo Araya-Salas

e-mail: marceloa27@gmail.com

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.4498928>.

Social group signatures in hummingbird displays provide evidence of co-occurrence of vocal and visual learning

Marcelo Araya-Salas^{1,2,3}, Grace Smith-Vidaurre¹, Daniel J. Mennill⁴, Paulina L. González-Gómez^{5,6}, James Cahill⁷ and Timothy F. Wright¹

¹Department of Biology, New Mexico State University, Las Cruces, NM, USA

²Laboratory of Ornithology, Cornell University, Ithaca, NY, USA

³Escuela de Biología, Universidad de Costa Rica, San Pedro, San José, Costa Rica

⁴Department of Biological Sciences, University of Windsor, Windsor, Ontario, Canada

⁵Department of Neurobiology, Physiology and Behavior, University of California Davis, Davis, CA, USA

⁶Universidad Autónoma de Chile, Santiago, Chile

⁷Laboratory of the Neurogenetics of Language, Rockefeller University, New York, NY, USA

MA-S, 0000-0003-3594-619X

Vocal learning, in which animals modify their vocalizations based on social experience, has evolved in several lineages of mammals and birds, including humans. Despite much attention, the question of how this key cognitive trait has evolved remains unanswered. The motor theory for the origin of vocal learning posits that neural centres specialized for vocal learning arose from adjacent areas in the brain devoted to general motor learning. One prediction of this hypothesis is that visual displays that rely on complex motor patterns may also be learned in taxa with vocal learning. While learning of both spoken and gestural languages is well documented in humans, the occurrence of learned visual displays has rarely been examined in non-human animals. We tested for geographical variation consistent with learning of visual displays in long-billed hermits (*Phaethornis longirostris*), a lek-mating hummingbird that, like humans, has both learned vocalizations and elaborate visual displays. We found lek-level signatures in both vocal parameters and visual display features, including display element proportions, sequence syntax and fine-scale parameters of elements. This variation was not associated with genetic differentiation between leks. In the absence of genetic differences, geographical variation in vocal signals at small scales is most parsimoniously attributed to learning, suggesting a significant role of social learning in visual display ontogeny. The co-occurrence of learning in vocal and visual displays would be consistent with a parallel evolution of these two signal modalities in this species.

1. Introduction

Vocal learning is an important cognitive trait that has evolved in a select few groups of vertebrates. It is well documented in three mammalian taxa (cetaceans, bats and humans), but is absent in the sister taxa of each of these clades, including the non-human primates [1] (but see [2]). A similar pattern occurs in birds, where songbirds, parrots and hummingbirds learn their vocalizations, but their closest relatives do not [3]. There has been extensive research on animal vocal learning in the last four decades, and a number of hypotheses have been proposed regarding the benefits conferred by learning on individuals with this trait [4–6]. In contrast there are relatively few hypotheses for how this trait evolved in the first place [4–6]. One exception is the motor theory for the origins of vocal learning. This hypothesis suggests that the neural substrates for vocal learning evolved from pre-existing circuits for general motor learning [7]. This hypothesis arose from the observations that (a) each of the interconnected neural centres used for vocal learning is embedded within neural regions used

for general motor movement, and (b) these specialized regions are connected in cortico-striatal basal ganglia loops similar to those used for general motor learning in a range of taxa [3,7]. One interesting prediction arising from this hypothesis is that species with vocal learning may also use learning to acquire other communication signals with complex motor patterns, such as dynamic visual signals (e.g. co-occurrence of vocal and visual learning). Remarkably, few studies have examined the role of learning in the development of dynamic visual signals in non-human animals that exhibit vocal learning [8].

The presence of vocal learning is often diagnosed by mapping geographical variation in signals [6,9]. Vocal variation at a small geographical scale that shows both sharp differences among localities and consistency within localities is indicative of social learning of local vocal traditions [10]. This type of vocal variation (often termed 'dialects') is found in many vocal learning taxa (e.g. whales [11], bats [12], parrots [13] and hummingbirds [14]). Although their absence does not necessarily imply a lack of learning, dialects in genetically homogeneous populations are most parsimoniously explained by the learning of motor patterns that produce these acoustic signals. By extension, patterns of geographical variation in visual signals, in the absence of genetic differentiation, could provide insight into the significance of learning in the development of the associated motor patterns.

Hummingbirds are an intriguing and little-explored group of animals in which to investigate these questions. They are known for their conspicuous visual displays used for mate attraction and territorial defence in males [15,16]. Songs are used for similar purposes and have been shown to be acquired through learning in several species [14,17–21]. One hummingbird species in which both visual displays and remarkable vocal learning skills have been characterized is the long-billed hermit (*Phaethornis longirostris*). This species forms leks of 5–20 males that sing and display on small, closely packed display territories. Females visit leks to select a mate from among the displaying males, but provide all subsequent parental care themselves. Leks constitute discrete social units, separated by approximately 1 km from each other within a matrix of continuous forest. Male long-billed hermits have a vocal repertoire consisting of a single song type sung by resident males from established perches within their lek territory, while visual displays are composed of sequences of stereotyped movements (hereafter display elements) and specific vocalizations. Displays occur mostly at specific perches, and are only performed when intruders approach singing residents and both males engage in protracted interactive displays [22]. Previous work has documented the basic form of nine visual display elements and the existence of song neighbourhoods within leks (sub-groups with different song types [22]), as well as an open-ended song-learning programme [14].

In this study we assessed whether social learning is involved in the acquisition of the visual motor displays in the long-billed hermit. We specifically asked: do visual displays exhibit a dialect pattern of geographical variation consistent with learning? If so, do vocal dialects and visual dialects occur at the same scales? To rule out the prime alternative hypothesis for a mechanism promoting geographical variation, we also examined genetic relatedness among individuals. Finally, we asked what attributes of visual displays seem to be acquired through learning. We

evaluated these questions by comparing the structural similarity of vocal and visual signals at geographical scales representative of four different social group levels: individual, song neighbourhood within a lek (i.e. subgroup of males within a lek that share a particular song type), lek and site. We applied parallel statistical analyses to vocal and visual signal parameters to assess whether these parameters were more similar within social groups versus between social groups at each social level. This concurrent analysis provides a strong test of our prediction for the co-occurrence of vocal and visual learning in this species.

2. Methods

Fieldwork was conducted during two breeding seasons (2013–2014) at six leks from three sites in northeastern Costa Rica (figure 1): La Selva (four leks: SUR, CCL, SJA and LOC), Hitoy-Cerere (one lek: HC1) and Tirimbina (one lek: TR2). Male long-billed hermits from the four leks at La Selva were mist-netted and marked with numbered bands. Tags with unique three-colour combinations were attached to the back and chest of each bird with nontoxic eyelash glue (LashGrip-Ardell) [18]. Tags typically stayed on the bird for two weeks before they fell off. At the other sites we determined identities of lekking males through behavioural observations and territory mapping. Perches of singing males were flagged and mapped at all sites.

We recorded 10 938 songs during 94 song recording sessions from 60 males across six leks (mean 10 males per lek, range 8–15) with an average of 4.3 (range = 1–9) recording sessions per male. For each male we selected the five songs with the highest signal-to-noise ratio among all its recordings (mean: 4.9 song per male, range: 2–5). Based on visual inspection of song spectrograms, we found two leks with a single song type and four with two song types (figure 1). No song type was observed at more than one lek. We recorded songs of lekking males on a Marantz PMD 660 and a Sennheiser ME62/K6 microphone (20–20 000 Hz frequency response) on a parabolic antenna (53 cm diameter) in .WAV format with a sampling frequency of 44 100 Hz and 32 bits sampling depth.

Video recordings were made on five Fujifilm HS30 and one Fujifilm HS10 cameras mounted on tripods placed at approximately 1–2 m in front of a flagged perch and, whenever possible, about the same height as the perch (some cameras had to be angled slightly upwards to capture displays at higher perches). All videos were recorded for a maximum capture period of 2 h at 29.97 frames per second. After a preliminary analysis of a subset of the videos, we estimated that at minimum, 36 video hours per individual were needed to capture at least five different display elements (electronic supplementary material, figure S1). Therefore, we attempted to record each male for 6 h a day over 6 days. Each day we recorded 4 h in the morning (07.00–11.00) and 2 h in the afternoon (14.00–16.00) during the peak activity periods [21]. We aimed to video record six territorial males per lek. Birds at Tirimbina were recorded in both 2013 and 2014; all other sites were only recorded in 2013.

(a) Acoustic and visual display analyses

Songs were compared using spectrographic cross-correlation (200-samples window length; 90% window overlap; Hanning window function; Pearson correlation). A mean pairwise cross-correlation was calculated for each individual, song neighbourhood or lek, when individuals/song neighbourhoods/leks were the sampling units, respectively. Song types were visually classified based on their spectrographic structure. Visual

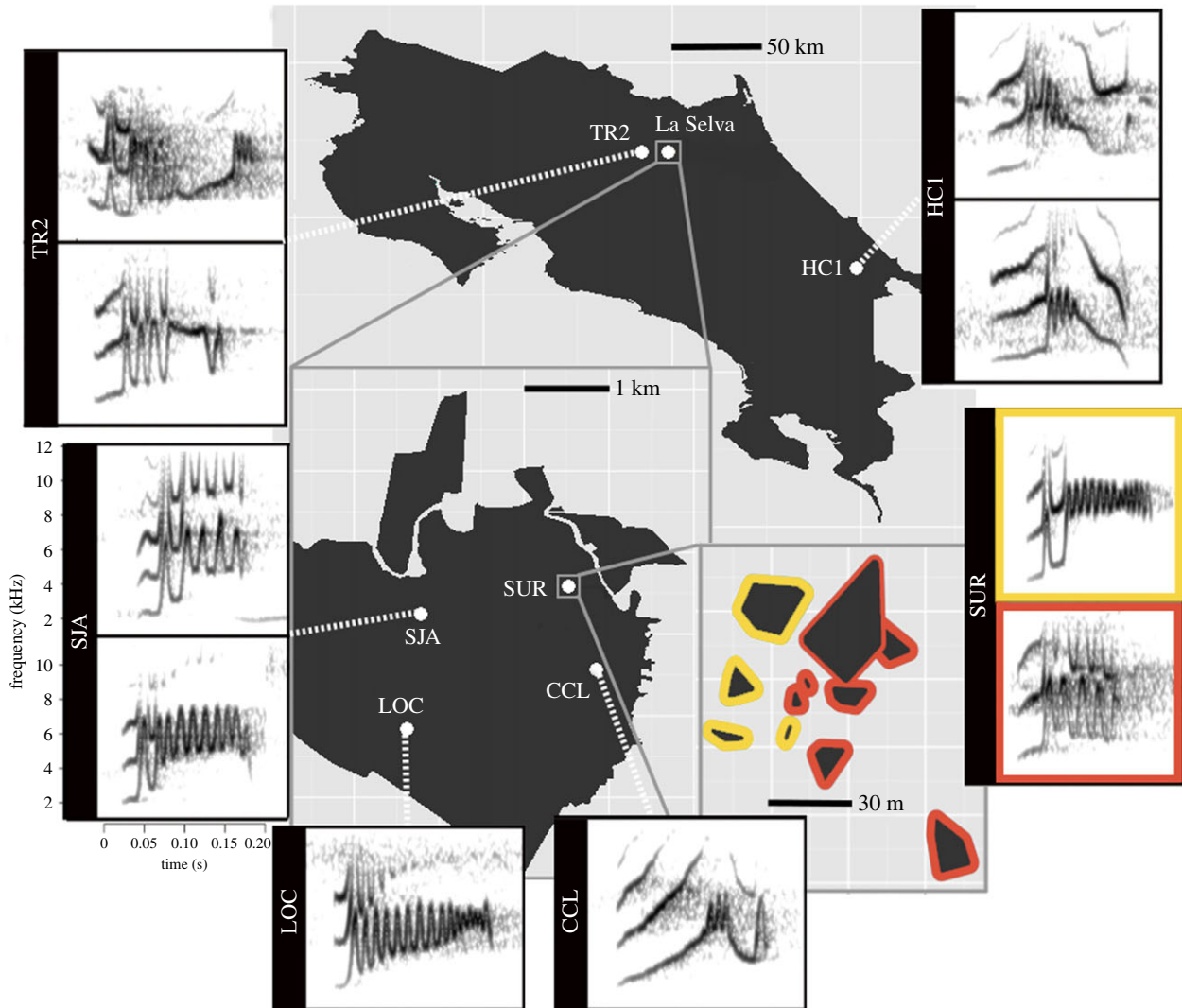


Figure 1. Location of three study sites in northeastern Costa Rica. Spectrograms of the 10 observed song types grouped by lek are also shown. Maps of leks at La Selva Biological Station are shown in greater detail in the lower left map. The two song neighbourhoods at lek SUR are shown in lower right map (similar song neighbourhoods were found at the other leks with 2 song types—SJA, TR2, and HC1—but are not shown); polygons represent lekking male territories and the coloured borders delineate the song neighbourhoods corresponding to the coloured borders of the spectrograms at the far right.

classification has been shown to be highly repeatable in this species [14]. Acoustic analyses were carried out using the R packages *tuneR* [23] (importing sound files into R), *seewave* [24] (creating spectrograms) and *warbleR* [25] (running cross-correlation).

We scanned all videos to identify display elements and compiled a descriptive ethogram (table 1). We also annotated the element sequence for each display. Both residents (identified by tags and consistent territorial behaviour) and intruders (identified as individuals other than the resident) were involved in some display elements (figure 2). As variation between individuals was used for evaluating lek-level signatures in visual displays, it was important to assess individual consistency in display characteristics for both intruders and residents (i.e. whether most of the variation could be attributed to one of these displayers). As we detected a resident signature, but not an intruder signature (see Results), for most visual display parameters, display sequences were assigned to resident individuals.

We measured four components of visual displays: repertoire composition (presence–absence of display elements), display element proportion (relative abundance of elements), display element transitions (probability of transition between elements) and display element fine-scale parameters (duration of specific elements and complete displays as well as counts and repetition

rate of specific movements repeated within elements). For the fine-scale parameters, collinear variables and variables with small sample size were excluded, leaving three parameters: display duration, float duration and repetition of bill openings in side by side displays.

(b) Genetic analysis

We examined the genetic relatedness among individuals and leks using single nucleotide polymorphism (SNP) yielded through genotyping-by-sequencing. High-throughput reduced-representation methods, including genotyping-by-sequencing, have been recommended for evaluating population-level genetic structure in general [26,27] and in birds in particular [28]. Genetic analysis was conducted for 20 of the 30 males videotaped at La Selva leks (mean \pm s.d. pairwise shared SNPs 1002 ± 990). Individuals were captured at the four La Selva leks using mist nets and whole blood samples were preserved on Whatman FTA Elute cards. We extracted genomic DNA using standard Whatman protocols and sexed individuals using PCR with the P0-P2-P8 primer combination [29]. Library preparation for genotyping-by-sequencing with the targeted capture approach was performed by the University of Wisconsin's Biotechnology Center DNA Sequencing Facility using the ApeKI kit (New

Table 1. Ethogram describing the 12 display elements of resident and intruder long-billed hermit males at lek perches. Some of these displays have been previously described elsewhere [22].

| visual display element | description |
|------------------------|---|
| tail fan | Tail lifted and rectrices spread, fan-like, separating the two white central rectrices. Tail lifted closer to 90 degrees with respect to body as conspecific draws closer, and sometimes wagged slowly up and down. |
| perch exchange | Complex and ritualized perch-switching duet. Intruder approaches perched resident, rapidly flies over resident and twists in air to land behind resident, similar to copulation posture. Resident immediately flies off perch as intruder perches and can initiate another perch exchange. Can be repeated various times and does not always follow an orderly intruder–resident–intruder sequence (figure 3c). |
| squeak | Sharp vocalization that is usually given during perch exchange, as one bird descends upon another in a pseudocopulatory position. |
| perch displacement | Perched bird is displaced without ritualized flyover associated with perch exchange. |
| float | Displaced conspecific flies slowly back and forth in front of perched bird, often with bill open. Perched bird follows floater's movements back and forth, sometimes with bill open and in an aggressive posture (figure 3a). |
| side by side | Intruder and resident perch side by side, often close together on the same perch. Resident usually puffs throat feathers, opens bill and lifts it vertically. Both resident and intruder can rapidly flutter wings and tail. Intruder will sometimes open bill as well. Both birds can sometimes hold wings and tails still, with feathers flattened in typical aggressive manner (figure 3b). |
| bill poke | Performed during side by side or float. Intruder pokes bill tip into resident's throat or abdomen while resident puffs out throat feathers, holding bill open and vertical. |
| bill pop | Bird performing bill pop hovers in front of perched bird, then simultaneously flips head upward while opening bill wide to produce a 'pop' sound. Bill pop can also occur while both birds are airborne during hover up (figure 3d). |
| gape display | Performer hovers in front of perched bird, then flips head upward while opening bill wide and displaying gape while producing a soft, buzzing 'shrrr . . . shrrr' vocalization. |
| chase | Resident or intruder will chase the opponent after an interactive display. |
| hover-up | Resident and intruder face each other while hovering in air, often with rectrices spread. Sometimes precedes a chase, although birds often move off-frame when holding this position. |
| departure | Intruder leaves perch without being chased by resident. Resident can stare in direction of intruder's departure and fly off perch shortly after. |

England Biolabs, Ipswich, MA) [30] and samples sequenced on an Illumina HiSeq 2500 with 100 bp length single-end reads.

We mapped reads to the Anna's hummingbird reference genome (assembly: *Canna_diploid_1.0*) [31] using *BWA MEM* v. 0.7.5 [32], and excluded reads with map quality scores less than 30 and removed duplicates using *SAMtools* version 0.1.19 [33]. Then we randomly sampled a single high-quality base call ($\text{baseQ} > 30$) at each position in the reference genome. At sites with multiple mapped reads, a single base call was selected to represent each individual. For each pair of individuals, we calculated the number of pairwise differences at biallelic sites across the entire genome using *trialn-report* (Paleogenomics/Chrom-Compare).

(c) Statistical analysis

We evaluated visual and vocal display distinctiveness in the four display components at the individual, song-neighbourhood, lek, and site level. We applied Mantel test correlations (100 000 iterations) to assess parameter similarity at each social group level. We chose this method because it directly tests the prediction that learning generates higher similarity among members of the same social groups, whereas alternative approaches (e.g. GLMMs) typically test whether at least one group differs from others. We used pairwise binary matrices to represent group membership, or individual membership when assessing individual signatures. We

used 0 to denote the same group/individual and 1 to denote a different group/individual [13,34]. Group membership matrices were used as predictors in Mantel correlations against dissimilarity matrices of acoustic or visual parameters. Statistical significance indicates that parameters are more similar between members of the same groups than to members of other groups (e.g. displays within individuals, individuals within leks). A genetic distance matrix was also used as a predictor for acoustic and visual display parameters and as a response against a lek membership binary matrix to directly assess genetic structure at the lek level (using only La Selva leks). The cross-correlation matrix was converted to a distance matrix by subtracting it from 1. We used *Morisita overlap index* [35] to calculate dissimilarity in repertoire proportions, *Jaccard dissimilarity index* [36] for dissimilarity of presence–absence of visual elements, and *Euclidean distance* for dissimilarity of log-transformed continuous variables (fine-scale parameters of visual signals and geographical distances). All fine-scale parameters were summarized in a single distance matrix calculated as the Euclidean distance between samples in the multidimensional space defined by *z*-transformed measures. Pairwise sequence dissimilarity was estimated using optimal matching analysis with substitution costs derived from the observed transition rates [37]. The *r* statistic of the Mantel test was used as a similarity measure between element transition matrices. Song recordings or visual displays were the sampling unit for analyses at the individual level. Individual

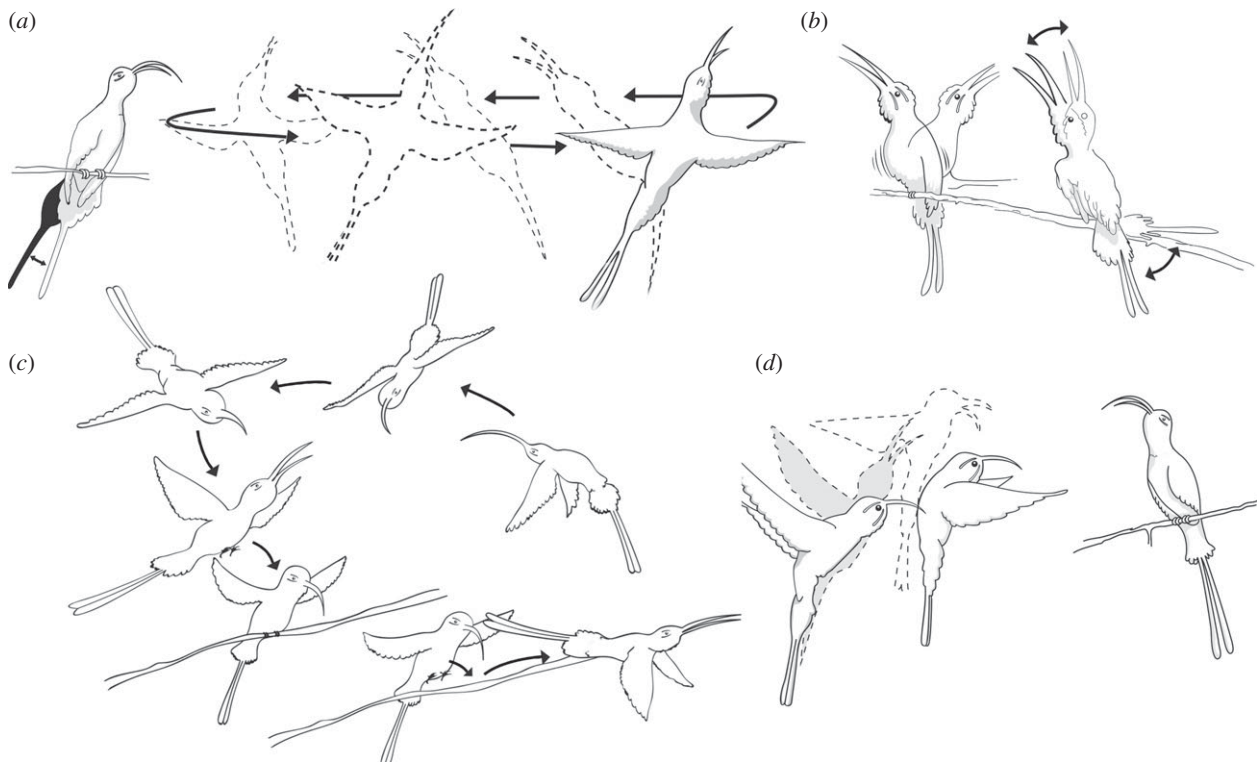


Figure 2. Four of the twelve observed visual display elements of long-billed hermits: (a) float, (b) side by side, (c) perch exchange and (d) bill pop. See also electronic supplementary material, video S1. Additional display elements are depicted elsewhere [22].

average values for continuous parameters, individual accumulated repertoire proportions across visual displays or mean pairwise dissimilarity between element transitions among individuals were used as the sampling units for analyses at song-neighbourhood and lek levels. Similarly, lek averages for continuous parameters, lek accumulated repertoire proportions across individuals, and mean pairwise dissimilarity between element transitions among leks were used as sampling units when evaluating variation at the site level. Variation of fine-scale parameters was not evaluated at the individual level, as there were not enough individuals with sufficient measures at the lek level.

For analyses in which individuals were the sampling unit, we focused on the 21 individuals with at least four elements in their repertoire and at least two display sequences (mean recording time = 37.26 h). As repertoire composition (presence-absence of display elements) did not vary considerably between leks, variation in this parameter was not evaluated. Males with only one display and displays with fewer than four elements were excluded from analyses in which individuals or displays were the sampling units, respectively. The associations between parameter similarity, geographical distance and genetic distance were evaluated at the lek level. All sequence analyses were done in R [38]. Manipulation of visual element sequences was done with the package TraMineR [37,39]. The package vegan [40] was used for Mantel tests and dissimilarity indices.

3. Results

(a) Vocal signals

We characterized song structure using pairwise spectrographic cross-correlation [41]. Song structure was more similar between individuals from the same lek than between individuals from different leks (Mantel $r = 0.49$, $n_{\text{birds}} = 60$, $n_{\text{lek}} = 6$, $p = 0.0001$; figure 3). Song structure was not significantly different between sites (Mantel $r = 0.05$, $n_{\text{birds}} = 60$, $p = 0.54$; figure 3) and did not correlate with geographical

distance (partial Mantel test controlling for lek-level similarity: Mantel $r = -0.20$, $n_{\text{birds}} = 60$, $p = 0.82$). Song structure did not correlate with genetic distance among the four La Selva leks (Mantel $r = -0.21$, $n_{\text{birds}} = 19$, $p = 0.99$). Songs also contained individual signatures; spectrographic structure was more similar between song renditions from the same individual than among individuals (Mantel $r = 0.25$, $n_{\text{songs}} = 293$, $n = 60$, $p < 0.0001$). Song structure was also more similar between individuals from the same song neighbourhood than from different neighbourhoods (partial Mantel test controlling for lek-level similarity, $r = 0.29$, $n_{\text{birds}} = 60$, $n_{\text{song types}} = 10$, $p < 0.0001$). Although this result is unsurprising given that song neighbourhoods are defined by the sharing of song types, it does demonstrate the utility of our analysis to detect signal variation at this social scale.

(b) Visual signals

Visual displays were recorded for 56 lekking males at the six leks (mean recording time = 29.76 h). We observed 12 display elements across all leks (figure 2 and table 1; electronic supplementary material, video S1), 10 of which were previously described [21], and two of which are described for the first time here (perch displacement and bill poke; table 1). The same visual repertoire of 12 elements was found in all but one of the leks; in lek CCL only 10 elements were observed (electronic supplementary material, figure S1). We found that the number of elements did not increase with sampling effort after sampling approximately 15 display sequences (electronic supplementary material, figure S1). 95% confidence intervals of bootstrapped richness did not differ from 12 display elements in any lek (including CCL) as well as for all leks combined (electronic supplementary material, figure S2).

For analyses at the individual level, song-neighbourhood level and lek level, we focused on the 21 individuals with at least four elements in their vocal repertoire and at least two

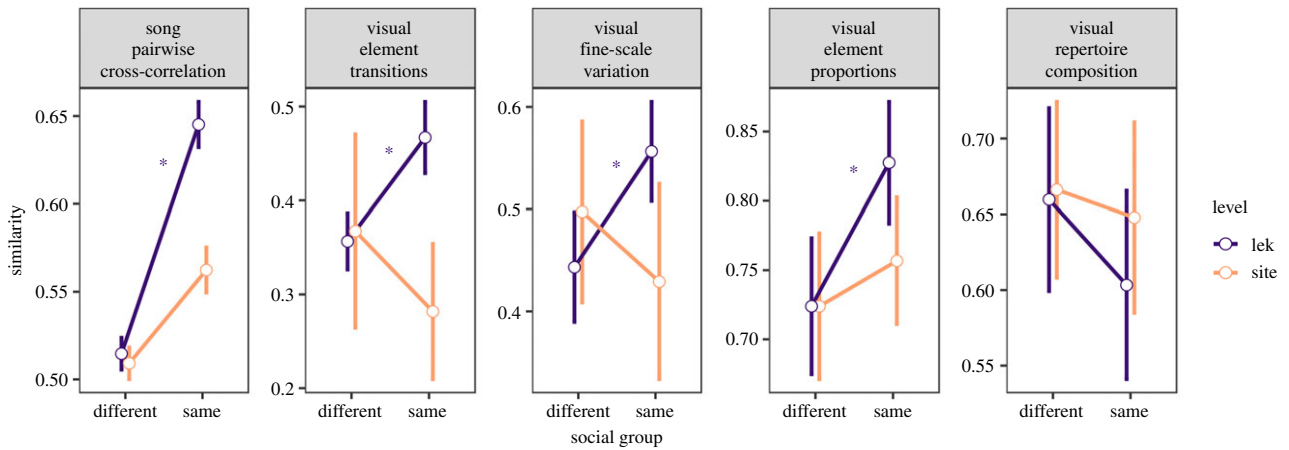


Figure 3. Dissimilarity of acoustic and visual displays (mean \pm s.e.) at the site and lek level. All parameters except for visual repertoire composition showed greater similarity within leks than among leks. No such pattern was observed across sites for any of the acoustic or visual display parameters. Leks and individuals were the sampling units for site- and lek-level tests respectively. Asterisks reflect significant differences at $p < 0.05$ from Mantel tests comparing within-lek versus between-lek similarities (i.e. a lek-level vocal or visual signature).

display sequences (mean recording time = 37.26 h, mean number of recording sessions = 2.53). We measured four components of visual displays: repertoire composition (presence–absence of display elements), element proportion (relative abundance of display elements), element transitions (probability of transition between display elements) and fine-scale parameters. Visual displays contained individual signatures of resident males in repertoire composition (Mantel $r = 0.09$, $n_{\text{sequences}} = 67$, $n_{\text{birds}} = 11$, $p = 0.046$), element proportions (Mantel $r = 0.10$, $n_{\text{sequences}} = 67$, $n_{\text{birds}} = 11$, $p = 0.0042$) and transitions (Mantel $r = 0.09$, $n_{\text{sequences}} = 62$, $n_{\text{birds}} = 7$, $p = 0.037$), suggesting that a representative sample of visual displays was taken within individuals. Intruder male signatures were also detected in transitions (Mantel $r = 0.08$, $n_{\text{sequences}} = 61$, $n_{\text{birds}} = 10$, $p = 0.021$) but not in repertoire composition (Mantel $r = 0.009$, $n_{\text{sequences}} = 61$, $n_{\text{birds}} = 5$, $p = 0.36$) or element proportions (Mantel $r = 0.02$, $n_{\text{sequences}} = 61$, $n_{\text{birds}} = 5$, $p = 0.27$).

At the lek level, similarity of visual displays was significantly higher between individuals from the same lek for element proportions (Mantel $r = 0.12$, $n_{\text{birds}} = 21$, $n_{\text{leks}} = 6$, $p = 0.045$; figure 2), transitions (Mantel $r = 0.27$, $n_{\text{birds}} = 21$, $n_{\text{leks}} = 6$, $p = 0.00014$; figure 2), and fine-scale parameters (Mantel $r = 0.21$, $n_{\text{birds}} = 15$, $n_{\text{leks}} = 5$, $p = 0.02$; figure 2), but not in repertoire composition (Mantel $r = -0.082$, $n_{\text{birds}} = 21$, $n_{\text{leks}} = 6$, $p = 0.38$). This pattern, based on data from all recorded birds, remained similar when we restricted our analysis to our best-sampled site, La Selva. In contrast to acoustic displays, visual displays did not show higher similarity between members of the same song neighbourhood within leks with multiple song types (partial Mantel tests controlling for lek-level similarity: repertoire composition: $r = -0.04$, $n_{\text{birds}} = 14$, $n_{\text{leks}} = 4$, $p = 0.66$; element proportions: $r = 0.13$, $n_{\text{birds}} = 14$, $n_{\text{leks}} = 4$, $p = 0.083$; element transitions: $r = 0.005$, $n_{\text{birds}} = 14$, $n_{\text{leks}} = 4$, $p = 0.46$; fine-scale parameters: $r = 0.13$, $n_{\text{birds}} = 15$, $n_{\text{leks}} = 4$, $p = 0.075$). None of the visual display parameters were correlated with genetic distance (repertoire composition: $r = -0.16$, $n_{\text{birds}} = 20$, $n_{\text{leks}} = 4$, $p = 0.76$; element proportions: $r = -0.08$, $n_{\text{birds}} = 15$, $n_{\text{leks}} = 4$, $p = 0.68$; element transitions: $r = -0.17$, $n_{\text{birds}} = 15$, $n_{\text{leks}} = 4$, $p = 0.87$; fine-scale parameters $r = -0.1$, $n_{\text{birds}} = 8$, $n_{\text{leks}} = 4$, $p = 0.64$).

In our analysis of visual displays, we did not find site-level signatures in either element proportions (Mantel

$r = -0.38$, $n_{\text{leks}} = 6$, $p = 0.99$), element transitions (Mantel $r = 0.005$, $n_{\text{leks}} = 6$, $p = 0.46$) or fine-scale parameters (Mantel $r = -0.16$, $n_{\text{leks}} = 6$, $p = 0.53$). There was no association between similarity in any visual parameter and geographical distance (all leks: element proportions: $r = 0.23$, $n_{\text{leks}} = 6$, $p = 0.41$; element transitions: $r = -0.26$, $n_{\text{leks}} = 6$, $p = 0.73$; fine-scale parameters: $r = -0.27$, $n_{\text{leks}} = 6$, $p = 0.62$; La Selva leks: element proportions: $r = 0.09$, $n = 4_{\text{leks}}$, $p = 0.12$; element transitions: $r = 0.16$, $n = 4_{\text{leks}}$, $p = 0.33$; fine-scale parameters: $r = 0.09$, $n = 4_{\text{leks}}$, $p = 0.41$). Finally, at La Selva leks, no significant association between genetic distance and lek membership ($r = -0.09$, $n_{\text{birds}} = 20$, $p = 0.84$, electronic supplementary material, figure S3), song-neighbourhood membership ($r = -0.08$, $n_{\text{birds}} = 19$, $p = 0.82$) or geographical distance was detected ($r = -0.144$, $n_{\text{birds}} = 20$, $p = 0.94$).

4. Discussion

We evaluated variation in visual signals at the same geographical scales and with the same statistical approach as variation in songs, a signal modality in which both vocal learning and the resulting geographical dialects have been documented in long-billed hermits and related hummingbird species [14,22]. We found that both visual and vocal signals vary at small geographical scales, suggesting a role for social learning in the acquisition of visual displays. Songs showed distinctive acoustic signatures at the lek and song-neighbourhood level, while three out of the four visual signal features (element proportions, transitions among elements, and fine-scale parameters) showed lek-level signatures. Genetic analysis indicated no significant genetic differences between leks or song neighbourhoods. Altogether, our results point to social learning as the most parsimonious explanation for the observed dialect-like spatial variation of visual displays, which would be consistent with a rarely tested prediction of the motor theory for the origins of vocal learning: that visual display learning should also be present in vocal learning species.

The motor theory for the origins of vocal learning proposes that neural circuits dedicated to vocal learning arose from general motor learning circuits through a process of brain pathway duplication [7,42]. Another theory for the

origins of learned vocal communication systems, the mirror-systems hypothesis, holds that learned gestural signals were the evolutionary precursor to spoken language in humans [8]. These hypotheses share the prediction that, in species that rely on learning to acquire vocal signals, visual displays may also be acquired, at least in part, through learning. To date, however, little evidence has existed demonstrating the co-occurrence of visual and vocal learning. Variation of gestures consistent with social learning has been described in bonobos [43], but vocal learning seems to be absent or very limited in most non-human primates [2,6,44,45]. Bottlenose dolphins show impressive vocal learning skills [46], but only inconclusive evidence exists on the learning of visual signals [47,48]. A learning-driven development of visual displays in long-billed hermits would represent the only known co-occurrence of motor learning in both visual and vocal realms in a non-human animal, which is consistent with an evolutionary link between vocal and general motor learning.

Song neighbourhoods (identified by visual inspection of spectrograms) showed a detectable signature in their acoustic structure. Similarly, individuals from the same lek shared acoustic features in their songs, even at leks with more than one song type. Altogether, the results support the view that vocal distinctiveness has arisen by social learning [14,22], as has been suggested previously [17,20,49]. Overall, the performance of the analytical approach provides confidence in the variation we detected in visual parameters over the same geographical scales using the same statistical approaches.

Our results suggest that genetic variation is not a strongly viable alternative explanation for lek-level signatures in vocal and visual displays. This lack of genetic structure that we identified among leks is consistent with documented patterns of individual movement between leks. Approximately 4% of individuals captured at leks have been found attending more than one lek, including both juvenile and adult males (M.A.-S. 2010–2015, unpublished data). Females have been also shown to visit several leks [22], suggesting high potential for gene flow among leks. In addition, leks have a small population size (typically less than 20 males), and are found in close proximity of neighbouring leks (approx. 1 km), at a distance similar to that covered during daily foraging trips in this species [21]. Altogether, our data suggest no genetic differentiation in the sampled population. Similarly, differences at the site level might plausibly be associated with genetic differentiation as gene flow can be expected to be lower among sites than among leks within a site, but we found no consistent behavioural differences among sites. Although we lack sufficient data to exclude the possibility that a small number of genes have differentially segregating alleles that might be responsible for the behavioural differences, it seems most likely that such behavioural differences are not related to genetic differences between individuals. Rather, our genomic results are more consistent with gene flow between leks and imply the observed behavioural differences in vocal and visual displays are learned.

An alternative scenario is that visual displays are innate, and males with similar display characteristics preferentially settle on the same lek. However, this scenario stands at

odds with what we know about the territory acquisition process in long-billed hermits. Juveniles typically move from lek to lek competing with established males until they finally take ownership of a territory [22], and the likelihood of becoming territorial is determined in part by foraging efficiency, body size and weapon size (elongated bill tips used to stab opponents [50]) relative to other males in a lek [51]. Overall, the evidence suggests that variation in visual display structure does not play a major role when establishing a lek territory.

Environmental factors, such as habitat differences, can contribute to variation in vocal [52] or visual [53] displays among animals in different locations. We consider environmental variation to be an unlikely explanation for the patterns seen in the long-billed hermit. All of the leks in this study were found in relatively close proximity within mature lowland tropical rainforest, and showed no obvious differences in vegetative structure [22]. Furthermore, environmental features are expected to influence animal signals that transmit long distances through vegetation, but hermit displays are performed in very close proximity—usually within a few centimetres—and thus any habitat differences that might exist are unlikely to exert a strong influence on signal form at different leks.

We expect our findings to provide grounds for further investigation of visual display learning. Applying our experimental design to other vocal learning and non-vocal learning species could bring insight into how widely learning is involved in the development of vocal and visual signals, and whether visual display learning indeed evolved first as predicted by the motor theory of learning [7]. Another fruitful research avenue would be assessing learning-driven expression patterns of the gene *FoxP2* in neural centres responsible for motor control, given that this gene has been shown to play an important role in promoting the plasticity required for vocal learning in a range of taxa [54–56]. Together, these approaches could shed further light on the evolutionary origins of vocal learning.

Ethics. All activities described were reviewed and authorized by the Institutional Animal Care and Use Committee at the New Mexico State University (IACUC-2011–020) and were performed under the research permits 152-2009-SINAC and 063-2011-SINAC from Costa Rican authorities.

Data accessibility. The data associated with this paper are available in the Dryad Digital Repository: <https://doi.org/10.5061/dryad.gn8qf6q> [57].

Authors' contributions. M.A.-S., P.L.G.-G., D.J.M. and T.F.W. conceived and designed the study, M.A.-S., G.S.-V. and T.F.W. collected field data, M.A.-S., G.S.-V., T.F.W. and J.C. conducted analyses and wrote the paper. All authors reviewed, edited and approved the paper.

Competing interests. We have no competing interests.

Funding. The study was funded by National Geographic Society (CRE grant no. 9169-12), College of Arts and Science and Biology Department at New Mexico State University, Organization for Tropical Studies, and Animal Behavior Society.

Acknowledgements. We thank Alejandro Rico-Guevara for logistical support, Jordan Price, Christopher J. Clark and Michael Arbib for thoughtful discussions, Alejandra Galindo, Agustin Vega and Leah Harper for assistance in fieldwork, the University of Wisconsin Biotechnology Center for DNA library preparation and sequencing services, and Mariano Araya for providing display drawings.

1. Janik V, Slater P. 2000 The different roles of social learning in vocal communication. *Anim. Behav.* **60**, 1–11. (doi:10.1006/anbe.2000.1410)
2. Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Cheng L, Slocombe KE. 2015 Vocal learning in the functionally referential food grunts of chimpanzees. *Curr. Biol.* **25**, 495–499. (doi:10.1016/j.cub.2014.12.032)
3. Jarvis E. 2004 Learned birdsong and the neurobiology of human language. *Ann. N. Y. Acad. Sci.* **1016**, 749–777. (doi:10.1196/annals.1298.038)
4. Nowicki S, Searcy WA. 2014 The evolution of vocal learning. *Curr. Opin. Neurobiol.* **28**, 48–53. (doi:10.1016/j.conb.2014.06.007)
5. Sewall KB, Young AM, Wright TF. 2016 Social calls provide novel insights into the evolution of vocal learning. *Anim. Behav.* **120**, 163–172. (doi:10.1016/j.anbehav.2016.07.031)
6. Tyack P. 2008 Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *J. Comp. Psychol.* **122**, 319–331. (doi:10.1037/a0013087)
7. Feenders G, Liedvogel M, Rivas M, Zapka M, Horita H, Hara E, Jarvis ED. 2008 Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS ONE* **3**, e1768. (doi:10.1371/journal.pone.0001768)
8. Arbib MA. 2012 *How the brain got language: the mirror system hypothesis*. New York, NY: Oxford University Press.
9. Wright TF, Dahlin CR. 2018 Vocal dialects in parrots: patterns and processes of cultural evolution. *Emu* **118**, 50–66. (doi:10.1080/01584197.2017.1379356)
10. Podos J, Warren P. 2007 The evolution of geographic variation in birdsong. *Adv. Study Behav.* **37**, 403–458. (doi:10.1016/S0065-3454(07)37009-5)
11. Weilgart L, Whitehead H. 1997 Group-specific dialects and geographical variation in coda repertoire in South Pacific sperm whales. *Behav. Ecol. Sociobiol.* **40**, 277–285. (doi:10.1007/s002650050343)
12. Esser KH, Schubert J. 1998 Vocal dialects in the lesser spear-nosed bat *Phyllostomus discolor*. *Naturwissenschaften* **85**, 347–349. (doi:10.1007/s001140050513)
13. Wright TF. 1996 Regional dialects in the contact call of a parrot. *Proc. R. Soc. Lond. B* **263**, 867–872. (doi:10.1098/rspb.1996.0128)
14. Araya-Salas M, Wright TF. 2013 Open-ended song learning in a hummingbird. *Biol. Lett.* **9**, 20130625. (doi:10.1098/rsbl.2013.0625)
15. Schuchmann K. 1999 Family Trochilidae (hummingbirds). In *Handbook of the birds of the world* (eds J Hoyo, A Del Elliot, J Sargatal), pp. 468–535. Barcelona, Spain: Lynx Edicions.
16. Clark CJ. 2009 Courtship dives of Anna's hummingbird offer insights into flight performance limits. *Proc. R. Soc. B* **276**, 3047–3052. (doi:10.1098/rspb.2009.0508)
17. Baptista LF, Schuchmann KL. 1990 Song learning in the Anna hummingbird (*Calypte anna*). *Ethology* **84**, 15–26. (doi:10.1111/j.1439-0310.1990.tb00781.x)
18. Gonzalez C, Ornelas JF. 2009 Song variation and persistence of song neighborhoods in a lekking hummingbird. *Condor* **111**, 633–640. (doi:10.1525/cond.2009.090029)
19. Jarvis ED, Ribeiro S, Da Silva ML, Ventura D, Vielliard J, Mello CV. 2000 Behaviourally driven gene expression reveals song nuclei in hummingbird brain. *Nature* **406**, 628–632. (doi:10.1038/35020570)
20. Gaunt SLL, Baptista LF, Sanchez JE, Hernandez D. 1994 Song learning as evidenced from song sharing in two hummingbird species (*Colibri coruscans* and *C. thalassinus*). *Auk* **111**, 87. (doi:10.2307/4088508)
21. Wiley R. 1971 Song groups in a singing assembly of Little Hermits. *Condor* **73**, 28–35. (doi:10.2307/1366121)
22. Stiles FG, Wolf LL. 1979 *Ecology and evolution of lek mating behavior in the long-tailed hermit hummingbird*. Washington, DC: American Ornithological Union.
23. Ligges U, Krey S, Mersmann O, Schnackenberg S. 2014 tuneR: analysis of music. R package version 1.2.1.
24. Sueur J, Aubin T, Simonis C. 2008 Equipment review: seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics* **18**, 213–226. (doi:10.1080/09524622.2008.9753600)
25. Araya-Salas M, Smith-Vidaurre G. 2017 warbleR: an R package to streamline analysis of animal acoustic signals. *Methods Ecol. Evol.* **8**, 184–191. (doi:10.1111/2041-210X.12624)
26. Helyar SJ *et al.* 2011 Application of SNPs for population genetics of non-model organisms: new opportunities and challenges. *Mol. Ecol. Resour.* **11**, 123–136. (doi:10.1111/j.1755-0998.2010.02943.x)
27. Morin PA, Luikart G, Wayne RK. 2004 SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* **19**, 208–216. (doi:10.1016/j.tree.2004.01.009)
28. Toews DP *et al.* 2015 Genomic approaches to understanding population divergence and speciation in birds. *Auk* **133**, 13–30. (doi:10.1642/AUK-15-51.1)
29. Han J-I, Kim J-H, Kim S, Park S-R, Na K-J. 2009 A simple and improved DNA test for avian sex determination. *Auk* **126**, 779–783. (doi:10.1525/auk.2009.08203)
30. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, 10. (doi:10.1371/journal.pone.0019379)
31. Korfach J, Gedman G, Kingan SB, Chin CS, Howard JT, Audet JN, Jarvis ED. 2017 De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *Gigascience* **6**, 85. (doi:10.1093/gigascience/gix085)
32. Li H. 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1403.2877. (<http://arxiv.org/abs/1303.3997>)
33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Durbin R. 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
34. Smouse PE, Long JC, Sokal RR. 1986 Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* **35**, 627–632. (doi:10.2307/2413122)
35. Morisita M. 1959 Measuring of the dispersion of individuals and analysis of the distributional patterns. *Mem. Fac. Sci. Kyushu Univ. Ser. E* **2**, 5–235.
36. Real R, Vargas JM. 1996 The probabilistic basis of Jaccard's index of similarity. *Syst. Biol.* **45**, 380–385. (doi:10.1093/sysbio/45.3.380)
37. Studer M, Ritschard G. 2016 What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *J. R. Stat. Soc. Ser. A Stat. Soc.* **179**, 481–511. (doi:10.1111/rssa.12125)
38. R Core Team. 2018 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
39. Gabadinho A, Ritschard G, Müller NS, Studer M. 2011 Analyzing and visualizing state sequences in R with TraMineR. *J. Stat. Softw.* **40**, 1–37. (doi:10.18637/jss.v040.i04)
40. Dixon P. 2003 VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930. (doi:10.1111/j.1654-1103.2003.tb02228.x)
41. Khanna H, Gaunt S, McCallum DA. 1997 Digital spectrographic cross-correlation: test of sensitivity. *Bioacoustics* **7**, 209–234. (doi:10.1080/09524622.1997.9753332)
42. Chakraborty M, Jarvis ED. 2015 Brain evolution by brain pathway duplication. *Phil. Trans. R. Soc. B* **370**, 20150056. (doi:10.1098/rstb.2015.0056)
43. Pika S, Liebal K, Tomasello M. 2005 Gestural communication in subadult bonobos (*Pan paniscus*): repertoire and use. *Am. J. Primatol.* **65**, 39–61. (doi:10.1002/ajp.20096)
44. Egnor SER, Hauser MD. 2004 A paradox in the evolution of primate vocal learning. *Trends Neurosci.* **27**, 649–654. (doi:10.1016/j.tins.2004.08.009)
45. Takahashi DY, Fenley AR, Teramoto Y, Narayanan DZ, Borjon JI, Holmes P, Ghazanfar AA. 2015 The developmental dynamics of marmoset monkey vocal production. *Science* **349**, 734–738.
46. Reiss D, McCowan B. 1993 Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): evidence for vocal learning. *J. Comp. Psychol.* **107**, 301–312. (doi:10.1037/0735-7036.107.3.301)
47. Herman L. 2002 *Exploring the cognitive world of the bottlenosed dolphin*. Cambridge, MA: MIT Press.

48. Bauer GB, Johnson CM. 1994 Trained motor imitation by bottlenose dolphins (*Tursiops truncatus*). *Percept. Mot. Skills* **79**, 1307–1315. (doi:10.2466/pms.1994.79.3.1307)
49. González C, Ornelas JF. 2014 Acoustic divergence with gene flow in a lekking hummingbird with complex songs. *PLoS ONE* **9**, e109241. (doi:10.1371/journal.pone.0109241)
50. Rico-Guevara A, Araya-Salas M. 2015 Bills as daggers? A test for sexually dimorphic weapons in a lekking hummingbird. *Behav. Ecol.* **26**, 21–29. (doi:10.1093/beheco/aru182)
51. Araya-Salas M, Gonzalez-Gomez P, Wojczulanis-Jakubas K, López V, Wright TF. 2018 Spatial memory is as important as weapon and body size for territorial ownership in a lekking hummingbird. *Sci. Rep.* **8**, 2001. (doi:10.1038/s41598-018-20441-x)
52. Boncoraglio G, Saino N. 2007 Habitat structure and the evolution of bird song: a meta-analysis of the evidence for the acoustic adaptation hypothesis. *Funct. Ecol.* **21**, 134–142. (doi:10.1111/j.1365-2435.2006.01207.x)
53. Arriero E, Fargallo JA. 2006 Habitat structure is associated with the expression of carotenoid-based coloration in nestling blue tits *Parus caeruleus*. *Naturwissenschaften* **93**, 173–180. (doi:10.1007/s00114-006-0090-5)
54. Haesler S, Wada K, Nshdejan A, Morrisey EE, Lints T, Jarvis ED, Scharff C. 2004 FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.* **24**, 3164–3175. (doi:10.1523/JNEUROSCI.4369-03.2004)
55. Whitney O, Pfenning AR, Howard JT, Blatti CA, Liu F, Ward JM, Sinha S. 2014 Core and region-enriched networks of behaviorally regulated genes and the singing genome. *Science* **346**, 1256780. (doi:10.1126/science.1256780)
56. Heston JB, White SA. 2015 Behavior-linked FoxP2 regulation enables zebra finch vocal learning. *J. Neurosci.* **35**, 2885–2894. (doi:10.1523/JNEUROSCI.3715-14.2015)
57. Araya-Salas M, Smith-Vidaurre G, Mennill DJ, González-Gómez PL, Cahill J, Wright TF. 2019 Data from: Social group signatures in hummingbird displays provide evidence of co-occurrence of vocal and visual learning. Dryad Digital Repository. (doi:10.5061/dryad.gn8qf6q)