






## Variability in humpback whale songs reveals how individuals can be distinctive when sharing a complex vocal display

Luca Lamoni,<sup>1,a)</sup>  Ellen C. Garland,<sup>1</sup>  Jenny A. Allen,<sup>2,b)</sup>  Jennifer Coxon,<sup>1</sup> Michael J. Noad,<sup>2</sup>  and Luke Rendell<sup>1</sup> 

<sup>1</sup>Centre for Social Learning and Cognitive Evolution/Sea Mammal Research Unit, School of Biology, University of St. Andrews, St. Andrews KY16 9TH, United Kingdom

<sup>2</sup>Cetacean Ecology and Acoustics Laboratories, School of Veterinary Science, The University of Queensland, Gatton, Queensland 4343, Australia

### ABSTRACT:

Individually distinctive acoustic signals in animal vocal communication are taxonomically widespread, however, the investigation of these signal types in marine mammals has focused only on a few species. Humpback whale songs are a stereotyped, hierarchically structured vocal display performed by males, and hence thought to be sexually selected. Within a population, whales conform to a common version of the song despite the song constantly evolving. While humpback songs have been studied extensively at the population level, individual level variation has been rarely described, with inconclusive results. Here, we quantified inter- and intra-individual variability at different levels in the song hierarchy using songs from 25 singers across two song types from the eastern Australian population song of 2002 (12 singers), and the revolutionary song introduced in 2003 (13 singers). Inter-individual variability was found heterogeneously across all hierarchical levels of the song structure. In addition, distinct and individually specific patterns of song production were consistently recorded across song levels, with clear structural differences between the two song types. These results suggest that within the constraints of song conformity, males can produce individually distinctive patterns that could function as an advertisement to females to convey individual qualities. © 2023 Acoustical Society of America. <https://doi.org/10.1121/10.0017602>

(Received 17 July 2022; revised 20 February 2023; accepted 25 February 2023; published online 12 April 2023)

[Editor: James F. Lynch]

Pages: 2238–2250

### I. INTRODUCTION

In animal vocal communication, senders often broadcast signals containing information about themselves such as species, local population, age, body size, dominance status, and individual identity (Bradbury and Vehrencamp, 2011). Individual information can be encoded in acoustic signals of variable complexity, from brief, function-specific “calls” produced by both sexes all year round, to longer, complex “songs” produced by males during the breeding season [Catchpole and Slater (2008b); for examples of inter-individual variability in contact and group cohesion calls in South Polar skuas see Charrier *et al.* (2001), in vampire bats see Carter *et al.* (2012), in hyenas see Mathevon *et al.* (2010), in gorillas see Salmi *et al.* (2014), in African elephants see Soltis *et al.* (2005), in giant pandas see Charlton *et al.* (2009), in sperm whales see Gero *et al.* (2016), in bottlenose dolphins see Janik and Sayigh (2013)]. Individual information can also be present in advertisement calls produced to broadcast various aspects of the quality of potential

mates to females and/or other males as seen for example in deer (Reby *et al.*, 1998; Vannoni and McElligott, 2008) and anurans (Pettitt *et al.*, 2013; Röhr *et al.*, 2020).

Songs generally have longer duration and more complex structural features compared to calls (Catchpole and Slater, 2008b). In birds, several studies have documented significant individual differences in structural (e.g., arrangement of utterances into song types) and acoustic (frequency content and duration) features of songs [e.g., Sandoval *et al.* (2014), Petrusková *et al.* (2016), Bauer and Nagl (1992), and Lovell and Lein (2004)]. Song features such as repetition rate, singing rate, and repertoire size, and acoustic features like frequency modulation, can be positively selected by females because they are indicative of good mate quality (Searcy, 1992; Catchpole and Slater, 2008a).

Among marine mammals, humpback whales (*Megaptera novaeangliae*) provide an interesting example of an apparently sexually selected vocal trait: a complex, stereotyped song display (Payne and McVay, 1971), transmitted across multiple populations over wide geographical scales (Garland *et al.*, 2011). The fundamental element of this hierarchically organized display is the shortest continuous sound the human ear can perceive: the “unit.” A stereotyped combination of units constitutes a “phrase” (Payne and McVay, 1971). A complex phrase, constituted of multiple “motifs” (Winn and Winn, 1978), can be subdivided into “sub-phrases” (Cholewiak *et al.*,

<sup>a)</sup>Current address: International Council for the Exploration of the Sea (ICES), H.C. Andersens Boulevard 44-46, 1553 Copenhagen V., Denmark. Electronic mail: luca.lamoni@ices.dk

<sup>b)</sup>Also at: Southern Ocean Persistent Organic Pollutants Program, Centre for Planetary Health and Food Security, Griffith University, Nathan, Queensland 4111, Australia.

2013). The repetition of a phrase forms a “theme” (Payne and McVay, 1971; Frumhoff, 1983), although these phrase repetitions are not always exact copies of one another. This variability can lead to the presence, within the same theme, of multiple “phrase variants” (Payne *et al.*, 1983). A series of themes sung in a stereotyped sequence is defined as a “song” or “song cycle” (Payne and McVay, 1971). Singers tend to sing in continuous bouts of several hours, termed “song sessions,” containing multiple cyclical repeats of the song. Songs composed using different themes are classified as distinct “song types” (Garland *et al.*, 2011).

A general finding among the literature is that, within a population, humpback whales show a high degree of individual conformity to a common version of the song, even if this is itself constantly evolving (Payne *et al.*, 1983; Winn and Winn, 1978; Cholewiak *et al.*, 2013). Song content can either change gradually [“song evolution,” Payne *et al.* (1983)] or more rapidly [“song revolution,” Noad *et al.* (2000)] over time. Several studies have tracked the gradual addition, deletion, and/or modification of sound units and themes both within breeding seasons and across multiple years [e.g., Cerchio *et al.* (2001), Payne and Payne (1985), Payne *et al.* (1983), Cato (1991), and Allen *et al.* (2018)]. Songs recorded during periods dominated by this gradual change are referred to here as “evolutionary songs.” Song revolutions, by contrast, are the rapid and complete replacement of the song in a population by the introduction of a novel “revolutionary song.” In all identified examples the revolutionary song came from an adjacent population (Garland *et al.*, 2011; Noad *et al.*, 2000).

While song conformity within a population is well established, sexual selection requires individual distinctiveness, and relatively few authors have focused their efforts on understanding inter-individual song variability. An additional difficulty faced by researchers in measuring inter-individual variability is that, due to the song’s hierarchical structure, variability can occur at multiple levels, and it can also occur over time as song evolves. For example, Hafner *et al.* (1979) measured six acoustic parameters of a single unit type from a sample of six individuals recorded in the Caribbean and suggested that unit features could be used to discriminate among the recorded individuals. However, the individuals were recorded over a period of two years, and thus the authors might have misinterpreted song evolution within and between breeding seasons as inter-individual variability (Cholewiak *et al.*, 2013; Macknight *et al.*, 2001). Subsequent studies of individual variation have mainly focused on investigating how the acoustic features of a subset of song units vary across different individuals, with contradictory results. Frankel (1996) found significant inter-individual differences among six unit types from eleven individuals in Hawaii recorded over an eight day period. In contrast, Cerchio (1993) measured variability within and among 24 individuals recorded across several months off Hawaii and Mexico for several acoustic parameters without finding significant differences. Similarly, a study conducted off eastern Australia found no consistent differences in the

features of two unit types that could help discriminate among six different individuals recorded over a 12 day period (Macknight *et al.*, 2001). More recently, a study conducted off the coast of Brazil found individual differences in the proportional use of two different units within a theme from ten different recording sessions likely belonging to ten distinct individuals (Arraut and Vielliard, 2004). Multiple studies have documented that one of the most variable features of songs is the number of phrase repeats produced in a theme (Arraut and Vielliard, 2004; Cerchio, 1993; Cholewiak *et al.*, 2013). Smith (2009) reported how phrase repetition increased in certain themes when a singer was recorded while escorting a female compared to when recorded alone, suggesting that this feature is variable within as well as between individuals, depending on the singer’s behavioural context. Finally, while looking at song complexity over a broad time scale, Allen *et al.* (2018) suggested that song individuality, defined by the presence of individually unique phrase variants, tended to increase during evolutionary song change.

The study of a sexually selected trait and its inter- and intra-individual variability can provide insights into how trait development and female preference might evolve over evolutionary time scales, as well as how trait expression and associated fitness variations might change over individual life histories (Johnstone, 1995; Griffith and Sheldon, 2001). We are currently far from a full understanding of the nature of this variability in humpback whale song, so more information is needed. Here, we report a detailed analysis of intra- and inter-individual song variability in the songs sung by the eastern Australian humpback population in 2002 and 2003. There were two song types, as in 2003 a new revolutionary song was adopted. Our aim was to understand how the individually distinctive song patterns expected under sexual selection could be found even in a highly conformist song system. We had two objectives. First, we quantified inter-individual variability at different hierarchical levels across two song types—a sexual selection account would expect to find consistent individual differences in at least some of these levels. Second, we measured the consistency of any individually distinct patterns found by examining the relationship between intra- and inter-individual variability across singers. Since song needs to be a reliable individual level signal for sexual selection to operate, we expected variation within individuals to be smaller than variation between.

## II. MATERIALS AND METHODS

### A. Summary of methods

Humpback whale song recordings were collected on the east coast of Australia in 2002 and 2003. The song units within the recordings were then manually classified by two people based on the visual patterns of the spectrograms and auditory characteristics of the sounds. The classification consistency was then evaluated using a Random Forest classification analysis using frequency and duration variables measured from the spectrograms of a subset of units of all

whales recorded. The results suggested that the manual classification was highly consistent. Units were then transcribed into alphanumeric strings and grouped into “motifs” (phrases or sub-phrases) and themes. Motif and song (sequence of themes) similarity was subsequently measured using the Levenshtein distance method both within and among singers. Finally, the presence of individually distinctive patterns of both motif and song production was statistically evaluated using partial Mantel tests.

## B. Study location and recordings

The songs used in this study were recorded in September and October of 2002 and 2003 off Peregrine Beach (26°30'S, 153°07'E), on the east coast of Australia, during the southward migration of the eastern Australian breeding population of humpback whales. These data were collected as part of a larger collaborative project known as the Humpback whale Acoustic Research Collaboration (HARC) (Noad *et al.*, 2004). A detailed description of the song recording system can be found in the supplementary material (supplementary methods section)<sup>1</sup>.

Our study area is part of this population’s migratory corridor and whales display a consistent southward migratory behaviour during the months of September and October (Noad *et al.*, 2004; Cato, 1991; Williamson *et al.*, 2016; Noad and Cato, 2007; Allen *et al.*, 2018; Kavanagh *et al.*, 2017). Because of this consistent behaviour we considered song sessions recorded on different days as belonging to different individuals. In one instance, a song session started in the late evening extended until the early hours of the next day. As this song session was uninterrupted, it was attributed to one individual (ID 6). Most song sessions used, however, we recorded during daylight hours and so were usually at least 14 h apart.

Several hundred hours of song were recorded over the two-year study period [172 h in 2002; 285 h in 2003; see Smith (2009), Garland *et al.* (2011), Rekdahl (2012), and Allen *et al.* (2018)]. From this database we selected recordings that fulfilled two criteria: (1) a signal to noise ratio of at least 6 dB so that all units could be clearly identified and (2) one recognisable singer could be followed for the entire duration of its song session without interruptions longer than one minute. We only sampled one song session per day. While six song cycles are generally considered representative of an individual’s current song repertoire (Allen *et al.*, 2018; Garland *et al.*, 2017a; Owen *et al.*, 2019), we transcribed an average of ten song cycles per individual (min = 2, max = 23, SD = 5).

Three main factors contributed to the wide range of song cycles transcribed. (1) The variability in the number of themes in each song cycle meant that shorter (or longer) songs cycles could be completed more rapidly (or more slowly), increasing (or decreasing) the overall song cycle count. (2) Good quality recordings could be sometimes interrupted abruptly due to multiple singers overlapping on top of each other, singers stopping singing or moving

outside the study area, resulting in a variable number of song cycles transcribed. (3) Finally, certain recordings with relatively fewer song cycles were retained in the analysis due to their biological relevance. An example of this are the two song cycles from the singer ID 27 which sung the 2002 song type during 2003.

## C. Unit, motif, and theme classification

For each recording, two human classifiers (L.L. and either J.C. or one of three experienced research assistants) aurally and visually inspected spectrograms produced with Adobe Audition 3.0 (Blackman-Harris; 4096 FFT; 90% window overlap; frequency range inspected: 0–5 kHz) and classified each unit present. This protocol was consistent with previous classification methods used in studies that looked at the same song types (Smith *et al.*, 2008; Garland *et al.*, 2011; Rekdahl, 2012; Allen *et al.*, 2018). To verify the repeatability and consistency of our qualitative unit classifications, we conducted a Random Forest classification analysis using R (R Core Team, 2016) with the *randomForest* package (Liaw and Wiener, 2002). We selected a subsample of units (N = 1,125; 26 unit types) out of the total number of units transcribed (N = 42 613). The selected subset represented each individual’s unit repertoire and all units in one phrase for each theme produced by each singer were measured. Twelve frequency and duration variables were measured for each unit using RAVEN PRO 1.4 (Hann window, 4096 FFT, 16 bits, 5.4 Hz frequency resolution and 75% overlap) following previous studies (Garland *et al.*, 2017a; Garland *et al.*, 2017b; Charif *et al.*, 2010). The Random Forest classification analysis was parametrised with  $m_{try} = 4$  (N of variables randomly sampled as candidates at each split) and  $n_{tree} = 10\ 000$  (N of trees to grow). These parameters were based on the stability of the classification of uncertainty of each tree [i.e., lowest OOB estimate of error rate; OOB, or “out-of-bag” data is the prediction of data which is not in the bootstrap sample; see Liaw and Wiener (2002), Rankin *et al.* (2013), Garland *et al.* (2017b), and Breiman (2001)]. There was 92.8% agreement between the model and the human classifiers in the classification of 1125 units into 26 unit types confirming the high consistency of the manual song transcriptions (see Table II in the supplementary material<sup>1</sup> for Random Forest confusion matrix and classification error for each unit).

Units were transcribed into alphanumeric strings and then grouped into phrases and themes following previously published classifications of these song types (Rekdahl *et al.*, 2013; Garland *et al.*, 2012; Rekdahl, 2012; Smith, 2009; Miksis-Olds *et al.*, 2008; Allen *et al.*, 2017). However, to capture the complexity present in some phrase types, in terms of the number and type of units, units were grouped into sub-phrases (instead of single phrase types) where these were evident. A sub-phrase is a “sequence of one or more units that is sometimes repeated in a series” within a single phrase [Cholewiak *et al.* (2013) and Winn and Winn (1978) defined these units of repetition “motifs”] and either term has been

used by several authors to describe phrase structure in detail (Perazio *et al.*, 2018; Magnúsdóttir and Lim, 2019; Kowarski *et al.*, 2019). For example, the phrase constituting theme 3a (Fig. 1 in the supplementary material) is composed by two motifs (i.e., sub-phrases): motif c (repeated twice; an “ascending moan” followed by several “grunts”) and motif f (an ascending moan followed by three “trumpets”). Where it was possible to divide phrases into sub-phrases, we did so (Fig. 1 in the supplementary material).<sup>1</sup> Where we refer to either indivisible phrases or sub-phrases, we use the term “motif” (Winn and Winn, 1978) throughout the manuscript.

Themes could consist of repeated or alternating motifs (phrases and sub-phrases, respectively), following the hierarchical structure described in the literature (Cholewiak *et al.*, 2013; Payne and McVay, 1971). Spectrograms of all classified motifs from the 2002 and 2003 song types can be found in Figs. 2 and 3 in the supplementary material.<sup>1</sup>

#### D. Levenshtein distance (LD) analysis

Song transcription resulted in (1) sequences (or strings) of units (forming motifs) and (2) sequences of themes (forming songs). Multiple song cycles from the same individual in the same recording constituted a song session. We quantified sequence similarities using the Levenshtein distance (LD) method, which returns a distance between a pair of sequences based on the number of insertions, deletions, and substitutions needed to turn one string into another (Kohonen, 1985; Garland *et al.*, 2012). These LDs are then standardised by the length of the longest string in the pair and subtracted from 1 to give a similarity measure referred to hereafter as the Levenshtein Similarity Index or LSI (Garland *et al.*, 2012; Garland *et al.*, 2013; Garland *et al.*, 2017b). We performed this analysis at two hierarchical levels. First, we used strings of units to look at motif composition, similar to Garland *et al.* (2017b) and Murray *et al.* (2018). Second, we used strings of themes to evaluate song composition; in both cases we used the routines developed by Garland *et al.* (2012) and Garland *et al.* (2017b) and the analysis was run unweighted in that all unit changes were treated equally irrespective of the units’ acoustic features. If multiple analysed strings are repetitions of a particular set (for example, multiple repetitions of a given motif from the same individual), a helpful step is to designate the string with the highest summed similarity to all the other strings in its set to be what we called the “median string” (Kohonen, 1985). We used this string as representative of its set when comparing with different sets of strings. We calculated, for each individual, (1) a median string of units for each motif type and (2) a median string of themes for each individual’s entire song.

All LSI analyses were run in R (R Core Team, 2016) using published code [Garland *et al.* (2017b); available online (Garland, 2023)]. The primary analysis output is an LSI matrix that includes all the pairwise string similarity coefficients for the set of provided strings. These comparisons could be both between and within individuals—details on the specific matrices calculated are given below. The LSI matrices generated in this study were visualised both using hierarchical cluster analysis

with average-linkage clustering and by assigning different colours to different levels (percentages) of similarity within a matrix plotted as an image (the latter type of plot was produced in MATLAB). Two methods were used to assess the statistical uncertainty of the cluster analysis. (1) A bootstrapping analysis run in R (R Core Team, 2016) over 1000 iterations using the package *pvclust*, generating the *p*-values for each cluster (Suzuki and Shimodaira, 2006; Garland *et al.*, 2012; Garland *et al.*, 2017a; Garland *et al.*, 2017b). This package provides approximately unbiased (AU) probability values (significance for  $p > 95\%$ ) using multi-scale bootstrap resampling as well as bootstrap probability (BP) values (significance for  $p > 70\%$ ). (2) Cophenetic correlation coefficients (CCC) (Sokal and Rohlf, 1962), which measure how well the pairwise distances among the similarity matrix is preserved ( $CCC > 0.8$  indicates good data representation), were also calculated with R (Garland *et al.*, 2017a).

#### E. Quantifying individual variation

To investigate individual distinctiveness at the motif and song levels, we analysed the LSI matrices using partial Mantel tests (Smouse *et al.*, 1986). Although there are some concerns about the validity of the partial Mantel test in spatial ecology studies (Guillot and Rousset, 2013; Legendre *et al.*, 2015), we have restricted our analyses to testing the relationship of dissimilarity matrices, for which this test is still considered useful (Legendre *et al.*, 2015). We measured the correlation between the LSI matrix and an individual categorical matrix (ICM) of the same size in which 1 represented a pair of motifs or songs that were produced by the same individual and 0 represented a pair of motifs or songs produced by different individuals (Araya-Salas *et al.*, 2019; González and Ornelas, 2014; Ruegg *et al.*, 2006; Fischer *et al.*, 2020). Humpback song evolves over time at the population level (Cerchio *et al.*, 2001; Payne *et al.*, 1983; Guinee *et al.*, 1983), which could confound individual variation when different individuals are recorded at different times in the breeding season. To account for these potential evolutionary song changes, we included a time difference matrix (TDM) in the partial Mantel test which contained the time differences (calculated in number of days, to the nearest hour) between the recording dates of pairs of singers producing the same song type. Thus, a significant positive correlation in these tests would indicate that songs sung by the same individual were significantly more similar to each other than songs sung by different individuals, while taking into account any song change over time.

The partial Mantel correlation scores *r* could range between  $-1$  (strong negative correlation),  $0$  (no correlation), and  $1$  (strong positive correlation). We performed these tests in R (R Core Team, 2016) using the package *vegan* (Oksanen *et al.*, 2017) with the Pearson correlation method and 9999 permutations (Ruegg *et al.*, 2006; Fischer *et al.*, 2020). Significance levels were adjusted using a Bonferroni correction for multiple comparisons in R using the *p.adjust* function (Ruegg *et al.*, 2006). Examples of motif and song string

sequences as well as relevant custom written code used for the analysis carried out in this paper are available online (Lamoni, 2023).

### III. RESULTS

The song transcription process resulted in data from 25 singers (12 from 2002 and 13 from 2003), comprising 260 song cycles and 42 613 units (Table III in supplementary material<sup>1</sup>). Whale ID 27 was recorded in 2003 (3rd September) but produced only themes from the 2002 song type—this whale had not made the switch to the revolutionary 2003 song at the time of recording. For this reason, whale ID 27 was included in the 2002 dataset for analysis so that the two song types were analysed separately. Therefore, the final dataset comprised 13 singers with the 2002 song type and 12 with the 2003 song type. A total of 5599 motifs were transcribed for the 2002 song type and grouped into 13 motif types, while the 2003 dataset comprised 3557 motifs organised into 7 motif types.

#### A. Variation between individuals at the motif level using individual motif median sequences

The unit of analysis here was the set median string calculated from the set of all productions of a given motif by a given individual (e.g., the median string for motif “a” for individual 1). LSI matrices were calculated separately for the 2002 and 2003 motif datasets. Results are grouped below following the three most distinct patterns that became apparent while investigating motif variability among individuals (Fig. 1).

The first distinct pattern was motifs that showed little variability among individuals. Median sequences of motifs i and g (2002) and n, o, and p (2003) generally presented extremely low levels of inter-individual variability (Fig. 1;

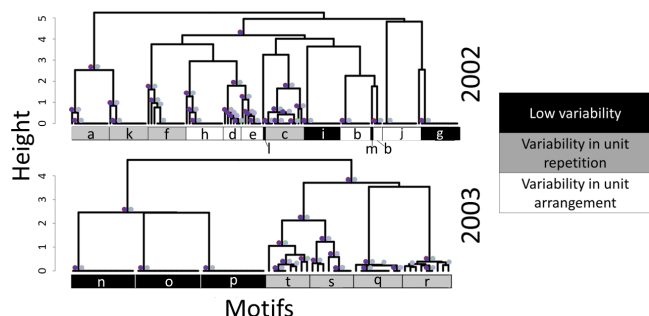


FIG. 1. (Color online) Summary of the inter-individual variability found among the singers' median strings for each motif in 2002 (upper panel) and 2003 (lower panel). The 2002 and 2003 motif dissimilarity matrices were bootstrapped (1000 times) and are represented by the respective dendrograms. Multiscale bootstrap resampling (AU, left, dark-purple dot indicates  $p > 95\%$ ) and normal bootstrap probabilities (BP, right, light-purple dot indicates  $p > 70\%$ ) are shown. Each dendrogram node represents a different individual. A horizontal black line above a motif type (height = 0) indicates an identical motif structure among the singers. The three patterns of motif variability between individuals are shown with different colour shadings: black for low variability, gray for variability in the number of repetitions of the same unit, and white for variability in the arrangement of the same units. Motifs l and m were sung only by ID 14 (2002). The CCC scores calculated for the 2002 and 2003 dendrograms were 0.98 and 0.99, indicating excellent data representation.

black shading; height = 0). This means that irrespective of the singer, these motifs were sung with a consistently high level of conformity throughout the sampling period.

The second distinct pattern was variability in unit repetition. Here, at least one singer varied in the number of repetitions of the same unit type within a motif. For example, in the 2002 song type, individuals varied in the number of grunt, trumpet, and “croak” repeats used in motifs a, f, c, and k, respectively (Figure 1; gray shading; height > 0). Likewise, in 2003, individuals varied in the number of grunts, “ratchets,” “high-squeaks,” and “bird-whistles” used in motifs q, r, s, and t, respectively (Fig. 1; gray shading).

The last distinct pattern identified in our data were individual variability in the unit arrangement within a motif. The inter-individual variability displayed in motifs b, d, e, h, and j in the 2002 evolutionary song was mainly due to variable arrangements of the same (or similar) unit types (Fig. 1; white shading; height > 0). For example, in 2002 whales ID 13, 14 and 26 made a substitution in the initial unit of motif b, singing a “modulated-moan” instead of the “ascending-moan” sung by most individuals. The cluster formed by motifs d and e displayed a larger number of branches compared to other 2002 motifs, indicating a relatively high degree of inter-individual variability (Fig. 1; white shading). Motifs d and e were formed by a similar pool of units (excluding “descending cries” and “short-shrieks” occasionally present in e). Motifs d and e could also be identified as forming a “transitional” phrase (Payne and Payne, 1985) due to the fact that they incorporated units of the preceding motif (f) with units of the following (h). When analysing population-wide patterns of song production, it is common practice to discard transitional phrases due to their potential variability (Allen *et al.*, 2018; Allen *et al.*, 2017; Garland *et al.*, 2017b). However, considering both our goal to quantify individual variability and the importance of transitional phrases in terms of song production and learning (Garland *et al.*, 2017a), we decided to keep these transitional elements in our analysis. Finally, it is noteworthy that in the revolutionary song of 2003 there were no examples of motif types showing this kind of individual variability.

#### B. Variation within and among individuals at the motif level using all motif sequences

We computed an LSI matrix for each motif type across all singers using all the strings available for each individual (not medians). Matrix size corresponded to the total number of times that motif occurred in the dataset. The relevant ICM (individual categorical matrix) and TDM (time difference matrix) constructed for each motif type were used, together with the corresponding LSI matrix, in the partial Mantel tests. These tested whether motifs produced by the same individual were more similar than those produced by different individuals while attempting to also account for the changes over time we know occur during breeding seasons. Four qualitative categories arose from this analysis (Fig. 2).

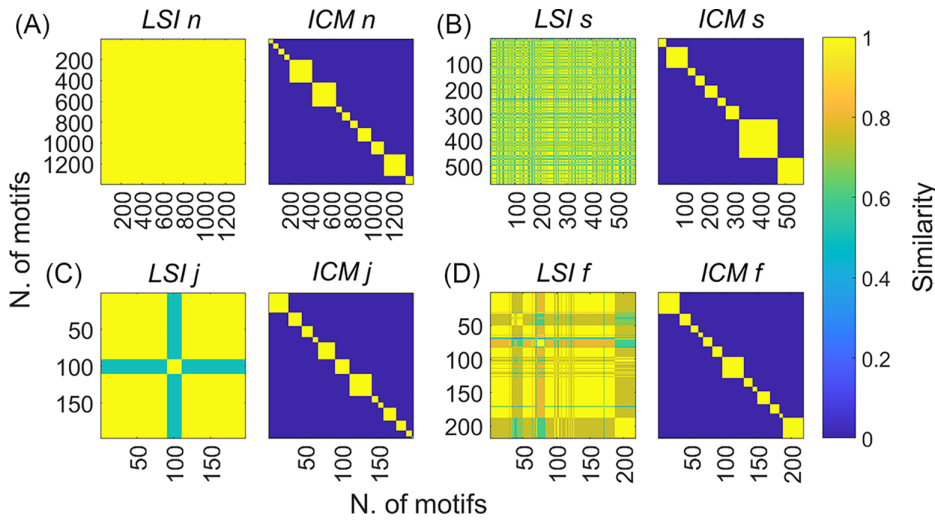


FIG. 2. (Color online) Four examples representative of the four categories of motif variation within and among individuals found in the analysis (A: motif n; B: s; C: j; D: f). In each example, the correlation between the motif LSI matrix (left) and its correspondent ICM matrix (right) is calculated considering the respective TDM matrix ( $r_A$ : NaN;  $r_B$ : 0.052;  $r_C$ : 0.151;  $r_D$ : 0.144). Each yellow square in the ICM represents one individual; its linear size indicates its sample size (i.e., number of times the motif was recorded). In a scenario of complete convergence (i.e., LSI matrix in A) all individuals display maximum similarity scores, whereas in a scenario of complete individuality the LSI matrix and ICM would be identical. TDMs are not represented here to show more clearly to the reader the correlations between LSI matrix and ICM.

**1. Category A—Complete conformity (correlation coefficient  $r$  not available as zero distances obtained)**

Motifs n, o, and p in the 2003 song type showed complete conformity across all individuals, corresponding to the low motif variability pattern shown by the set-median analysis (black shading; Fig. 1). A clear example is motif n [Fig. 2(A)], sung with complete consistency, both within and among all 12 singers, more than 1200 times in total. Due to this complete lack of variation, it was not possible to carry out the partial Mantel tests for these motifs (Table I).

**2. Category B—Equal variability within and between individuals ( $-0.003 < r < 0.096$ )**

For motifs a, b, g, h, i, q, r, and s inter- and intra-individual variability were equivalent across all individuals with no specific pattern present [Fig. 2(B)]. This led to low correlation scores in the partial Mantel test ( $-0.003 < r < 0.096$ , only 4 out of 8 tests had  $p < 0.05$ ; Table I). As an example, motif s presented the same level of variability within and among individuals [partial Mantel correlation score  $r = 0.052$ ; Fig. 2(B) and Table I]. This means that our

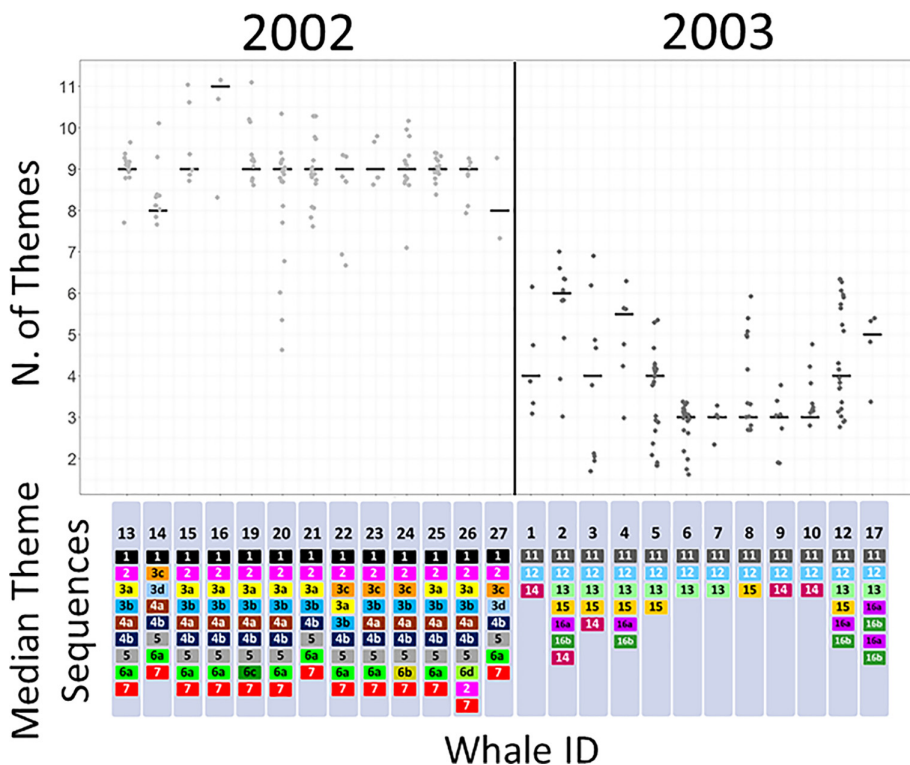


FIG. 3. (Color online) Upper panel: number of themes (including phrase variants) used in each song by each individual in 2002 (light gray dots on the left) and 2003 (dark gray dots on the right). The horizontal lines represent the median of each individual's set of songs. Data points, each representing one song cycle, were jittered to aid visualisation and hence do not all line up precisely with the number of themes. Lower panel: individual median song sequences calculated using the LSI analysis for all individuals. Whale ID's are listed in row 1 and are consistent for both panels.

TABLE I. Mantel correlation coefficients and their corresponding  $p$ -values for all motif types of 2002 and 2003 song. Significant correlation coefficients are shown in bold. Due to the large number of statistical tests, all  $p$ -values were adjusted using a Bonferroni correction. There are no results for motifs n, o, and p because all individuals producing these motifs showed complete conformity, hence every distance was zero—this can be read as an extreme lack of correlation with the individual categorical matrix (ICM), but mathematically cannot be computed.

Song type	Motifs	LSI matrix size	Category	Mantel correlation $r$	$p$ -value
2002	a	1408 × 1408	<b>B</b>	<b>0.023</b>	<b>0.001</b>
2002	b	749 × 749	<b>B</b>	<b>0.096</b>	<b>0.001</b>
2002	c	600 × 600	<b>C</b>	<b>0.109</b>	<b>0.001</b>
2002	d	138 × 138	<b>D</b>	<b>0.179</b>	<b>0.001</b>
2002	e	134 × 134	<b>D</b>	<b>0.183</b>	<b>0.001</b>
2002	f	218 × 218	<b>D</b>	<b>0.144</b>	<b>0.001</b>
2002	g	738 × 738	B	−0.003	0.750
2002	h	307 × 307	<b>B</b>	<b>0.069</b>	<b>0.001</b>
2002	i	959 × 959	B	0.016	0.276
2002	j	197 × 197	<b>C</b>	<b>0.151</b>	<b>0.001</b>
2002	k	135 × 135	<b>C</b>	<b>0.156</b>	<b>0.001</b>
2003	n	1395 × 1395	A	NA	NA
2003	o	1377 × 1377	A	NA	NA
2003	p	438 × 438	A	NA	NA
2003	q	168 × 168	B	0.014	0.750
2003	r	172 × 172	B	0.016	0.750
2003	s	189 × 189	<b>B</b>	<b>0.052</b>	<b>0.047</b>
2003	t	115 × 115	<b>C</b>	<b>0.141</b>	<b>0.001</b>

pool of singers produced variable versions of these motifs with little to no individual distinctiveness.

**3. Category C—Individual distinctiveness (one singer; 0.109 < r < 0.156)**

This category corresponded to both variability patterns observed previously: variability in unit repetition (motifs c, k, and t; gray shading; Fig. 1) and variability in unit arrangement (motif j; white shading; Fig. 1). Here, conformity in motifs such as j and k was high across all repetitions of the motif by all individuals except one (ID 20, 2002). This whale consistently sang a modified version of the common motifs. For example, it sang motif j 20 times, in all renditions with a “long-bark” followed by a “mini-siren” instead of a bark [Fig. 2(C)]. Motif types in which these individual modifications occurred presented higher, and statistically significant, partial Mantel correlation scores (0.109 < r < 0.156, all  $p < 0.05$ ; Table I).

**4. Category D—Individual distinctiveness (multiple singers; 0.144 < r < 0.183)**

In motifs such as f (corresponding to variability in unit repetition; gray shading; Fig. 1), d and e (corresponding to variability in unit arrangement; white shading; Fig. 1) multiple singers displayed individual distinctiveness. For example, each of three individuals consistently sang their own versions of the common f motif, sung typically by all the other singers with an ascending-moan and three trumpets.

While singer ID 14 (2002) omitted the initial ascending-moan, ID 19 (2002) and ID 27 (the singer in 2003 producing the 2002 song) consistently performed the motif with different numbers of trumpets [Fig. 2(D)]. The partial Mantel correlation coefficients for this category of motifs were the highest encountered in the motif dataset (0.144 < r < 0.183, all  $p < 0.05$ ; Table I). The partial Mantel tests carried out on this category of motif types allowed us to reject the null hypothesis of no relationship between LSI and ICM matrices, suggesting a higher similarity within rather than among individuals for certain motifs.

**C. Variation among individuals at the song level using individual song median sequences**

The dataset at the song level contained 260 song cycles, 129 from 13 singers for the 2002 song type and 131 from 12 singers for the 2003 song type (see Table III in supplementary material<sup>1</sup>). Themes could be formed by either a single motif (a phrase without sub-phrases or phrase variants) or by multiple motifs (more than one sub-phrase and/or phrase variants). Generally, themes from the 2002 song type included more numerous and variable motif combinations [i.e., phrase variants; Garland *et al.* (2017a) and Allen *et al.* (2018)] compared to the 2003 song type (Table IV in supplementary material<sup>1</sup>). Before conducting the LSI analysis, phrase repetitions (formed by one or more motifs) that formed each theme within each theme sequence were removed to avoid the potential confounding effects of the singers’ behavioural context so each phrase was represented once at the position the first repetition occurred in the song (Smith, 2009).

In 2002, seven main themes were consistently sung by the majority of sampled individuals (Table IV in supplementary material<sup>1</sup>). Three of the common themes (3, 4, and 6) also had phrase variants that were commonly sung (3a, 3b, 3c, 4a, 4b, and 6a). Different phrase variants were assigned based on the different motif combinations that characterised them; for example, 3a and 3b both started with a repetition of motif c, but while the former ended with motif f, the latter ended with motifs d and e. Three themes (8, 9, and 10) and four phrase variants of the main themes (3d, 6b, 6c, and 6d) were “uncommon” in the 2002 dataset being found in less than 15% of all song cycles and/or sung by less than half of the singers (Table IV in supplementary material<sup>1</sup>). Whale ID14 (2002) was the only singer that produced the uncommon themes 9 and 10. In 2003, all 6 themes were used by more than half (8/12) of the whales, including the common phrase variants of theme 16 (16a and 16b).

Considering both “main” themes and those containing phrase variants, the theme repertoire of the 2002 song type was larger than the one present a year later in the revolutionary song of 2003 (Table IV in supplementary material<sup>1</sup>). However, this difference became even more pronounced when examining the number of themes used in each song cycle by singers in both years (Fig. 3). In 2002, 10 out of 13 individuals used a median of nine themes (including those with phrase variants) in each cycle, and none had a median lower than 8, while 9 of 12 whales in 2003 produced a

median of 3 or 4 themes per song, with none having a median greater than 6. Of the 129 song cycles from 2002, only in 7 did the number of themes used in the song cycle fall within the range observed in 2003.

In both song types individuals varied in the extent to which consecutive song cycles had different number of themes, confirming that the songs produced within a single song session by the same individual were not exact copies of one another. Individual median theme sequences were much longer in 2002 (minimum of eight themes) compared to 2003 (minimum of three themes; Fig. 3). While some of the 2003 song type singers produced predominately long theme sequences (for example ID 2, 4, and 17) others showed a much more limited repertoire of themes (ID 6, 7, and 9) resulting in a clear dichotomy (absent in 2002) in the grouping of 2003 median theme sequences (Fig. 3).

#### D. Variation within and among individuals at the song level using all song sequences

The partial Mantel tests carried out with the 2002 LSI song matrix [Fig. 4(A)] and the corresponding ICM [Fig. 4(B)], while accounting for time differences, returned a significant correlation coefficient of 0.254 ( $p = 0.0002$ ). The same analysis carried out with the 2003 matrix [Figs. 4(C) and 4(D)] resulted in a lower correlation coefficient of 0.101 ( $p = 0.0002$ ). This allowed us to reject for both song types the null hypothesis that the LSI and ICM matrices were unrelated, suggesting, for some of the individuals in our study, a high degree of individual distinctiveness. However, it is clear there is more individual distinctiveness in the 2002 data, with double the correlation with individual

identity compared to 2003. In addition, the inclusion of the motif (i.e., phrase and sub-phrase) repetitions that formed each theme, in the theme sequences used in the LSI analysis, led to almost the same Mantel correlation coefficients (2002:  $r = 0.254$ ,  $p = 0.0002$ ; 2003:  $r = 0.088$ ,  $p = 0.0002$ ). This indicated that the presence of within-theme motif repetitions did not lead to increased individual distinctiveness.

### IV. DISCUSSION

In this study we investigated intra- and inter-individual song similarity among humpback whales recorded off eastern Australia over two successive breeding seasons that encompassed a song revolution. The large variation in theme production by individuals in 2003 as well as the fine scale, individually distinct, motif modifications seen in the 2002 song type suggested that while conformity to a general song type at all hierarchical levels was a consistent feature of humpback whale songs, individually distinctive patterns of production were also present. These patterns were noticeably stronger in the pre-revolution 2002 songs compared to the analyses of the 2003 revolutionary song.

#### A. Multiple sources of inter-individual variability

Our results highlight two main sources of motif variability. The first and most common type of variability occurred in motifs that included varying repetitions of the same unit type. The repetition of the same sound element might have different functions. For example, repetition can be used as a proxy to convey fitness information in rock hyraxes (*Procapra capensis*) where larger individuals are able to produce more numerous and longer “chuck” call bouts compared to smaller individuals (Koren and Geffen, 2009). Moreover, repeating the same sound element can also function as a compensation mechanism to maintain communication efficiency in unfavourable conditions (Lengagne *et al.*, 1999; Miller *et al.*, 2000). In humpback whales, variation in phrase repetition (rather than repetition of song units within phrases or sub-phrases) has been linked to both behavioural context (Smith, 2009), and exposure to anthropogenic noise [low-frequency active sonar; Miller *et al.* (2000)]. Although we did not account for singers’ behavioural context in our data, our results suggest that motif (i.e., phrase and sub-phrase) repetition does not increase individual distinctiveness. Within motifs, repetitions of the same unit might be variable both within and between individuals, yet a few examples were also present where specific unit repetition patterns resulted in some individual distinctiveness.

The second source of inter-individual variability in motifs, such as d and e, was combinatorial, with different individuals arranging the pool of units of the same motif in different ways, resulting in motif variants. This type of variability was also found in humpback whale songs off the Brazilian coast where different singers produced distinctive arrangements of the same phrase type (Arraut and Vielliard, 2004). The act of forming different combinations of sounds

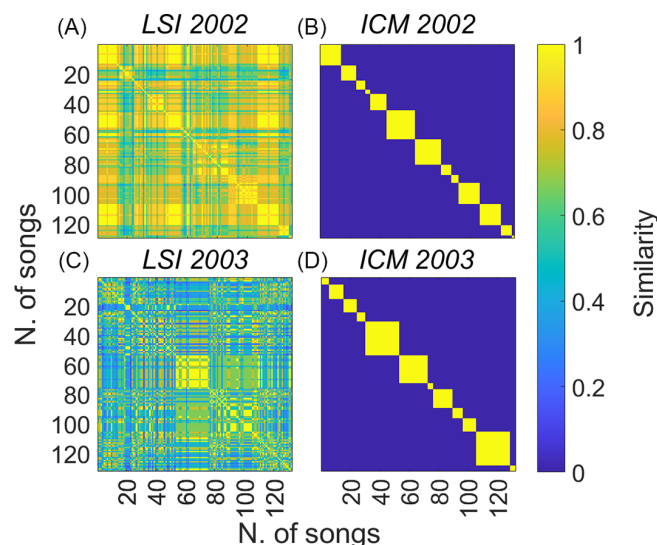


FIG. 4. (Color online) Matrices used for the partial Mantel tests at the song level. The correlation between the 2002 (A) and 2003 (C) song LSI matrices with their corresponding ICM (B and D respectively) were calculated also taking into account the respective TDM. Each yellow square in the ICM represents one individual; its linear size indicates its sample size (i.e., number of song cycles recorded). In a scenario of complete convergence all individuals in the LSI matrix display maximum similarity scores whereas in a scenario of complete individuality the LSI matrix and ICM would be identical.



by an animal can be meaningless (“phonological syntax”; Kroodsma and Momose, 1991; Collier *et al.*, 2014; Berwick *et al.*, 2011) or, more rarely, may convey different information to conspecifics depending on the combination of meaningful sounds [“lexical syntax”; e.g., Campbell’s monkey (*Cercopithecus campbelli*), Ouattara *et al.* (2009a) and Ouattara *et al.* (2009b); banded mongoose (*Mungos mungo*), Jansen *et al.* (2012); Japanese great tit (*Parus minor*), Suzuki *et al.* (2016) and Collier *et al.* (2014)]. Furthermore, the salience of different combinations of sounds varies among species. While combinatorial changes in familiar songs may elicit behavioural changes in Japanese great tits (Suzuki *et al.*, 2016) and Carolina chickadees [*Poecile carolinensis*; Clucas *et al.* (2004)], they are poorly discriminated by species like zebra finches [*Taeniopygia guttata*; Mol *et al.* (2021)] and indigo buntings [*Passerina cyanea*; Emlen (1972)]. It is beyond the scope of this paper to conclusively place humpback whale songs on the continuum between phonological and lexical syntax. However, the concurrent presence of motifs with low and high inter-individual variability recorded here fits well within a recent hypothesis provided by Murray *et al.* (2018), who suggested that humpback whale songs might be a multi-message signal in which different phrases of a song convey multiple types of information, such as species/population recognition and signaller quality (Murray *et al.*, 2018; Hebets and Papaj, 2005).

The song level analysis showed a clear structural contrast between the theme sequences that formed the song of 2002 and those of the revolutionary song of 2003. The total number of commonly sung themes was 7 in 2002 and 6 in 2003. However, the 2002 song type displayed a higher number of phrase variants (6 in common themes and 4 in uncommon ones; Table IV in supplementary material<sup>1</sup>) and uncommon themes (3) compared to the 2003 song type (2 phrase variants and 0 uncommon themes; Table IV in supplementary material<sup>1</sup>). Despite using different methods for both transcription and analysis, Allen *et al.* (2018) found a similar pattern in which the percentage of individually unique phrases was halved in the 2003 song (31.4% in 2002 vs 12.3% in 2003).

The number of themes used in each song, as well as for median theme sequences in our data (Fig. 3), highlighted a structural distinction between the two song types. Individuals from 2002 invariably used more themes per song than the 2003 singers. Inter-individual song variability also differed between the two song types. In 2002, there was a general degree of similarity among individuals, with few exceptions like whale ID 14, and ID 27, the individual recorded in 2003 producing the 2002 song type. By contrast, 2003 had two quite dissimilar variations of the basic song: a short and a long version that varied in terms of number of themes used (3 vs 4–7; Fig. 3). Despite this dichotomy, some individuals that presented a short median theme sequence also sporadically sang the longer song version.

The disparity in the theme content of the two songs and the different levels of inter-individual variability in our sample can be attributed to the different nature of these two song types (Allen *et al.*, 2018). Allen *et al.* (2018) provided a long-

term analysis (from 2002 to 2014) of how song complexity and individuality fluctuated based on song revolution occurrences. To characterise intra- and inter-individual song variability more thoroughly, we focused on a single revolution event. This gave us an increased sample size, in terms of number of singers and number of song cycles, compared to Allen *et al.* (2018). During the time period between revolutions, songs undergo cultural evolution through the addition of new units, the lengthening of existing ones, and the modification of phrases (Payne and Payne, 1985; Payne *et al.*, 1983; Cerchio *et al.*, 2001). This evolutionary process usually results in a general increase in complexity (Allen *et al.*, 2018). As Allen *et al.* (2018) suggest, the relatively rich repertoire of units/motifs/themes and the higher presence of individually distinctive patterns displayed by singers in 2002 compared to 2003 likely resulted from the cultural evolution of this song type.

The origin of the 2002 song type is uncertain but can be traced back to at least 2000 (Garland *et al.*, 2011). Therefore, based on our sample, it is plausible to assume that the relatively long evolution time of the 2002 song type allowed most singers of this population to learn accurately a long and stable sequence of themes—all 2002 singers in our sample sang all the most common themes (Fig. 3). Furthermore, this evolution period might have helped some singers to develop individually distinctive patterns at both the motif and song level (Figs. 2 and 4). In contrast, our analysis of the 2003 singers showed a general lack of individual distinctiveness (both at the motif and song level; Figs. 2 and 4) and an instability of theme sequencing represented by some singers producing predominantly the long version of the 2003 song (6 themes) and others singing mostly a short version (3 themes) and, only occasionally, the long one (Fig. 3). By the same reasoning used to interpret the 2002 results, the differences displayed by the 2003 singers could represent a song type not stabilised yet due to its recent acquisition (hence short time scale for cultural evolution and diversification among individuals at motif level) during the breeding season and/or migration prior to recording. However, in future studies it will be necessary to compare more song types (before and after song revolutions) to verify that the differences discussed above are consistent and not resulting from limited sampling, random learning and/or production errors.

## B. Individual distinctiveness

Two critical requirements for individual distinctiveness are the occurrence of individually distinct features and their consistency through time (Sandoval *et al.*, 2014). In our study, the relationship between inter- and intra-individual variability both at the motif and song level have highlighted the presence of individually distinctive patterns among the singers. Furthermore, these distinctive patterns were mostly consistent through the duration of the recordings. Ideally, it would have been optimal to extend the recording time for each individual over multiple days in order to track the consistency of the individual patterns over a longer sampling period and to have a more balanced sample size across

singers. However, the migratory nature of the singers' movement through the study area did not allow this type of sampling. Therefore, if consistent across the recorded song cycles, we assumed that the individually distinctive patterns recorded were representative of the male singing behaviour at other times.

At the motif level, our partial Mantel test produced a heterogeneous scenario. While small and non-significant correlations indicated a lack of individual distinctiveness for some motif types [Figs. 2(A) and 2(B); Table I], other motif types displayed higher and significant correlations suggesting the presence of multiple individually distinct patterns of production [Figs. 2(C) and 2(D); Table I]. At the song level, the 2002 song cycles showed a higher presence of individually distinctive patterns compared to 2003, regardless of whether phrase repetitions within themes were considered (Fig. 4; Table II) and despite overall song similarity among individuals being higher in 2002 than 2003. These results suggest that fine scale changes both at the motif and song level are present and may help a singer/male "stand out" acoustically from his competitors. Furthermore, it is possible that acoustic features such as pitch and duration may help convey information about the singer as well, although previous studies that compared several acoustic parameters among a comparable pool of singers did not record any individually distinctive pattern (Cerchio, 1993; Macknight *et al.*, 2001).

Our results are consistent with the suggestion of Arraut and Vielliard (2004) that individuals may differ in their ability to learn and compose songs. In general, factors such as age, personality traits, developmental stress, and reproductive state are known to affect learning abilities across species (Mesoudi *et al.*, 2016; Thornton and Lukas, 2012). Unfortunately, individual life histories are unknown for this pool of singers. It would be consistent with findings from bird species [e.g., great reed warbler, *Acrocephalus arundinaceus*: Hasselquist *et al.* (1996); zebra finch, *Taeniopygia guttata*: Neubauer (1999); Boogert *et al.* (2011)] if, within the inherent cognitive constraints of song production (Cerchio *et al.*, 2001), a wider (or more novel) repertoire of units and/or themes were to be positively selected for by females (Noad *et al.*, 2000; Smith *et al.*, 2008) as a potential indication of male cognitive ability ["cognitive capacity hypothesis"; Boogert *et al.* (2008) and Templeton *et al.* (2014)]. Considering the continuously evolving nature of humpback whale songs and the potential key role of novelty in song selection by females, a singer on the leading edge of song change (i.e., innovating and/or learning quicker than the

others and therefore demonstrating potentially better quality mate characteristics) may benefit from higher reproductive success compared to "slower" learners—the singer ID 27 could be considered a slow learner because it sung the 2002 song type in 2003, albeit being a sample of one individual other factors could be at play. This is part of the novelty-threshold hypothesis formulated by Noad (2002), a theoretical framework into which the individual patterns observed in this study fit well. Future studies carried out on larger pools of singers will be able to test more effectively the presence of "fast" and "slow" learners and quantify other potential factors.

Humpback whale songs appear then to be characterised by a tension between conformity and individual distinctiveness. In an established song like the one of 2002, the cultural evolutionary process seemed to decrease the individual variation at the song level while increasing the occurrences of individually distinct patterns at the motif level. Thus, in this context, innovations appeared more prominent at low hierarchical levels. Across multiple years, song evolution increases song complexity and individuality (Allen *et al.*, 2018). The motif-level individual patterns observed here are the "song embellishments" suggested by Allen *et al.* (2018) to allow males to stand out amidst population-wide song conformity (Noad, 2002), as also seen in other singing species such as *Hypocnemis* antbirds (Seddon and Tobias, 2010). By contrast, the occurrence of song revolutions, like the one of 2003, leads to a decrease in song complexity (Allen *et al.*, 2018) and to a disruption of the balance between conformity and individual distinctiveness we observed in 2002. The individual variation in this case becomes more macroscopic, at the song level, and likely reflects different learning abilities and/or exposure to the novel song type.

## V. CONCLUSIONS

We quantified intra- and inter-individual song variability, identifying fine-scale individual patterns at both the level of motif and song theme sequence, within the context of the general population-level song conformity observed in all humpback whale populations. The results presented here reveal a complex scenario in which inter-individual variability is not found homogeneously across the song structure. While some motif types displayed high variability between individuals, others showed a complete lack of it, and variation both within and among individuals could indicate potentially different motif functions. Distinct individual patterns were found in both motif and songs, with differences between the evolutionary and revolutionary song types analysed here. These results suggest that within the constraints of a conformist song system, male humpback whales are able to produce individually distinctive patterns that could advertise some index of their quality as potential mating partners. Equally however, they could reflect individual idiosyncrasy resulting from learning and/or production errors that have no significance at all for females. Future studies will therefore be

TABLE II. Partial Mantel correlation coefficients and their corresponding *p*-values for all songs sung in 2002 and 2003. Significant correlation coefficients are shown in bold. All the original *p*-values were  $\leq 0.0001$ ; they have been subsequently adjusted using a Bonferroni correction.

Song type	Matrix size	Mantel Correlation <i>r</i>	<i>p</i> -value
2002	129 × 129	<b>0.254</b>	<b>0.0002</b>
2003	132 × 132	<b>0.101</b>	<b>0.0002</b>

needed to explore the biological significance of these results. Analysis of a larger and more diverse song sample size than we were able to here, incorporating multiple revolutions and longer periods of evolutionary change, is one obvious requirement. Moreover, to understand what constitutes a whale being at the forefront of song evolution, as opposed to being simply aberrant, it will be critical to look at individual life histories, correlating song characteristics and idiosyncrasies over long individually focused sampling periods combined with other data such as age, paternity success, and body size.

## ACKNOWLEDGMENTS

We would like to thank Oliver Leedham, Malcolm Kennedy, and Michael McLoughlin for their help in the song transcription process. L.L. was supported by a Leverhulme Trust Grant to L.R. (among other recipients; Grant Reference No. RPG-2013-367). L.R. was supported by the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) and their support is gratefully acknowledged. MASTS is funded by the Scottish Funding Council (Grant Reference No. HR09011) and contributing institutions. E.C.G. was funded by a Royal Society Newton International Fellowship and a Royal Society University Research Fellowship. J.A.A. was funded by an Australian Government Research Training Programme Scholarship and the Australian American Association University of Queensland Fellowship. We thank all those involved in HARC for providing the data analysed here, particularly Doug Cato and John Noad. The HARC project was funded by the U.S. Office of Naval Research, the Australian Defence Science and Technology Organisation, and the Australian Marine Mammal Centre.

<sup>1</sup>See supplementary material at <https://www.scitation.org/doi/suppl/10.1121/10.0017602> for a list of units' names and acronyms (Table I; SuppPub1.docx), a description of the song recording system (Supplementary methods; SuppPub1.docx), the Random Forest analysis confusion matrix (Table II; SuppPub1.docx), spectrograms illustrating song structure and classification method (Fig 1; SuppPub1.docx), a summary of all song transcriptions (Table III; SuppPub1.docx), theme list and composition (Table IV; SuppPub1.docx), spectrograms of all motifs (Figs. 2 and 3; SuppPub1.docx), and individual median motif sequences (Tables V and VI; SuppPub1.docx).

Allen, J. A., Garland, E. C., Dunlop, R. A., and Noad, M. J. (2018). "Cultural revolutions reduce complexity in the songs of humpback whales," *Proc. R. Soc. B: Biol. Sci.* **285**, 20182088.

Allen, J. A., Murray, A., Noad, M. J., Dunlop, R. A., and Garland, E. C. (2017). "Using self-organizing maps to classify humpback whale song units and quantify their similarity," *J. Acoust. Soc. Am.* **142**, 1943–1952.

Araya-Salas, M., Smith-Vidaurre, G., Mennill, D. J., González-Gómez, P. L., Cahill, J., and Wright, T. F. (2019). "Social group signatures in hummingbird displays provide evidence of co-occurrence of vocal and visual learning," *Proc. Biol. Sci.* **286**, 20190666.

Arraut, E. M., and Viellard, J. M. E. (2004). "The song of the Brazilian population of humpback whale *Megaptera novaeangliae*, in the year 2000: Individual song variations and possible implications," *An. Acad. Bras. Ciênc.* **76**, 373–380.

Bauer, H.-G., and Nagl, W. (1992). "Individual distinctiveness and possible function of song parts of short-toed treecreepers (*Certhia brachydactyla*)," *Evid. Multivar. Song Anal. Ethol.* **91**, 108–121.

Berwick, R. C., Okanoya, K., Beckers, G. J. L., and Bolhuis, J. J. (2011). "Songs to syntax: The linguistics of birdsong," *Trends Cogn. Sci.* **15**, 113–121.

Boogert, N. J., Fawcett, T. W., and Lefebvre, L. (2011). "Mate choice for cognitive traits: A review of the evidence in nonhuman vertebrates," *Behav. Ecol.* **22**, 447–459.

Boogert, N. J., Giraldeau, L.-A., and Lefebvre, L. (2008). "Song complexity correlates with learning ability in zebra finch males," *Anim. Behav.* **76**, 1735–1741.

Bradbury, J. W., and Vehrencamp, S. L. (2011). *Principles of Animal Communication*, 2nd ed. (Sinauer Associates, Sunderland, MA).

Breiman, L. (2001). "Random forests," *Mach. Learn.* **45**, 5–32.

Carter, G. G., Logsdon, R., Arnold, B. D., Menchaca, A., and Medellin, R. A. (2012). "Adult vampire bats produce contact calls when isolated: Acoustic variation by species, population, colony, and individual," *PLoS One* **7**, e38791.

Catchpole, C. K., and Slater, P. J. B. (2008a). "Sexual selection and female choice," in *Bird Song: Biological Themes and Variations* (Cambridge University Press, Cambridge), pp. 171–202.

Catchpole, C. K., and Slater, P. J. B. (2008b). "The study of bird song," in *Bird Song: Biological Themes and Variations* (Cambridge University Press, Cambridge), pp. 1–18.

Cato, D. H. (1991). "Songs of humpback whales: The Australian perspective," *Mem. Queensland Museum* **30**, 277–290.

Cerchio, S. (1993). Cultural evolution and geographic variation in songs of humpback whales in the eastern North Pacific, Moss Landing, CA.

Cerchio, S., Jacobsen, J. K., and Norris, T. F. (2001). "Temporal and geographical variation in songs of humpback whales, *Megaptera novaeangliae*: Synchronous change in Hawaiian and Mexican breeding assemblages," *Anim. Behav.* **62**, 313–329.

Charif, R. A., Waack, A. M., and Strickman, L. M. (2010). "Raven Pro 1.4 user's manual," Cornell Laboratory of Ornithology, Ithaca, NY.

Charlton, B. D., Zhang, Z., and Snyder, R. J. (2009). "Vocal cues to identity and relatedness in giant pandas (*Ailuropoda melanoleuca*)," *J. Acoust. Soc. Am.* **126**, 2721–2732.

Charrier, I., Jouventin, P., Mathevon, N., and Aubin, T. (2001). "Individual identity coding depends on call type in the South Polar skua *Catharacta maccormicki*," *Polar Biol.* **24**, 378–382.

Cholewiak, D. M., Sousa-Lima, R. S., and Cerchio, S. (2013). "Humpback whale song hierarchical structure: Historical context and discussion of current classification issues," *Mar. Mam. Sci.* **29**, E312–E332.

Clucas, B. A., Freeberg, T. M., and Lucas, J. R. (2004). "Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*)," *Behav. Ecol. Sociobiol.* **57**, 187–196.

Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B., and Townsend, S. W. (2014). "Language evolution: Syntax before phonology?," *Proc. R. Soc. B: Biol. Sci.* **281**, 20140263.

Emlen, S. T. (1972). "An experimental analysis of the parameters of bird song eliciting species recognition," *Behavior* **41**, 130–171.

Fischer, J., Wegdell, F., Trede, F., Pesco, F. D., and Hammerschmidt, K. (2020). "Vocal convergence in a multi-level primate society: Insights into the evolution of vocal learning," *bioRxiv:2020.2024.005504*.

Frankel, A. S. (1996). "Interindividual variation in the songs of humpback whales," *J. Acoust. Soc. Am.* **99**, 2556–2574.

Frumhoff, P. (1983). "Aberrant songs of humpback whales (*Megaptera novaeangliae*): Clues to the structure of humpback songs," in *Communication and Behavior of Whales*, edited by R. Payne (Westview Press, Boulder, CO), pp. 81–128.

Garland, E. C. (2023). "leven," <https://github.com/ellengarland/leven> (Last viewed March 13, 2023).

Garland, E. C., Goldizen, A. W., Rekdahl, M. L., Constantine, R., Garrigue, C., Hauser, N. D., Poole, M. M., Robbins, J., and Noad, M. J. (2011). "Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale," *Curr. Biol.* **21**, 687–691.

Garland, E. C., Lilley, M. S., Rekdahl, M. L., Goldizen, A. W., Noad, M. J., and Garrigue, C. (2012). "Improved versions of the Levenshtein distance method for comparing sequence information in animals' vocalisations: Tests using humpback whale song," *Behaviour* **149**, 1413–1441.

Garland, E. C., Noad, M. J., Goldizen, A. W., Lilley, M. S., Rekdahl, M. L., Garrigue, C., Constantine, R., Daeschler Hauser, N., Poole, M. M., and Robbins, J. (2013). "Quantifying humpback whale song sequences to

- understand the dynamics of song exchange at the ocean basin scale," *J. Acoust. Soc. Am.* **133**, 560–569.
- Garland, E. C., Rendell, L., Lamoni, L., Poole, M. M., and Noad, M. J. (2017a). "Song hybridization events during revolutionary song change provide insights into cultural transmission in humpback whales," *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7822–7829.
- Garland, E. C., Rendell, L., Lilley, M. S., Poole, M. M., Allen, J. A., and Noad, M. J. (2017b). "The devil is in the detail: Quantifying vocal variation in a complex, multi-levelled, and rapidly evolving display," *J. Acoust. Soc. Am.* **142**, 460–472.
- Gero, S., Whitehead, H., and Rendell, L. (2016). "Individual, unit and vocal clan level identity cues in sperm whale codas," *R. Soc. Open Sci.* **3**, 150372.
- González, C., and Omelas, J. F. (2014). "Acoustic divergence with gene flow in a lekking hummingbird with complex songs," *PLoS One* **9**, e109241.
- Griffith, S. C., and Sheldon, B. C. (2001). "Phenotypic plasticity in the expression of sexually selected traits: Neglected components of variation," *Anim. Behav.* **61**, 987–993.
- Guillot, G., and Rousset, F. (2013). "Dismantling the Mantel tests," *Methods Ecol. Evol.* **4**, 336–344.
- Guinee, L. N., Chu, K., and Dorsey, E. M. (1983). "Changes over time in the songs of known individual humpback whales (*Megaptera novaeangliae*)," in *Communication and Behavior of Whales. AAAS Selected Symposium*, edited by R. Payne (Westview Press, Boulder, CO), Vol. 76, pp. 59–80.
- Hafner, G. W., Hamilton, C. L., Steiner, W. W., Thompson, T. J., and Winn, H. E. (1979). "Signature information in the song of the humpback whale," *J. Acoust. Soc. Am.* **66**, 1–6.
- Hasselquist, D., Bensch, S., and von Schantz, T. (1996). "Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler," *Nature* **381**, 229–232.
- Hebets, E. A., and Papaj, D. R. (2005). "Complex signal function: Developing a framework of testable hypotheses," *Behav. Ecol. Sociobiol.* **57**, 197–214.
- Janik, V. M., and Sayigh, L. S. (2013). "Communication in bottlenose dolphins: 50 years of signature whistle research," *J. Comp. Physiol. A* **199**, 479–489.
- Jansen, D. A., Cant, M. A., and Manser, M. B. (2012). "Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls," *BMC Biol.* **10**, 97.
- Johnstone, R. A. (1995). "Sexual selection, honest advertisement and the handicap principle: Reviewing the evidence," *Biol. Rev. Camb. Philos. Soc.* **70**, 1–65.
- Kavanagh, A. S., Noad, M. J., Blomberg, S. P., Goldizen, A. W., Knies, E., Cato, D. H., and Dunlop, R. A. (2017). "Factors driving the variability in diving and movement behavior of migrating humpback whales (*Megaptera novaeangliae*): Implications for anthropogenic disturbance studies," *Mar. Mamm. Sci.* **33**, 413–439.
- Kohonen, T. (1985). "Median strings," *Pattern Recogn. Lett.* **3**, 309–313.
- Koren, L., and Geffen, E. (2009). "Complex call in male rock hyrax: A multi-information distributing channel," *Behav. Ecol. Sociobiol.* **63**, 581–590.
- Kowarski, K., Moors-Murphy, H., Maxner, E., and Cerchio, S. (2019). "Western North Atlantic humpback whale fall and spring acoustic repertoire: Insight into onset and cessation of singing behavior," *J. Acoust. Soc. Am.* **145**, 2305–2316.
- Kroodsma, D. E., and Momose, H. (1991). "Songs of the Japanese population of the winter wren (*Troglodytes troglodytes*)," *Condor* **93**, 424–432.
- Lamoni, L. (2023). "humpback\_whale\_song\_variability," available at [https://github.com/lucalamoni/humpback\\_whale\\_song\\_variability](https://github.com/lucalamoni/humpback_whale_song_variability) (Last viewed March 13, 2023).
- Legendre, P., Fortin, M.-J., and Borcard, D. (2015). "Should the Mantel test be used in spatial analysis?," *Methods Ecol. Evol.* **6**, 1239–1247.
- Lengagne, T., Aubin, T., Lauga, J., and Jouventin, P. (1999). "How do king penguins (*Aptenodytes patagonicus*) apply the mathematical theory of information to communicate in windy conditions?," *Proc. R. Soc. London, Ser. B: Biol. Sciences* **266**, 1623–1628.
- Liaw, A., and Wiener, M. (2002). "Classification and regression by randomForest," *R News* **2**, 18–22.
- Lovell, S. F., and Lein, M. R. (2004). "Song variation in a population of Alder Flycatchers," *J. Field Ornithology* **75**, 146–151.
- Macknight, F. L., Cato, D. H., Noad, M. J., and Grigg, G. C. (2001). "Qualitative and quantitative analyses of the song of the east Australian population of humpback whales," *Mem. Queensland Museum* **47**, 525–537, available at <https://biostor.org/reference/239799>.
- Magnúsdóttir, E. E., and Lim, R. (2019). "Subarctic singers: Humpback whale (*Megaptera novaeangliae*) song structure and progression from an Icelandic feeding ground during winter," *PLoS One* **14**, e0210057.
- Mathevon, N., Koralek, A., Weldele, M., Glickman, S. E., and Theunissen, F. E. (2010). "What the hyena's laugh tells: Sex, age, dominance and individual signature in the giggling call of *Crocuta crocuta*," *BMC Ecol.* **10**, 9.
- Mesoudi, A., Chang, L., Dall, S. R. X., and Thornton, A. (2016). "The evolution of individual and cultural variation in social learning," *Trends Ecol. Evol.* **31**, 215–225.
- Miksis-Olds, J. L., Buck, J. R., Noad, M. J., Cato, D. H., and Stokes, M. D. (2008). "Information theory analysis of Australian humpback whale song," *J. Acoust. Soc. Am.* **124**, 2385–2393.
- Miller, P. J. O., Biassoni, N., Samuels, A., and Tyack, P. L. (2000). "Whale songs lengthen in response to sonar," *Nature* **405**, 903–903.
- Mol, C., Bolhuis, J. J., and Moorman, S. (2021). "Vocal learning in songbirds: The role of syllable order in song recognition," *Philos. Trans. R. Soc. B* **376**, 20200248.
- Murray, A., Dunlop, R. A., Noad, M. J., and Goldizen, A. W. (2018). "Stereotypic and complex phrase types provide structural evidence for a multi-message display in humpback whales (*Megaptera novaeangliae*)," *J. Acoust. Soc. Am.* **143**, 980–994.
- Neubauer, R. L. (1999). "Super-normal length song preferences of female zebra finches (*Taeniopygia guttata*) and a theory of the evolution of bird song," *Evol. Ecol.* **13**, 365–380.
- Noad, M. J. (2002). *The Use of Song by Humpback Whales (Megaptera Novaeangliae) during Migration off the East Coast of Australia* (University of Sydney, Sydney, NSW, Australia).
- Noad, M. J., and Cato, D. H. (2007). "Swimming speeds of singing and non-singing humpback whales during migration," *Mar. Mamm. Sci.* **23**, 481–495.
- Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., and Jenner, K. C. S. (2000). "Cultural revolution in whale songs," *Nature* **408**, 537–537.
- Noad, M. J., Cato, D. H., and Stokes, M. D. (2004). "Acoustic tracking of humpback whales: Measuring interactions with the acoustic environment," in *Acoustics 2004: Transportation Noise and Vibration—The New Millennium: Proceedings of the Annual Conference of the Australian Acoustical Society*, Darlinghurst, Australia, pp. 353–358.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. (2017). *Vegan: Community ecology package*.
- Quattara, K., Lemasson, A., and Zuberbühler, K. (2009a). "Campbell's monkeys concatenate vocalizations into context-specific call sequences," *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22026–22031.
- Quattara, K., Lemasson, A., and Zuberbühler, K. (2009b). "Campbell's monkeys use affixation to alter call meaning," *PLoS One* **4**, e7808.
- Owen, C., Rendell, L., Constantine, R., Noad, M. J., Allen, J., Andrews, O., Garrigue, C., Poole, M. M., Donnelly, D., Hauser, N., and Garland, E. C. (2019). "Migratory convergence facilitates cultural transmission of humpback whale song," *R. Soc. Open Sci.* **6**, 190337.
- Payne, K., and Payne, R. (2010). "Large scale changes over 19 years in songs of humpback whales in Bermuda," *Z. Tierpsychol.* **68**, 89–114.
- Payne, K., Tyack, P. L., and Payne, R. (1983). "Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): A detailed analysis of two seasons in Hawaii," in *Communication and Behavior of Whales*, edited by R. Payne (Westview Press, Boulder, CO), pp. 9–57.
- Payne, R., and McVay, S. (1971). "Songs of humpback whales," *Science* **173**, 585–597.
- Perazio, C. E., Zapetis, M. E., Roberson, D., Botero, N., and Kuczaj, S. (2018). "Humpback whale, *Megaptera novaeangliae*, song during the breeding season in the Gulf of Tribugá, Colombian Pacific," *Madagascar Conserv. Dev.* **13**, 83–90.
- Petrusková, T., Pišvejcová, I., Kinštová, A., Brinke, T., and Petrussek, A. (2016). "Repertoire-based individual acoustic monitoring of a migratory passerine bird with complex song as an efficient tool for tracking territorial dynamics and annual return rates," *Methods Ecol. Evol.* **7**, 274–284.

- Pettitt, B. A., Bourne, G. R., and Bee, M. A. (2013). "Advertisement call variation in the golden rocket frog (*Anomaloglossus beebei*): Evidence for individual distinctiveness," *Ethology* **119**, 244–256.
- Rankin, S., Archer, F., and Barlow, J. (2013). "Vocal activity of tropical dolphins is inhibited by the presence of killer whales, *Orcinus orca*," *Mar. Mamm. Sci.* **29**, 679–690.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
- Reby, D., Joachim, J., Lauga, J., Lek, S., and Aulagnier, S. (1998). "Individuality in the groans of fallow deer (*Dama dama*) bucks," *J. Zool.* **245**, 79–84.
- Rekdahl, M. L. (2012). *Humpback Whale Vocal Communication: Use and Stability of Social Calls and Revolutions in the Songs of East Australian Whales*, University of Queensland.
- Rekdahl, M. L., Dunlop, R. A., Noad, M. J., and Goldizen, A. W. (2013). "Temporal stability and change in the social call repertoire of migrating humpback whales," *J. Acoust. Soc. Am.* **133**, 1785–1795.
- Röhr, D. L., Camurugi, F., Paterno, G. B., Gehara, M., Juncá, F. A., Álvares, G. F., Brandão, R. A., and Garda, A. A. (2020). "Variability in anuran advertisement call: A multi-level study with 15 species of monkey tree frogs (Anura, Phyllomedusidae)," *Can. J. Zool.* **98**, 495–504.
- Ruegg, K., Slabbekoorn, H., Clegg, S., and Smith, T. B. (2006). "Divergence in mating signals correlates with ecological variation in the migratory songbird, Swainson's thrush (*Catharus ustulatus*)," *Mol. Ecol.* **15**, 3147–3156.
- Salmi, R., Hammerschmidt, K., and Doran-Sheehy, D. M. (2014). "Individual distinctiveness in call types of wild western female gorillas," *PLoS One* **9**, e101940.
- Sandoval, L., Méndez, C., and Mennill, D. J. (2014). "Individual distinctiveness in the fine structural features and repertoire characteristics of the songs of white-eared ground-sparrows," *Ethology* **120**, 275–286.
- Searcy, W. A. (1992). "Song repertoire and mate choice in birds," *Am. Zool.* **32**, 71–80.
- Seddon, N., and Tobias, J. A. (2010). "Character displacement from the receiver's perspective: Species and mate recognition despite convergent signals in suboscine birds," *Proc. R. Soc. London, Ser. B: Biol. Sciences* **277**, 2475–2483.
- Smith, J. N. (2009). *Song Function in Humpback Whales (Megaptera Novaeangliae): The Use of Song in the Social Interactions of Singers on Migration* (The University of Queensland, Brisbane).
- Smith, J. N., Goldizen, A. W., Dunlop, R. A., and Noad, M. J. (2008). "Songs of male humpback whales, *Megaptera novaeangliae*, are involved in intersexual interactions," *Anim. Behav.* **76**, 467–477.
- Smouse, P. E., Long, J. C., and Sokal, R. R. (1986). "Multiple regression and correlation extensions of the Mantel test of matrix correspondence," *System. Biol.* **35**, 627–632.
- Sokal, R. R., and Rohlf, F. J. (1962). "The comparison of dendrograms by objective methods," *Taxon* **11**, 33–40.
- Soltis, J., Leong, K., and Savage, A. (2005). "African elephant vocal communication II: Rumble variation reflects the individual identity and emotional state of callers," *Anim. Behav.* **70**, 589–599.
- Suzuki, R., and Shimodaira, H. (2006). "Pvclust: An R package for assessing the uncertainty in hierarchical clustering," *Bioinformatics* **22**, 1540–1542.
- Suzuki, T. N., Wheatcroft, D., and Griesser, M. (2016). "Experimental evidence for compositional syntax in bird calls," *Nat. Commun.* **7**, 10986.
- Templeton, C. N., Laland, K. N., and Boogert, N. J. (2014). "Does song complexity correlate with problem-solving performance in flocks of zebra finches?," *Anim. Behav.* **92**, 63–71.
- Thornton, A., and Lukas, D. (2012). "Individual variation in cognitive performance: Developmental and evolutionary perspectives," *Philos. Trans. R Soc. B* **367**, 2773–2783.
- Vannoni, E., and McElligott, A. G. (2008). "Low frequency groans indicate larger and more dominant fallow deer (*Dama dama*) males," *PLoS One* **3**, e31113.
- Williamson, M. J., Kavanagh, A. S., Noad, M. J., Kniest, E., and Dunlop, R. A. (2016). "The effect of close approaches for tagging activities by small research vessels on the behavior of humpback whales (*Megaptera novaeangliae*)," *Mar. Mamm. Sci.* **32**, 1234–1253.
- Winn, H. E., and Winn, L. K. (1978). "The song of the humpback whale *Megaptera novaeangliae* in the West Indies," *Mar. Biol.* **47**, 97–114.